Journal of Financial Econometrics

Volume 19, Number 1, Winter 2021

Special Issue Articles on Predictive Modeling, Volatility, and Risk Management in Financial Markets: In Memory of Peter F. Christoffersen (Part II)

Articles

Continued on back cover

https://academic.oup.com/jfec

ISSN (Print) 1479-8409
ISSN (Online) 1479-8417

OXFORD UNIVERSITY PRESS

Official journal of the Society for Financial Econometrics

Aims and Scope

The goal of *Journal of Financial Econometrics* is to reflect and advance the relationship between econometrics and finance, both at the methodological and at the empirical levels. Estimation, testing, learning, prediction and calibration in the framework of asset pricing or risk management represent the core focus. The scope includes topics relating to volatility processes, return modelling, dynamic conditional moments, machine learning, big data, fintech, extreme values, long memory, dynamic mixture models, endogenous sampling transaction data, and microstructure of financial markets.

JOURNAL OF FINANCIAL ECONOMETRICS

VOLUME 19, NUMBER 1
WINTER 2021

Contents

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 1–38
doi: 10.1093/jjfinec/nbaa038
Advance Access Publication Date: 21 December 2020
Article

News and Idiosyncratic Volatility: The Public Information Processing Hypothesis*

Robert F. Engle[1], Martin Klint Hansen[2], Ahmet K. Karagozoglu[3] and Asger Lunde[4]

[1]New York University, [2]Novo Nordisk, [3]Hofstra University; New York University and [4]CREATES, Aarhus University

Address correspondence to Asger Lunde, Professor of Economics, Department of Economics and Business, Aarhus University, Fuglsangs Allé 4, 8210 Aarhus V, Denmark, or e-mail: alunde@econ.au.dk.

Received September 4, 2019; revised August 22, 2020; editorial decision September 15, 2020; accepted October 1, 2020

Abstract

Motivated by the recent availability of extensive electronic news databases and advent of new empirical methods, there has been renewed interest in investigating the impact of financial news on market outcomes for individual stocks. We develop the information processing hypothesis of return volatility to investigate the relation between firm-specific news and volatility. We propose a novel dynamic econometric specification and test it using time series regressions employing a machine learning model selection procedure. Our empirical results are based on a comprehensive dataset comprised of more than 3 million news items for a sample of 28 large U.S. companies. Our proposed econometric specification for firm-specific return volatility is a simple mixture model with two components: public information and private processing of public information. The public information processing component is defined by the contemporaneous relation with public information and volatility, while the private processing of public information component is specified as a

* This article was previously circulated under the title "And Now, The Rest of the News: Volatility and Firm Specific News Arrival." We thank René Garcia, the Editor, and two anonymous referees for their helpful comments. We are grateful to Torben Andersen, Tim Bollerslev, Christian Brownlees, Nikolaus Hautsch, Yu Jun, Andrew Patton, Peter Phillips, Kevin Sheppard, and Carsten Tanggaard for helpful discussions. We are also thankful for helpful comments received at the HUKU 2011 conference in Copenhagen, and at seminars at CREATES, Copenhagen Business School and Duke University Singapore Management University, Brunel University, Lancaster University and at the 2015 ERIM seminar at Erasmus University Rotterdam. Earlier version of this article was completed while the second author was at Aarhus University. Engle thanks the Volatility and Risk Institute at NYU and its sponsors NSF (#2018923) and NBIM (S0611 - A Financial Approach to Climate Risk 2019-06-12) for financial and intellectual support. Hansen acknowledges financial support from the Bikuben, Oticon, and Tuborg foundations. Hansen and Lunde acknowledge financial support from the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation. Karagozoglu acknowledges the special research leave afforded by Hofstra University. All errors are ours.

general autoregressive process corresponding to the sequential price discovery mechanism of investors as additional information, previously not publicly available, is generated and incorporated into prices. Our results show that changes in return volatility are related to public information arrival and that including indicators of public information arrival explains on average 26% (9–65%) of changes in firm-specific return volatility.

Key words: firm-specific news, realized volatility, public information arrival, information processing hypothesis of return volatility

JEL classification: G14

Empirical research confronting the paradigm that changes in stock prices are related to the arrival of new economic public information has reached mixed conclusions. While both financial theory and empirical research suggest that unanticipated public information affects stock prices, early research using all observable macroeconomic and industry-specific information has found high levels of inexplicable price volatility (Roll, 1988).

Initially, research, spurred by the advent of event study methodology, explored how corporate information events relate to changes in stock prices. Among others, these efforts led to the first studies using aggregate news counts to examine the relationship between market activity and the flow of public information (Ederington and Lee, 1993, Berry and Howe, 1994, Mitchell and Mulherin, 1994). Motivated by the recent availability of extensive electronic news databases and proliferation of text analysis methods, current branches of research investigate the impact of financial news on market outcomes for individual stocks and market indexes. Except for relatively few articles, literature on the role of media in financial markets focuses on the relationship between news and direction of returns or other market activity.[1]

Our research belongs to a fairly recent branch of literature that examines how news impacts firm-specific stock price volatility, for example, Boudoukh et al. (2019), Calomiris and Mamaysky (2019), Glasserman and Mamaysky (2019) and, in a special case, to the research by Jeon, McCurdy, and Zhao (2019) who investigate the relation between news and jumps in the stock returns, especially the jump size volatility.

In this article, we develop the information processing hypothesis of return volatility to investigate the relation between firm-specific news and volatility. We propose a novel dynamic econometric specification and test it using time series regressions employing a machine learning model selection procedure. Our empirical results are based on a comprehensive dataset comprised more than 3 million news items identified with 83 subject categories, as well as intraday prices and firm-specific information obtained from five

1 For example, Tetlock (2007), (2011), Tetlock, Saar-Tsechansky, and Macskassy (2008), Hillert, Jacobs, and Müller (2014) investigate investor sentiment and market activity. Fang and Peress (2009), Dougal et al. (2012), Heston and Sinha (2017), Engelberg, Mclean, and Pontiff (2018) examine news coverage and return predictability. Engelberg (2008), Dougal et al. (2012), Brogaard et al. (2014), Foucault et al. (2016) focus on news and trading. While Kalev et al. (2004), Maheu and McCurdy (2004), Lumsdaine (2010), Dzielinski and Hasseltoft (2017), Neuhierl, Scherbina, and Schlusche (2013) are among the earlier studies that consider the how news impacts price volatility.

databases covering the period from January 2001 through July 2009 for a sample of 28 large U.S. companies. Our primary contribution originates from the unique hypothesis we develop to directly test the impact of firm-specific news on firm-specific volatility for which we present economically significant results.

The economic model behind our hypothesis is motivated by market microstructure and behavioral finance as well as the information asymmetry and asset pricing literature. We build our hypothesis by synthesizing market microstructure theory, among others, which relates trading and news: news-induced information asymmetry among market participants due to information processing skills (Kim and Verrecchia, 1991, 1994), differences in opinion (Kandel and Pearson, 1995) as well as behavioral finance theories of momentum that relates market participants' reaction and news: slow diffusion of information (Hong and Stein, 1999, Hong, Lim, and Stein, 2000), attention (Barber and Odean, 2008), information processing time (Engelberg, 2008), cognitive biases (Antoniou, Doukas, and Subrahmanyam, 2013).

According to Rubinstein (1993), "differences in consumer behavior are often attributed to varying intelligence and ability to process information; and agents reading the same morning newspapers with the same stock price lists will interpret the information differently." Grossman (1986) suggests that the "private processing of public information" can lead to private information. Green (2004) indicates that the release of public information raises the level of information asymmetry, which is consistent with the interpretation that some market participants have an advantage at information processing.

Financial information research faces two key challenges (Boudoukh et al., 2007). The first challenge is to identify and observe all relevant fundamental information for a specific financial asset. The second challenge consists of correctly quantifying and measuring the fundamental. This article builds on existing financial information research by investigating the relationship between economic information arrival and changes in stock return volatility. One of our contributions is to identify a collectively exhaustive measure of firm-specific news flow by collecting all firm-specific news in the cross-section of approximately 30,000 different news sources accessible through the Dow Jones Factiva database. Our dataset is one of the most comprehensive news datasets employed in the literature.[2]

Although closely related, our analysis differs substantially from the investigations by Boudoukh et al. (2019), who examine the stock return variance ratios in periods with relevant news, and by Glasserman and Mamaysky (2019), who investigate monthly volatilities interacted with news sentiment.[3]

Glasserman and Mamaysky (2019) report that unusual news is reflected in volatility slower at the aggregate level than at the company-specific level which leads them to pose the following questions: "Why is volatility-relevant information in news not absorbed more quickly; and why is it absorbed more slowly at the aggregate level than at the company-specific level?" Glasserman and Mamaysky (2019) suggest that the "observed responses of volatility to news may be explained by attention constraints on investors." Our empirical

2 Boudoukh et al. (2019) use only the news articles contained in the Dow Jones Newswire while the Dow Jones Factiva database we use covers substantially more news sources in addition to the Dow Jones Newswire.

3 Calomiris and Mamaysky (2019) focus on a forecasting model using the context-specific measures of newsflow into risk and return in different equity markets around the world.

results for testing the information processing hypothesis of return volatility provide convincing explanation for the question posed by Glasserman and Mamaysky (2019).

Public information is incorporated contemporaneously while additional information, previously not publicly available, is generated from the private processing of public information and therefore incorporated sequentially. Alternative interpretation is to distinguish between the announcement effects of public news, which are contemporaneous and lagged effects due to the processing of those announcements, which would be sequential.[4] Our terminology of "private processing of public information" is private in the sense that investors do not publish how they process and interpret the public information. We refer to this specification as the information processing hypothesis of return volatility and test it using time series regressions. For a sample of 28 large U.S. stocks, we construct indicators of economic information arrival and investigate their relationship with measures of firm-specific realized variance by use of time series regression. Our approach allows us to test whether public information arrival is related to increases in volatility.

We find that the model identified by our machine learning model selection procedure includes lagged news indicators alongside contemporaneous news indicators. We interpret this effect as evidence that after the arrival of public information, additional information, previously not publicly available, is generated and incorporated into asset prices via superior processing of this public information.

Results from the three robustness checks we perform confirmed that our measures of information arrival capture news relevant for a particular stock and that, in general, time series regressions of firm-specific volatility on firm-specific information arrival provide evidence that, *ex-post*, large changes in return volatility.

Section 1 develops our information processing hypothesis of return volatility. In Section 2, we present a novel time series representation for estimating the components of return volatility, which we use for testing the relationship between the arrival of economic information and firm-specific return volatility. Section 3 discusses how we measure the economic information arrival using our vast news database. In Section 4, we present our data including our sources, details of variables, news indicator choices, and data mining techniques used in the analysis. Section 5 presents our results and contrasts them with previous studies. We support our investigation with a series of robustness checks in Section 6 and conclude in Section 7.

1 Information Processing Hypothesis of Return Volatility

This section introduces a return specification describing sources of equity volatility. We conclude with a description of the model's intuition and its relationship with established financial information theory.

In this investigation, we use a simple market model to decompose total stock return variance into a component that is firm-specific and a component that is common to the market. Throughout our investigation, we assume that this approach is capable of isolating the variance associated with revisions in expected future cash flows. In other words, we equate idiosyncratic variance from a market model with the variance of stock returns associated with firm-specific revisions in expected future cash flows.

4 We thank the anonymous referee for providing this interpretation for our hypothesis.

Before proceeding, it is important to note that this investigation uses the term information to mean facts, knowledge, or intelligence. This is an important definition since it is important to distinguish it from the concept of information in an econometric sense. Although the term information set will not be used throughout the article, it is an implicit assumption that everything is conditioned on the econometricians' information set. That said, it is also important to note that different types of information will be discussed. Public information will refer to facts, knowledge, or intelligence that is published into the public domain by news sources such as news wires, newspapers, press release wires, and others. Private information will refer to information generated by investors from the processing of information in the public domain. Information processing in turn refers to the collection of facts, knowledge, or intelligence and its subsequent examination, investigation, study, or analysis in the context of setting new expectations for future cash flows of the company in question. This study does not equate private information with inside information. Inside information may, however, be a source of contextual information for the processing of new public information. However, our investigation does not treat inside information as a separate object of analysis.

In this article, we consider equities, financial instruments linked directly to the economic performance of a company. Generally, when we think of information about a company's economic performance, we use a broad typology with three main categories: General and macroeconomic; industry; and company-specific. Changes in expectations about firm's future cash flows and returns occur due to the arrival of new information. The expected return includes a contemporaneous relation with market wide returns, $r_{m,t}$, such that unexpected returns only correspond to a firm-specific component. Let, i denote a given company, $r_{i,t}$ the return on its common stock, and $\varepsilon_{i,t}$ the unexpected firm-specific return component, then a specification of the return generating process we consider is

$$r_{i,t} - E_{t-1}(r_{i,t}) = r_{i,t} - \mu_t - \beta_t r_{m,t} = \varepsilon_{i,t} \tag{1}$$

where β_t measures the response of $r_{i,t}$ to the return on the chosen market portfolio, $r_{m,t}$ and μ_t is an excess return.[5]

Following French and Roll (1986), we include a distinction between private and public information so that returns have a public information component and a component due to private processing of public information. Applying this conceptual breakdown and specifying the dynamics of each component, $\varepsilon_{\text{private},t}$ and $\varepsilon_{\text{public},t}$, we consider the following simple mixture model in terms of the firm-specific return variance, for simplicity we drop the notation, i, identifying a given company:

$$\begin{aligned} \varepsilon_t &= \varepsilon_{\text{private},t} + \varepsilon_{\text{public},t} \\ \varepsilon_{\text{private},t} &= \sigma_t \epsilon_{1,t} \text{ with } \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \theta \sigma_{t-1}^2 \\ \varepsilon_{\text{public},t} &= \sqrt{\left(\sum_{k=1}^{K} \delta_k n_{t,k}\right)} \epsilon_{2,t} \end{aligned} \tag{2}$$

5 We have written μ_t and β_t as time varying to be consistent with the evidence regarding temporal variation in betas as demonstrated by, for example, Bali and Engle (2010), Patton and Verardo (2012), and Hansen, Lunde, and Voev (2014).

$$\epsilon_{1,t}, \epsilon_{2,t} \sim iid(0,1),$$

where $n_{t,k}$, for the time being, is defined broadly as an indicator of public information arrival. In the case where we have only one type of public information and a time-invariant effect on expected returns and expected cashflows, we have

$$\varepsilon_t = \sigma_t \epsilon_{1,t} + \sqrt{\delta n_t}\epsilon_{2,t}, \tag{3}$$

which is a specification that captures the private processing component of public information, σ_t, as a GARCH (1,1) process and the relationship with the arrival of public information as being contemporaneous. Any mispricing and other noise components are left entangled and unidentified in the two error terms $\epsilon_{1,t}$ and $\epsilon_{2,t}$. The conditional variance of the idiosyncratic return component will then be

$$\mathrm{Var}(\varepsilon_t|\Omega_{t-1}, n_t, x_t) = \sigma_t^2 + \delta n_t, \tag{4}$$

where Ω_{t-1} contain the history of the firm's stock return, and x_t is a vector of exogenous information such as general market conditions as captured by the return on the market, $r_{m,t}$. Since ε_t is defined as per Equation (1), its variance will also be conditioned on x_t although it does not appear explicitly on the right-hand side of Equation (4). Our model does not imply a causal relationship between volatility and news items. Rather, we suggest that news items and volatility share a common latent factor, new economic information.

While there are several possibilities for econometric implementation of the specification suggested in Equation (4), time series regression provides the flexibility necessary for our research question. Our chosen approach, which we present in the next section, is inspired by the ability to rewrite the specification from Equation (4) in terms of an autoregressive distributed lag model, where we simply replace $\mathrm{Var}(\varepsilon_t|\Omega_{t-1}, n_t, x_t)$ with its realized counterpart, the firm-specific realized variance.

We refer to this specification as the information processing hypothesis of return volatility. The intuition follows in two steps. First, market participants evaluate the signal inherent in new economic information and, due to intense competition between numerous informed traders, incorporate an imprecise and unbiased estimate into market prices almost immediately. Next, investors pursuing active information-based strategies conduct further economic assessments and revise their expectations and portfolios accordingly. In what can be characterized as an adaptive price discovery mechanism triggered by the arrival of public information, this behavior results in stock return volatility decreasing over time. Lagged effects of economic information arrival will be related to information processing activities by investors.

Our information processing hypothesis is in line with three mechanisms modeled by financial information theory. First, when multiple informed traders observe the exact same signal, prices may reflect new information almost instantaneously (Holden and Subrahmanyam, 1992). Second, skilled processing of public information may identify profitable trading opportunities (Admati and Pfleiderer, 1988). Third, a large information advantage of multiple informed traders, combined with the subsequent arrival of informative public information, may lead to decreasing return volatility on an interday basis (Foster and Viswanathan, 1990).

Admati and Pfleiderer (1988) examine how informed traders and liquidity traders interact. The information structure of their model allows the interpretation that skilled

information processors will be able to identify profitable trading opportunities due to faster and more efficient processing of public information. They show that endogenous information acquisition intensifies the concentration of trading and that increased competition among informed traders increases price informativeness. Foster and Viswanathan (1990) present a model where the quality of public information available to noninformation-based traders determines the decay rate of the information advantage.

Theory and anecdotal evidence support our specification for the information processing hypothesis of return volatility. To summarize, it suggests that information observed by a broad set of market participants is likely to be incorporated almost instantaneously. In addition, the arrival of public information creates a trading opportunity for investors with exceptional information processing capabilities. The information advantage is largest in the period after public information arrival and decreases with time as multiple information-based investors compete with each other. The speed of variance decline may be further increased by the arrival of public information providing analysis of previously disclosed economic information. Conclusively, our specification captures the intuition that, after the arrival of public information, the change in stock return volatility will reflect changes in the rate at which additional information, previously not publicly available, is generated via superior processing of new public information, and is revealed in prices.

In a broader context, established financial research suggests that there are several possible sources of volatility. The prevailing hypothesis involves the arrival of new information. The mixture of distributions hypothesis (MDH) suggests that volume and volatility are driven by a common factor, which can be interpreted as the rate of information arrival (Clark, 1973). If the underlying information flow is clustered, then this will lead to clustering in volatility. The MDH is general and can be thought of as subsuming several more detailed hypotheses that differ primarily in their suggested explanation for the source of volatility clustering.

An alternate perspective interprets volatility as economic uncertainty related to the fundamental value of the underlying assets of the firm (Merton, 1974; Black, 1976; Christie, 1982). Some theories suggesting volatility as a measure of economic uncertainty reverse the causal relation and propose that news is the result of the media producing news items to resolve uncertainty or information asymmetry (Veldkamp, 2006). A possible source of volatility not directly related to information is noise. Noise as a source of volatility arises from trading by noninformation-based traders (Black, 1986). Noise therefore includes any irrational reason for trading as well as trades motivated by liquidity needs. We can consider the noise hypothesis as covering all irrational and noninformation-based sources of volatility.

Based on the market microstructure models of Glosten and Milgrom (1985) and Andersen (1996), Chua and Tsiaplias (2019) indicate that the asset returns are largely due to the relationship between the arrival of news and the underlying mechanism that processes the new information into prices. While referring to the same models, Maheu and McCurdy (2004) suggest that varying information content of news may have different effects on volatility.

In this light, our model captures the salient hypotheses of the sources of return volatility. The specification allows us to nest all of the above hypotheses in one model. This article presents a simple empirical investigation where firm-specific volatility is regressed on firm-specific news. While our investigation will not enable us to say anything about the specific

behavior of agents subject to public information arrival, results from time series regressions of our two measures of interest, news, and volatility, provide evidence of whether the arrival of information is related to changes in volatility.

2 Estimating Components of Return Volatility under Information Processing Hypothesis

In our analysis, the variable of interest is idiosyncratic variance, which captures firm-specific price movements within the trading day. We refer to this measure as the daily firm-specific variance (FV_t) and suggest that it corresponds to a realized version of $Var(\varepsilon_t|\Omega_{t-1}, n_t, x_t)$. We estimate it by:

$$\mathrm{FV}_{i,t} = \mathrm{RV}_t^i - \beta_t^2 \mathrm{RV}_t^{\mathrm{SPY}}, \text{ using } \beta_t = \frac{R\mathrm{Cov}_t^{i,\mathrm{SPY}}}{\mathrm{RV}_{\mathrm{t}}^{\mathrm{SPY}}}, \tag{5}$$

where the individual parts are computed using intraday data so that $\mathrm{RV}_{\mathrm{t}}^{\mathrm{i}}, \mathrm{RV}_t^{\mathrm{SPY}}$, and $R\mathrm{Cov}_t^{i,\mathrm{SPY}}$ represent the covariance matrix between the individual asset and the S&P 500 index as represented by the SPY exchange traded fund, ensuring FV_t is always positive. We interpret β_t as a realized beta and $\mathrm{RV}_{\mathrm{t}}^{\mathrm{SPY}}$ as the realized variance of the market index.[6]

We now present a novel time series specification for testing the information processing hypothesis of return volatility. First, we discuss how our time series specification allows us to estimate and test the significance of two components of equity volatility: public information and private processing of public information.

We estimate the public information component in order to test whether the arrival of economic information is related to contemporaneous changes in FV_t . We interpret a significant proportion of covariation between realized variance and contemporaneous indicators as implying that prices are moving in conjunction with the arrival of relevant information about the firm's economic prospects.

In addition, we test if private processing of public information occurs following the arrival of economic information. We argue that any lagged effects of economic information arrival will be related to information processing activities of information-based investors. For example, institutional investors and investment managers pursuing active investment management strategies will process newly arrived information, evaluate its impact on company value, infer the initial impact estimate already incorporated in prices, and finally they will devise trading strategies to exploit the derived signal.

Furthermore, recent research into media, news, and financial markets has brought forth an alternate hypothesis for the origin of price movements and market activity. It has been suggested that a proportion of stock returns and market activity may be driven by a media effect reflecting the attention that a company achieves in the media (Huberman and Regev, 2001, Barber and Odean, 2008, Engelberg and Parsons, 2011, Hillert, Jacobs, and Müller,

6 The realized measures are computed by aggregating squared 5-min returns within each trading day as done in Andersen et al. (2001). In addition, the realized covariance matrix between the individual asset and the market index is estimated using *Refresh Time* sampling as discussed in Barndorff-Nielsen et al. (2011). All our realized measures are based on transaction prices according to the cleaning rules presented in Barndorff-Nielsen et al. (2009).

2014[7]). In order to control for media effects, we include a set of indicators measuring increases in media attention. Our aim is to disentangle media attention effects from economic content effects.

We test the information processing hypotheses with the following autoregressive distributed lag specification:

$$\ln(\mathrm{FV_t}) = \omega + \sum_{k=1}^{K} \rho_k \ln(\mathrm{FV_{t-k}}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t \tag{6}$$

where we add indicators of public information arrival, *ni, t*, that we believe may either have economic relevance or measure media attention. This empirical approach is motivated both by the aim of our investigation and the complexity arising from the number of indicators, which we intend to consider in our estimation efforts. The specification in Equation (6) has three components of particular interest: an autoregressive part, $\sum_{k=1}^{K} \rho_k \ln(\mathrm{FV_{t-k}})$; contemporaneous news; $\sum_{i=1}^{M} \gamma_i n_{i,t}$; and lagged effects of news, $\sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j}$.

This time series representation in Equation (6) is a convenient way of investigating how FV_t is related to the arrival of different types of economic information. Our central hypothesis is that public information is incorporated contemporaneously while additional information, previously not publicly available, resulting from the processing of new public information is incorporated sequentially through an adaptive mechanism where market prices gradually reveal this additional information. While the representation in Equation (6) is convenient for estimation, it is not the most intuitive for interpretation. The issue is that in Equation (6) lagged news can both influence the FV_t directly and indirectly through the lags of the dependent variable. To this end it is, after estimation, helpful to use the so-called final form for interpretation (Theil and Boot, 1962). This is obtained by eliminating all lagged values of $\ln(\mathrm{FV_t})$ by means of lags of the same equation. For simplicity, we just show what this looks like for $K = 1,\ M = 1$, and $J = 1$ (the general expression is of similar form):

$$\ln(\mathrm{FV}_t) = \omega + \rho \ln(\mathrm{FV}_{t-1}) + \gamma_1 n_{1,t} + \tilde{\gamma}_{1,1} n_{i,t-1} + \epsilon_t \tag{7}$$

which gives us

$$\ln(\mathrm{FV}_t) = \frac{\omega}{1-\rho} + \sum_{s=0}^{\infty} \rho^s \gamma_1 n_{1,t-s} + \sum_{s=0}^{\infty} \rho^s \tilde{\gamma}_{1,1} n_{1,t-1-s} + \sum_{s=0}^{\infty} \rho^s \epsilon_{t-s} \tag{8}$$

That we rewrite as

$$\ln(\mathrm{FV}_t) = \frac{\omega}{1-\rho} + \gamma_1 n_{1,t} + \sum_{s=0}^{\infty} (\rho^{s+1}\gamma_1 + \rho^s \tilde{\gamma}_{1,1}) n_{1,t-1-s} + \sum_{s=0}^{\infty} \rho^s \epsilon_{t-s} \tag{9}$$

So $\gamma_1 n_{1,t}$ represent the explanation provided by the contemporaneous news, and $\sum_{s=0}^{\infty} (\rho^{s+1}\gamma_1 + \rho^s \tilde{\gamma}_{1,1}) n_{1,t-1-s}$ what is explained by lagged news. Besides the R^2 from the estimated model we can compute a partial R^2 for these two terms that will tell us how much of the variation in the FV_t is explained with news (we denote this R^2_{news}). Also, of

7 Hillert et al. (2014) use news articles, between 1989 and 2010, obtained from the LexisNexis database relying on the "relevance score" provided by that database without identify the information/subject categories for each new item.

particular interest are the partial R^2 of the contemporaneous term (denoted by this $R^2_{\text{cont.}}$). We will use this formulation to enable our interpretation of the dynamic relationship between realized variance and economic information arrival.

3 Measuring Economic Information Arrival

Previous empirical work by Ederington and Lee (1993), Mitchell and Mulherin (1994), Berry and Howe (1994), and DeGennaro and Shrieves (1997) suggests that a relevant measure of public information arrival is a simple count of the number of news items. Hence, our indicator corresponds to

$$n_{i,t} \equiv c_{i,t} \tag{10}$$

where $c_{i,t}$ counts from time $t-1$ to t the number of news items for a given subject or media attention category.[8] Higher values will reflect a higher level of information arrival and content materiality. If media industry participants reliably evaluate the materiality of new information and consistently initiate editing and distribution of news items of material economic content, then it is reasonable to assume that a positive change in the number of news items will proxy for the revisions in expectations about future cash flows of the firm.

The structure of the media industry implies that most news items are the result of information release. This is consistent with the description of the news disclosure process provided by Thompson, Olsen, and Dietrich (1987). They describe how firms typically initiate firm-specific news stories through press releases and direct contact with journalists. New economic information is created by participants in the information environment of a firm.

Findings by Tetlock, Saar-Tsechansky, and Macskassy (2008) as well as by Engelberg (2008) suggest that there is relevant "soft" information in news stories that is not immediately reflected in prices. Furthermore, Dougal et al. (2012) show that, at least over short time horizons, journalists have the potential to influence investors' behavior, which suggests that the interpretation of public news is important. In summary, news items are the result of the activities of media industry participants as they edit, aggregate, and distribute raw economic information. Media industry participants choose the degree that items are edited and aggregated in order to fit the medium's distribution frequency (i.e., continuously, daily, weekly, etc.) and distribution form (e.g., electronic or print). As a result, positive changes in firm-specific newsflow, as we measure in the full cross-section of news sources, are a sufficient indicator reflecting the arrival of material unexpected information.

4 Data and Variable Construction

4.1 Data Sources

This subsection describes the characteristics of the data used in our investigation. We employ data from five databases covering the period from January 29, 2001 through July 31, 2009 for a sample of 28 large U.S. companies included in the Dow Jones Industrial Average for which we could obtain a sufficient amount of intraday returns and news counts between 2001 and 2009. Table 1 provides an overview of the characteristics of our sample. Table 1

8 All news items published after close of the exchange are transferred to the news count observation for the next trading day.

Table 1 Sample description

Company name	Exchange	Ticker	Headquarters	Largest business segment	(1)	(2)	(3)	(4)
Exxon Mobil Corp.	NYSE	XOM	Irving, TX, USA	Downstream–nonUnited States	70	United States	31	349.2
General Electric Co.	NYSE	GE	Fairfield, Connecticut, USA	Infrastructure	34	United States	55	310.8
Microsoft Corporation	NASDAQ	MSFT	Redmond, WA, USA	Windows and windows live	30	United States	65	285.4
Wal-Mart Stores Inc.	NYSE	WMT	Bentonville, AR, USA	Walmart United States	67	United States	79	214.9
Citigroup, Inc.	NYSE	C	New York, NY, USA	Institutional clients	51	North America	43	190.7
Pfizer Inc.	NYSE	PFE	New York, NY, USA	Biopharmaceutical	91	United States	54	187.9
Johnson and Johnson	NYSE	JNJ	New Brunswick, NJ, USA	Pharmaceutical	43	United States	56	177.3
Procter and Gamble Co.	NYSE	PG	Cincinnati, OH, USA	Household, fabric and home care	29	United States	45	164
IBM Corp.	NYSE	IBM	Armonk, NY, USA	Global services	54	United States	38	152.6
Bank of America Corp.	NYSE	BAC	Charlotte, NC, USA	Consumer and commercial banking	57	United States	92	146.8
Intel Corporation	NASDAQ	INTC	Santa Clara, CA, USA	PC client group	73	United States	20	143.3
AT&T, Inc.	NYSE	T	Dallas, TX, USA	Wireline	78	United States	100	134.7
Chevron Corporation	NYSE	CVX	San Ramon, CA, USA	Downstream	85	United States	51	128.9
JPMorgan Chase & Co.	NYSE	JPM	New York, NY, USA	Investment bank	34	United States	74	118.8
The Coca-Cola Company	NYSE	KO	Atlanta, Georgia, USA	Non-alcoholic beverages	100	North America	29	115.4
Verizon Comm. Inc.	NYSE	VZ	New York, New York, USA	Wireline	54	United States	99	105.9
Merck & Co. Inc.	NYSE	MRK	Whitehouse St., NJ, USA	Pharmaceutical	92	United States	59	100.2
Hewlett-Packard Comp.	NYSE	HPQ	Palo Alto, CA, USA	Personal systems group	30	United States	37	84.3
The Home Depot, Inc.	NYSE	HD	Atlanta, GA, USA	Retail-home improvement	99	United States	95	72.4
3M Co.	NYSE	MMM	St Paul, MN, USA	Industrial and transportation	28	United States	40	55.5

(continued)

Table 1 Continued

Company name	Exchange	Ticker	Headquarters	Largest business segment	(1)	(2)	(3)	(4)
American Express Comp.	NYSE	AXP	New York, NY, USA	U.S. card services	47	United States	72	54.9
United Technologies Corp.	NYSE	UTX	Hartford, CT, USA	Carrier	28	United States	53	52
Walt Disney Co.	NYSE	DIS	Burbank, CA, USA	Media networks	41	United States and Canada	79	51.2
McDonald's Corp.	NYSE	MCD	Oak Brook, IL, USA	Food service industry	100	United States	35	47.8
Boeing Co.	NYSE	BA	Chicago, IL, USA	Commercial airplanes	48	United States	65	44.3
EI DuPont de Nemours & Co.	NYSE	DD	Wilmington, DE, USA	Agriculture and nutrition	23	United States	42	40.2
Caterpillar Inc.	NYSE	CAT	Peoria, IL, USA	Machinery and engines	78	North America	45	31.5
Alcoa, Inc.	NYSE	AA	Pittsburgh, PA, USA	Primary metals	36	United States	58	24.1
				Avg. =	57.1		57.5	128.0

The table describes some characteristics of the 28 companies in our sample. They are among the firms included in the Dow Jones Industrial Average for which we could obtain a sufficient amount of intraday returns and news counts between 2001 and 2009. We identify the largest business units by calculating mean percentage revenue contribution over the sample period. In addition, we present the percentage revenue contribution of North American or U.S. revenues to total firm revenues excluding corporate adjustments. These two measures provide an overview of the activities, which the individual firm is involved in. The table is based on SEC filings obtained from the S&P Capital IQ database. The last four columns correspond to: (1) Average % of Total Revenues of Largest Business Segment; (2) Name of the Largest Geographic Revenue Segment; (3) Average % of Total Revenues of the Largest Geographic Revenue Segment; and (4) Average Market Capitalization in $billion over the years 2001–2009.

illustrates that the sample companies on average have 57.1% of revenues from their largest business unit, while 57.5% of revenues are from North American operations. In addition, the average market capitalization over the sample period across firms is $128 billion. These characteristics underscore the large size, business diversification, and international activities of companies in the sample.

Intraday data for each company's common stock are extracted from the Trade and Quote database. Firm-specific news items are collected from the Dow Jones Factiva database using the Dow Jones company codes, which we match with CUSIP identification codes as well as permanent company and security identification numbers from the Center for Research in Security Prices (CRSP) database. Daily returns and trading volume (TV) used for robustness investigations are from the CRSP. For our robustness analysis, we also use implied volatilities (IVs), obtained from Bloomberg, that are calculated from 1 month at-the-money call option prices. Company-specific information used in our discussion section comes from sources such as the U.S. Securities and Exchange Commission (SEC) filings and other reference data extracted from the S&P Capital IQ database.

4.2 Indicators of Economic Information Arrival

Our choice of information arrival indicators is linked directly to the economic rationale inherent in our model describing the microeconomic sources of equity volatility. This model states that unexpected returns may arise from changes in expected returns and changes in expected future cash flows. Our chosen variables are all considered indicators of the arrival of new economic information related to changes in expected returns or expected cashflows. We limit our choice of indicators of economic information arrival to a set of 83 subject categories, assigned by the Dow Jones Intelligent Indexing system, presented in Table 2.[9] These categories reflect a mutually exclusive set of corporate information categories covering an extensive set of corporate information events. In this way, news response coefficients can be partially disentangled into economic content effects as suggested in Engelberg and Parsons (2011). The subject categories are, in a taxonomic sense, mutually exclusive; however, this does not preclude that they may occur simultaneously or on the same day.

To control for the media attention effect, we include the 16 indicators displayed at the bottom of Table 2. These indicators measure the change in the level of news item counts in newspapers and other types of source formats. While the ideal measure would entail a measure of investor readership, we argue that our chosen variables control for the media attention effect, since they proxy for the distribution and proliferation of company-specific news. The media attention hypothesis and the information processing hypothesis are not mutually exclusive and may well coexist. Hence, we include proxies for media attention by measuring changes in the arrival of news. This allows us to explicitly test the importance of aggregate measures of media attention alongside proxies for different corporate information events.

Variation in the information content of news items suggests the relationship between market impact and news may vary depending on the nature of the news. We use the Dow Jones Intelligent Indexing categorization to distinguish between the content of news items. For example, one news item related to a corporate merger may have a larger impact than a

9 Heston and Sinha (2017) use subject codes assigned by an alternative data source, Thomson Reuters NewsScope, in predicting stock returns using news stories.

Table 2 Choice of news variables

Main DJII code and category name	No. of selected variables	List of subject category names
GCAT-general/ political	12	Crime/Courts, Disasters/Accidents, Environmental News, Climate Change, Global/World Issues, Health, Int. Relations, Domestic Politics, Regional Politics, Science/Technology, Weather, Labor Issues, Demographics
ECAT-economic	27	Economic Growth, Money Supply, Inflation/Prices, Personal Income/ Average Earnings, Consumer Sentiment, Consumer Credit/ Expenditure/Savings, Budget Account, Government Borrowing Requirement, Agricultural Production, Industrial Production, Capacity Utilization, Inventories, Factory Orders/Durable Goods, Employment Costs/ Productivity, Employment/Unemployment, Business Sentiment, Reserve Assets, Mortgage Applications/ Refinancing, Car Registrations/Vehicle Sales, Bankruptcy Figures, Index of Leading Economic Indicators, Housing Starts/Construction Figures, Home Sales/Housing Affordability, Economic/Monetary Policy, Government Finance, Trade/External Payments, Euro Zone/ Currency
MCAT-financial/ commodity markets	8	Money Markets, Foreign-Exchange News, Soft Commodity Markets, Metals Markets, Energy Markets, Fund Markets
CCAT-corporate/ industrial	36	Plans/Strategy, Corporate Crime/Legal/Judicial, Regulation/ Government Policy, Annual Meetings, Dividends, Sales Figures, Earnings Surprises, Analyst Comment/Recommendation, Internal Audit, Bankruptcy, Share Capital, Corporate Debt Instruments, Financing Agreements, Corporate Credit Ratings, Acquisitions/ Mergers/Takeovers, Divestitures/Asset Sales, Privatizations/ Nationalization, Joint Ventures, Output/Production, New Products/ Services, Research/Development, Capacity/Facilities, Information Technology, Product Safety, Marketing/Market Research, Government Contracts, Defense Contracts, Non-government Contracts, Licensing Agreements, Franchises, Outsourcing, Competition Issues, Management Issues, Labor Disputes, Lay-offs/ Redundancies, Natural Reserves/Resources Discovery
Economic information indicators	83	
Media attention indicators	16	Dow Jones News Service, Reuters Newswire, Washington Post, New York Times, Wall Street Journal, Associated Press Newswires, USA Today, Financial Times, Aggregate Newspaper, Aggregate Newswire, Aggregate Industry Publication, Aggregate Newsletters, Aggregate Press Release Wire, Aggregate Other Wire, Aggregate General News and Business Publication, Aggregate Government and Politics Publication
Total	99	

This table provides an overview of the subject categories from the Dow Jones Intelligent Indexing system that we use to create content-specific indicators of new economic information arrival. When creating indicators for a given company, we include only indicator series that have news arrival on at least 1% of the days in the sample period. Dow Jones Factiva provides more information on their Dow Jones Intelligent Indexing system on the following webpage: https://www.dowjones.com/products/factiva/#overview.

news item related to the launch of a new product. The information typology used by Roll (1988) provides a starting point for thinking about news categorization. General and macroeconomic information may include facts about the state of consumer, industrial, and other markets as well as news related to politics, government policy, regulation, and demographics. Industry-specific information may include news about access to resources, mergers, and acquisitions, and other industry information events. Firm-specific news may include information about corporate issues and activities, such as research and development, restructuring, litigation and arbitration, business/franchise performance, funding, as well as other corporate information events. With respect to Roll's (1988) information typology, the news in our dataset mainly covers industry- and firm-specific news items.

Since we aim to estimate the proportion of stock return volatility that is related to the arrival of economic information, we are concerned with the issue of news item endogeneity. Endogenous news is news items that include references to stock market information such as TV and price change alerts. We are only interested in news items that contain reference to some form of economic information related to the firm. In other words, we wish to filter out items that are automatically generated based on market activity. Other studies have approached this issue using filtering rules. Of all news considered to contain discussion of share price movements a subset also discusses events in the broad equity markets and the rest discuss events in other financial markets (steel, oil, natural gas, etc.).[10] Our examination of market news items reveals that purely endogenous news items, created due to abnormal market activity, account for ~5% of the corporate news flow and is more prevalent among a specific set of news sources. We find that news containing references to price movements can either be purely endogenous, triggered by trading activity, or include a summary or reference to corporate information events. The Dow Jones Intelligent Indexing system enables identification of the topics covered in an individual news item; hence, we can filter out purely endogenous news and focus on news items that provide economic content.

4.3 Sample Characteristics

Table 3 presents the characteristics of our measure of firm specific realized variance while Table 4 describes the characteristics of our news item dataset. Our sample starts after the decimalization occurred at the New York Stock Exchange (NYSE) and the National Association of Securities Dealers Automated Quotations exchange (NASDAQ). The NYSE completed decimalization on January 29, 2001 while NASDAQ completed the transition on April 9, 2001. Alternatively, the sample period for each firm starts from when the last major corporate event occurred as marked by a name change, merger, initial public offering, or similar. As can be seen in the last column of Table 3, this results in 919–2137 observations per company.

Table 3 shows that our variable of interest, daily FV_t, generally accounts for between 58% and 77% of all intraday volatility (refer to Column 8). This is consistent with results from market models at lower frequencies, for example, using daily data. For example, Roll (1988) finds that the market model's average R^2 across 96 large firms is 24%,

10 We therefore choose to remove all news that the Dow Jones Intelligent Indexing system labels as covering events in equity markets (M11), contains stock market pricing information (C1522), and does not contain any other type of economic information, in the form of a reference to other corporate information events.

Table 3 Characteristics of firm specific realized variance

	RV		FV		β		FV/RV		Start date	No. of observations
Ticker	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2		
AA	5.84	114.85	3.73	30.78	1.08	0.23	0.75	0.02	January 29, 2001	2137
AXP	5.23	108.87	2.98	35.37	1.05	0.21	0.65	0.03	January 29, 2001	2136
BA	3.29	21.09	2.13	5.78	0.89	0.09	0.72	0.03	January 29, 2001	2136
BAC	7.04	635.13	3.7	161.56	1	0.45	0.64	0.03	January 29, 2001	2136
C	9.34	1308.2	5.74	530.02	1.18	0.34	0.62	0.02	January 29, 2001	2136
CAT	3.58	32.9	2.06	6.02	1.02	0.14	0.68	0.03	January 29, 2001	2136
CVX	2.57	28.04	1.38	3	0.84	0.14	0.69	0.04	December 10, 2001	1963
DD	2.97	20.75	1.64	3.24	0.94	0.08	0.65	0.03	January 29, 2001	2136
DIS	3.71	45.83	2.36	19.5	0.89	0.1	0.72	0.03	January 29, 2001	2136
GE	3.78	71.19	1.98	20.06	0.98	0.1	0.6	0.03	January 29, 2001	2136
HD	3.87	31.36	2.28	6.69	1.04	0.12	0.68	0.02	January 29, 2001	2136
HPQ	3.82	29.76	2.52	11.29	1.02	0.11	0.71	0.02	May 6, 2002	1822
IBM	2.28	13.98	1.24	2.73	0.87	0.05	0.62	0.02	January 29, 2001	2136
INTC	4.46	27.64	2.34	6.01	1.39	0.18	0.58	0.02	April 9, 2001	2087
JNJ	1.55	6.89	1.06	2.87	0.56	0.06	0.75	0.02	January 29, 2001	2136
JPM	6.14	194.08	3.39	54.18	1.16	0.23	0.63	0.03	January 29, 2001	2136
KO	1.66	5.99	1.09	1.62	0.62	0.06	0.74	0.02	January 29, 2001	2136
MCD	2.58	14.49	1.85	5.95	0.75	0.09	0.76	0.02	January 29, 2001	2136
MMM	2.1	11.66	1.22	2.74	0.81	0.06	0.65	0.03	January 29, 2001	2136
MRK	2.8	20.52	1.98	8.76	0.71	0.11	0.77	0.02	January 29, 2001	2136
MSFT	2.75	14.78	1.4	2.71	1.02	0.1	0.6	0.02	April 9, 2001	2087
PFE	2.48	11.37	1.65	3.55	0.77	0.09	0.74	0.02	January 29, 2001	2136
PG	1.53	7.32	0.98	1.61	0.63	0.05	0.72	0.02	January 29, 2001	2136
T	3.43	38.86	1.86	8.29	0.88	0.1	0.68	0.03	December 1, 2005	919
UTX	2.55	18.68	1.51	5.05	0.86	0.08	0.67	0.03	January 29, 2001	2137
VZ	2.88	19.52	1.78	5.19	0.83	0.09	0.71	0.03	January 29, 2001	2137
WMT	2.23	10.36	1.34	3.25	0.84	0.06	0.66	0.02	January 29, 2001	2136
XOM	2.48	43.34	1.28	5.28	0.89	0.11	0.64	0.03	January 29, 2001	2136

This table reports descriptive statistics for our sample. The end date is July 31, 2009 for all stocks. In our analysis, the variable of interest is idiosyncratic variance, which captures firm-specific price movements within the trading day. We refer to this measure as the daily FV_t and estimate it as:

$$\mathrm{FV}_{\mathrm{i,t}} = \mathrm{RV}_{\mathrm{t}}^{\mathrm{i}} - \beta_t^2 \mathrm{RV}_{\mathrm{t}}^{\mathrm{SPY}}, \text{ using } \beta_t = \frac{R\mathrm{Cov}_{\mathrm{t}}^{\mathrm{i,SPY}}}{\mathrm{RV}_{\mathrm{t}}^{\mathrm{SPY}}},$$

$\mathrm{RV}_{\mathrm{t}}^{\mathrm{SPY}}$ and $R\mathrm{Cov}_{\mathrm{t}}^{\mathrm{i,SPY}}$ represent the covariance matrix between the individual asset and the index as represented by the SPY exchange traded fund, ensuring FV_t is always positive. We interpret β_t as a realized beta and $\mathrm{RV}_{\mathrm{t}}^{\mathrm{SPY}}$ as the realized variance of the market index. Realized measures are computed by aggregating squared 5-min returns within each trading day following Andersen et al. (2001). In addition, the realized covariance matrix between the individual asset and the market index is estimated using *Refresh Time* sampling as discussed in Barndorff-Nielsen et al. (2011). All realized measures are based on transaction prices according to the cleaning rules presented in Barndorff-Nielsen et al. (2009).

Table 4 News dataset characteristics

Ticker	No. of News sources	Gini (%)	Trading μ (S.E.)	Holiday μ (S.E.)	Weekend μ (S.E.)
AA	1922	85	14.1 (0.5)	8.5 (1.2)	2.4 (0.1)
AXP	2799	80	15.4 (0.3)	9 (1)	3.7 (0.2)
BA	3529	87	55.3 (1)	34.5 (3)	12 (0.5)
BAC	3143	86	45.5 (1.3)	19.5 (2.3)	8.4 (0.4)
C	3425	88	72.7 (1.5)	39.2 (3.9)	12.8 (0.5)
CAT	1682	83	7.8 (0.2)	3.9 (0.4)	2 (0.1)
CVX	2777	85	25.9 (0.5)	14.7 (1.6)	4.4 (0.2)
DD	2292	81	10.5 (0.2)	6.9 (0.7)	1.9 (0.1)
DIS	3618	87	38.8 (0.7)	27.4 (2.3)	14.5 (0.4)
GE	4716	86	66.1 (1.1)	48.4 (4.4)	15.7 (0.5)
HD	1668	83	11.2 (0.4)	5.8 (0.7)	3.1 (0.1)
HPQ	3328	84	35.3 (0.8)	23.2 (2.5)	6.6 (0.3)
IBM	3799	86	56.2 (1)	38.7 (3.6)	9.1 (0.4)
INTC	3304	87	53.6 (1.1)	31.2 (3.1)	8.5 (0.3)
JNJ	3130	86	27.9 (0.6)	20.9 (2.1)	6.3 (0.2)
JPM	3437	90	80.1 (1.4)	50.5 (4.8)	14.1 (0.5)
KO	2974	81	21.5 (0.4)	14.4 (1.4)	5.7 (0.2)
MCD	3303	80	20.3 (0.4)	14.4 (1.5)	6.3 (0.2)
MMM	1656	83	6.7 (0.2)	3.9 (0.5)	1.2 (0.1)
MRK	1795	84	12.8 (0.4)	7.7 (0.8)	2.2 (0.1)
MSFT	4493	87	102 (1.8)	63.6 (5.3)	22.6 (0.8)
PFE	2806	86	29.4 (0.7)	18.7 (1.9)	6.1 (0.3)
PG	3323	83	24 (0.6)	17.4 (1.7)	5.8 (0.4)
T	1868	86	16.7 (0.5)	7.3 (1.3)	2.9 (0.2)
UTX	2184	87	14.1 (0.3)	8.5 (0.8)	2.4 (0.1)
VZ	2468	88	34.1 (0.7)	17.2 (1.6)	6.1 (0.2)
WMT	3206	86	41.7 (0.8)	26.3 (2.2)	13.2 (0.4)
XOM	3373	86	46.5 (0.9)	27.9 (2.6)	7.9 (0.3)

This table provides an overview of our news data. Number of sources is computed based on the unique number of source codes available in our sample. News items are computed based on the unique items in our sample identified by headline, date, timestamp, sourcecode, and wordcount. Gini refers to the Gini coefficient computed based on the amount of unique news items per source code for a given company. The Gini coefficient measures concentration of news among observed sources and is calculated as $G = 1 - \frac{2}{n-1}\left(n - \frac{\sum_{i=1}^{n} iy^i}{\sum_{i=1}^{n} y^i}\right)$, where $y_i \leq y_{i+1}$. Trading, Holiday, and Weekend refer to the mean daily number of news items over the sample period occurring on each type of day.

corresponding to idiosyncratic return variability of 76%. The mean realized variance across stocks corresponds to annualized volatility in the range from 19.6 for Procter & Gamble (NYSE: PG) to 48.5 for Citigroup (NYSE: C). On the other hand, mean firm-specific realized variance translates to annualized volatility in the range from 15.7 for PG to 38.0 for C. Mean realized betas correspond to expected magnitudes.

Table 4 contains descriptive statistics for our measure of aggregate firm-specific news over the sample period. The table shows the number of daily news items for each company. Our dataset contains all news items identified by the Dow Jones Intelligent Indexing system as being related to a particular firm. It may also include news where the news agency has specifically marked news as being of importance to the firm. Our sample includes news items from approximately 30,000 different news sources and more than 3 million news items for the 28 companies.

Figure 1 illustrates the aggregate firm-specific news flow for Alcoa (NYSE: AA) across the full cross-section of news sources accessible from the Dow Jones Factiva database. Each company's news flow is composed of news items from a wide group of sources, between 1656 and 4716 individual sources per stock over the sample period.[11] A review of the top 10 sources in terms of aggregate news counts reveals that three key newswires reappear consistently across the sample: Associated Press Newswires, Dow Jones News Service, and Reuters News.[12] The remaining top news sources vary considerably but tend to reflect characteristics of a particular company. The presence of several local news publications reflects the location of a company's corporate headquarters or a large local presence, examples include *SeattlePI.com* for Boeing, *Peoria Journal Star* for Caterpillar, *Charleston Gazette* for Dupont, *The Atlanta Journal* for Coca Cola, *The Cincinnati Post* for Procter & Gamble, and *The Arkansas Democrat Gazette* for Walmart.

Similarly, industry-specific sources are also among the 10 largest contributors to the firm-specific news flow: *Metal Bulletin News Alert Service* for Alcoa; *Moody's Ratings Delivery Service* for various financial institutions (NYSE: AXP, BAC, C, JPM); and corporations with large financing arms (NYSE: CAT, GE); *Health & Medicine Week*, *Biotech Week*, *Pharma Business Week*, and *Drug Week* for healthcare companies such as Pfizer, Johnson & Johnson, and Merck; *Upstream* and *The Oil Daily* for Exxon and Chevron; *TR Daily* and *TR's State Newswire* for telecommunication services companies such as AT&T and Verizon; *FedBizOpps* for companies regularly engaging as contractors with the U.S. Government such as Boeing and United Technologies Corporation; *The Grocer* for Walmart and *Just-Drinks* for Coca Cola.

Table 4 provides insight on general timeliness characteristics of aggregate news items. While we observe a consistent pattern of less news on holidays and weekends as opposed to on trading days, the intensity of aggregate news varies substantially across companies.

In Table 4, we also include a measure of the contribution of the largest news sources to total news flow. The measure is computed as a Gini coefficient using information on the news contribution of each individual source to the total newsflow over the sample period. A Gini coefficient of 0.85 for Alcoa (NYSE: AA) implies that 85% of the firm-specific news flow is contributed by the 15% largest news sources. We can think of the Gini coefficient as a concentration ratio, a high Gini coefficient implies that the firm-specific newsflow is concentrated among relatively fewer sources. As a reference point, a Gini coefficient of 0.50 implies that each source contributes equally to the total newsflow.

Across the sample of companies, the concentration ratio ranges from 80% to 90%. This pattern is partially due to the differences in distribution frequencies across different sources. For example, there are relatively fewer newswires, but they account for a larger proportion

11 Hansen (2012) illustrates how the number of sources changes over time for each stock.

12 The source code identifiers are APRS, DJ, and LBA.

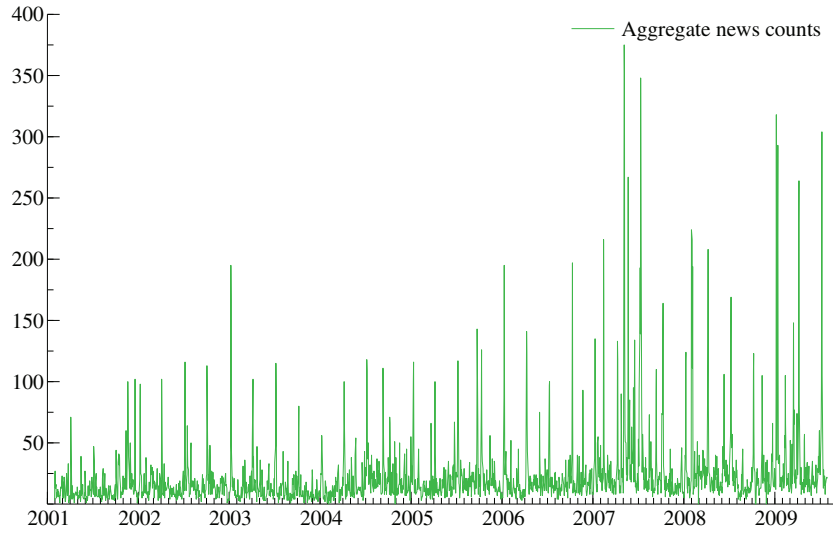

Figure 1 Aggregate news counts for Alcoa (NYSE: AA), January 29, 2001 to July 31, 2009.

Notes: The chart presents the aggregate newsflow of Alcoa Inc in terms of number of news items registered in the Dow Jones Factiva database, using the Dow Jones Intelligent Indexing code for Alcoa.

of the newsflow on an aggregate level. The pattern may be reinforced by the specialization of different sources in terms of their coverage and content focus. For example, financial wires and industry publications are more likely to cover companies on a continuous basis than mainstream media.

The timeliness of news items deserves a final remark before proceeding. We observe a general trend in the number of average daily news items and find that certain months have higher means of daily news items due to quarterly earnings announcements. This has implications for our time series regressions. We include a set of dummy variables and a time trend to account for effects from increasing overdispersion in aggregate news items over time.[13]

4.4 Model Selection Algorithm

With our chosen set of explanatory variables, there are a vast range of possible models we could settle upon, that is, for the estimation of Equation (6) with $K = 5$, which represents up to five lags of the firm-specific variance; $M = 99$ corresponding to possible combinations of 83 economic information indicators and 16 media attention indicators; $J = 3$ which represents inclusion of up to three lags of the corresponding information arrival indicator. To ensure that model selection is free from our subjective judgment, we resort to the general-to-specific model selection procedure as surveyed in Campos, Ericsson, and Hendry (2005) and implemented in Doornik (2009).

13 These variables are rarely significant when news is included in the model. This is consistent with the results in Andersen and Bollerslev (1998), suggesting that day of the week dummies imperfectly substitute for announcement effects such as unanticipated news.

Our use of Doornik's (2009) Autometrics procedure is an effort to keep our model selection procedure as objective as possible. Among alternative machine learning model selection procedures, Autometrics is particularly well-suited for our investigation since according to Castle, Qin, and Reed (2013), it is distinguished by its emphasis on congruency and the use of multi-path searches to select variables based on significance rather than goodness of fit.[14]

It is important to explain the way we conduct estimation of the various sub-models we compare in this article. Whenever we estimate a particular specification Autometrics gets to select the model all over. This implies that the terminal models that we end up with may not be nested, even though the initial models were nested. This explains why the R^2 may decrease when going from a smaller to a larger nesting initial model.

5 Evidence of Information Processing Components of return volatility

This section presents our estimation results and contrasts these with the findings in existing literature. First, we summarize general measures of fit for each estimated model. We provide an overview of the type of economic information included in the estimated time series regressions across stocks in subsection 6.1 and review the role of media attention indicators in subsection 6.2.

Table 5 presents the total as well as partial R^2's from time series regressions used to estimate Equation (6) and expressed in terms of the final form Equation (9), where measures of information are included. We estimate four versions of the model in Equation (6) and then using the estimated parameters, we construct a variable for each of the news components of the final form. For each model, in addition to the total R^2, we present partial R^2 for the variation in the firm-specific variance that is explained with news (R^2_{news}) and the partial R^2 of the contemporaneous term ($R^2_{\text{cont.}}$).

Table 5 provides evidence that, for all 28 stocks, the model identified by our machine learning model selection procedure includes lagged news indicators alongside contemporaneous news indicators. This result suggests that the arrival of unexpected public information triggers information processing activities leading to additional information, previously not publicly available, being incorporated sequentially. A close examination of Table 5 reveals that both the contemporaneous news component and the lagged news component are significant by themselves as well as together, suggesting that the public information and

14 Doornik's (2009) approach, Autometrics, uses a path search through the space of possible models determined by the variables introduced in the initial model specification. The starting point for the model reduction procedure is a general unrestricted model with the full information set. The model has been specified in order to be a statistically well-behaved process. Each insignificant variable in the initial model is a possible reduction path. The first path corresponds to removing the variable with the lowest absolute *t*-value. The process continues in this way until we reach a terminal model. Along the way, each model is subjected to a series of encompassing and diagnostic tests. If model reduction fails, then the current model is considered a terminal model, and the model selection procedure continues. The procedure may arrive at multiple terminal models in which case we use an information criterion to choose the model that best fits the data. We have also approached the estimation using an alternative machine learning model selection procedure: The Least Absolute Shrinkage and Selection Operator implemented using the Least Angle Regression algorithm. This approach reached similar results.

Table 5 The partial R^2 of firm specific indicators of news arrival

Ticker	AR(5)	AR(5) + contemporaneous news			AR(5) + contemporaneous news + lagged effects of news			Contemporaneous news + lagged news	
	R^2_{Total}	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$	R^2_{Total}	$R^2_{\text{Cont.}}$
AA	68	71	16	3	73	30	3	40	8
AXP	84	85	19	1	86	9	1	47	11
BA	66	70	17	4	72	27	4	51	13
BAC	88	89	34	1	90	48	1	66	12
C	89	90	39	1	91	65	1	77	10
CAT	65	68	14	4	71	19	5	31	12
CVX	65	67	10	2	69	17	2	28	6
DD	68	70	10	3	72	20	3	35	9
DIS	73	75	11	1	77	36	1	50	8
GE	81	82	30	2	84	46	2	66	13
HD	70	72	10	2	73	14	2	26	5
HPQ	68	71	14	3	73	21	4	45	12
IBM	73	74	18	2	76	28	2	47	9
INTC	72	73	7	2	74	13	2	43	9
JNJ	70	74	20	3	76	37	3	45	10
JPM	85	86	41	2	87	58	3	70	13
KO	71	73	11	2	74	16	2	33	10
MCD	59	65	14	6	66	16	6	29	12
MMM	62	66	12	4	68	11	4	21	9
MRK	52	55	11	2	56	15	3	27	6
MSFT	71	72	21	2	74	31	2	52	10
PG	65	68	10	3	70	13	3	30	11
PFE	59	63	16	4	66	23	6	44	11
T	65	69	20	4	71	30	4	48	15
VZ	69	72	16	3	73	20	3	37	9
UTX	68	71	10	3	72	14	2	29	11
WMT	64	67	13	3	69	23	3	31	7
XOM	67	68	6	2	71	19	2	33	5

This table reports the R^2 from contemporaneous and lagged news indicators. The model estimated is given in Equation (6):

$$\ln(\text{FV}_t) = \omega + \sum_{k=1}^{K} \rho_k \ln(FV_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t$$

$\text{R}^2_{\text{Total}}$ denotes the R^2 of the model. R^2_{news} denotes the partial R^2 for the variation in the firm-specific variance that is explained with news, and $R^2_{\text{cont.}}$ represents the partial R^2 of the contemporaneous news. AR(5) corresponds to five lags of the log-level realized FV_t. FV_t is estimated using intraday data. In our analysis we consider 83 economic information indicators and 16 media attention indicators (i.e., $M = 99$) and include up to three lags of the corresponding information arrival indicator (i.e., $J = 3$). Sample length generally spans from January 29, 2001 to July 31, 2009. CVX, HPQ, INTC, T, MSFT, start at the following dates: October 10, 2001; April 9, 2011; May 6, 2002; December 1, 2005; April 9, 2001.

effects of private processing of new economic information arrival appear to be equally important in accounting for changes in firm-specific realized variance.

Our finding, that news is related to increases in volatility, is of particular interest since earlier research using measures of news arrival found small or insignificant relationships with return volatility. Roll (1984) examined the variability in orange juice futures prices and found a substantial amount of unexplained volatility. While he found higher volatility on days with *Wall Street Journal* articles covering oranges than on days without, he argued that news was of little importance given the low frequency of such information events.

However, more recently, Boudoukh et al. (2019) suggest that "public information, when appropriately identified, is a much more important source of volatility than previously found." Boudoukh et al. (2019) contrast their results to those of French and Roll (1986) who find "that the major source for return volatility is not public information, but instead private information revealed either by trading or by noise trading." Results from testing our private processing of public information hypothesis of return volatility using a full cross-section of news sources further reinforce the notion that return volatility is related to public information arrival.

The idea that public information arrival generates additional information, previously not publicly available, through private processing of new public information is related to Roll's (1988) finding that the probability of information arrival is higher on days without public news arrival. Roll (1988) describes the prevailing paradigm about changes in asset prices and suggests that through observation and measurement of unanticipated economic information a large proportion of changes in stock prices should be explainable.

Glasserman and Mamaysky (2019), based on empirical results using measures of sentiment and unusualness of news articles contained in the Thomson Reuters database for the top 50 global banks between January 1996 and December 2014, find that unusual news with negative sentiment predicts an increase in volatility while unusual positive news forecasts lower volatility at both the company-specific and aggregate levels.

In this article, by exploiting the full cross-section of news sources, we have sought a measure of firm-specific news that is as collectively exhaustive as possible. As presented in the seventh column (R^2_{news}) of Table 5, we find that on average 26% (9–65%) of changes in firm-specific realized variance are related to information arrival. Considering that 58–77% of realized variance is firm-specific,[15] we argue that our ability to, *ex-post*, relate 9–65% of changes in variance to basic measures of information arrival is of economic importance. Even so, our conclusion is similar to Roll (1988) since we find that public news cannot account for all variation in returns. Interestingly, our study shows that lagged news effects, which were not accounted for in Roll (1988), are significant alongside contemporaneous effects. To summarize, our results favor the paradigm that changes in volatility can be related to the arrival of unanticipated economic information.

Across stocks, we find that on average 26% (9–65%) of changes in firm-specific variance are related to the arrival of unexpected firm-specific information. Our finding contributes to the discussion raised by French and Roll (1986) and supports the conclusions of Boudoukh et al. (2019).[16] French and Roll (1986) use the distinction between trading and

15 Refer to Column 8 of Table 3.

16 Boudoukh et al. (2019) report that fundamental information in news accounts for 12.4% of idiosyncratic volatility during trading hours versus 49.6% during overnight periods.

nontrading periods as a proxy for the rate of information arrival in order to study information processing in financial markets. French and Roll (1986) suggest three hypotheses for patterns in stock return variances: public information is more likely to arrive during trading hours; private information is incorporated into prices only through trading by informed investors during trading hours; and pricing errors occur during trading hours. They conclude that variation in the information flow is the most likely determinant of volatility, and that private information is likely to be the largest component. We measure the rate of information arrival directly and our results suggest that the relationship between public and private information is more complex and, based on the relative size of the partial R^2 in column seven (R^2_{news}) of Table 5 for regressions including indicators of information arrival, that public information and related processing of public information are of economic importance for stock return variances.

Since on average 26% (9–65%) of changes in firm-specific return volatility can be explained by such a basic measure of information arrival, it is likely that news is an important missing piece in Shiller's (1981) volatility puzzle, which states that stock price volatility is too high to be accounted for by new information about the economic performance of the firm.

The results in Table 5 also have implications for a large body of literature exploring different methods to estimate the underlying information flow of financial assets by using observed returns and TV. Past studies have proposed models that explain characteristics of the return generating process best described as autoregressive properties in the variance of returns. The MDH is one theoretical explanation for this phenomenon. The MDH suggests that a serially correlated mixing variable, measured as the rate of information arrival, causes the autoregressive properties of the variance of returns of financial assets. Maheu and McCurdy (2004) put forth a version of the MDH where the return generating process is assumed to be directed by a latent news process. The conditional variance of returns is specified to have a smoothly evolving component, related to the diffusion of past news arrival, and a component related to the contemporaneous information arrival process that generates jumps. Our empirical approach can be considered a direct test of the MDH, where measures of new economic information are related to changes in the rate of price informativeness. Again, our results support our proposed specification of the return generating process and in particular mixture models with lagged effects of information arrival such as the diffusion of past information arrival in Maheu and McCurdy (2004).

Jeon, McCurdy, and Zhao (2019) find that news flow measures explain 24% of the jump size volatility of top 20 firms in their sample. Our finding, using a sample of 28 large U.S. companies included in the Dow Jones Industrial Average, that on average 26% of changes in firm specific realized variance are related to information arrival, further supports the economic importance of the impact of information contained in firm-specific news on volatility as well as on return jump intensities.

5.1 Categories of Economic Information

We now turn to characterizing the firm-specific news that we find is related to changes in stock return volatility. Table 6 shows that news items categorized as covering *Management Issues* are most commonly found in regressions explaining changes in firm-specific realized variance. In time series regressions that only include contemporaneous news series, our

Table 6 Parameter overview—Corporate and industrial indicators

Indicator	Number of stocks for which each information arrival indicator is included the model					Indicator	Number of stocks for which each information arrival indicator is included the model				
	News only	News and lagged news					News only	News and lagged news			
	γ_0	γ_0	γ_1	γ_2	γ_3		γ_0	γ_0	γ_1	γ_2	γ_3
Management issues	11	9	3	1	5	Inflation/prices	2	1	2	1	1
Earnings surprises	9	9	4	6	3	Consumer sentiment	2	1	6	4	0
Money markets	7	0	2	4	0	Factory orders/durable goods	2	2	2	2	1
Financing agreements	7	4	2	2	1	Employment costs/productivity	2	2	0	1	1
Acquisitions/mergers/takeovers	7	7	2	1	2	Mortgage applications/refinancing	2	3	2	1	2
Economic growth	6	3	1	0	4	Index of leading economic indicators	2	1	2	1	1
Corporate crime/legal/judicial	6	6	1	1	4	Euro zone/currency	2	2	2	2	1
Dividends	6	2	3	1	1	Plans/strategy	2	1	5	3	4
Analyst comment/recommendation	6	5	0	1	0	Annual meetings	2	3	5	1	1
Nongovernment Contracts	6	7	1	1	2	Internal audit	2	2	1	1	0
Government finance	5	3	3	3	0	Corporate debt instruments	2	2	1	2	1
Divestitures/asset sales	5	2	2	0	0	New products/services	2	6	5	3	1
Competition issues	5	2	1	1	1	Capacity/facilities	2	1	2	5	3
Lay-offs/redundancies	5	4	0	6	3	Product safety	2	3	1	0	0
International relations	4	2	3	1	0	Marketing/market research	2	3	3	4	2
Domestic politics	4	3	0	4	2	Franchises	2	2	3	1	0
Labor issues	4	4	2	2	0	Environmental news	1	0	1	2	4
Foreign exchange news	4	4	3	1	3	Global/world issues	1	3	3	6	0
Soft commodity markets	4	3	2	2	0	Health	1	2	2	1	2
Energy markets	4	3	1	1	5	Weather	1	2	2	2	2
Fund markets	4	3	4	5	2	Demographics	1	1	0	1	1

(continued)

Table 6 Continued

Indicator	Number of stocks for which each information arrival indicator is included the model					Indicator	Number of stocks for which each information arrival indicator is included the model				
	News only	News and lagged news					News only	News and lagged news			
	γ_0	γ_0	γ_1	γ_2	γ_3		γ_0	γ_0	γ_1	γ_2	γ_3
Bankruptcy	4	3	2	2	0	Money supply	1	1	0	0	1
Privatizations/nationalization	4	2	2	3	2	Personal income/average earnings	1	1	0	3	3
Joint ventures	4	0	3	5	3	Industrial production	1	3	2	1	4
Output/production	4	4	2	2	1	Capacity utilization	1	1	0	1	0
Climate change	3	2	2	3	5	Business sentiment	1	1	4	2	3
Budget account	3	2	0	2	3	Bankruptcy figures	1	1	1	0	0
Employment/unemployment	3	3	0	2	2	Housing starts/construction figures	1	2	0	2	0
Reserve assets	3	5	0	0	1	Trade/external payments	1	0	2	3	4
Economic/monetary Policy	3	3	1	3	0	Metals markets	1	1	4	2	2
Sales figures	3	4	6	4	4	Research/development	1	1	3	3	1
Share capital	3	4	6	1	2	Information technology	1	2	1	1	2
Corporate credit ratings	3	4	2	2	0	Government contracts	1	1	2	1	1
Licensing agreements	3	1	0	2	2	Defense contracts	1	2	0	1	2
Outsourcing	3	1	2	3	2	Natural reserves/resources discovery	1	2	0	1	2
Crime/courts	2	3	1	2	0	Consumer credit/expenditure/savings	0	1	1	2	1
Disasters/accidents	2	3	2	1	1	Inventories	0	0	2	1	1
Regional politics	2	1	2	1	1	Regulation/government policy	0	1	2	3	1
Science/technology	2	2	4	1	2	Labor disputes	0	0	4	2	2

This table summarizes which indicators of economic information arrival generally are included in the time series regressions in the cross section of our sample of 28 large U.S. stocks. The *Number* column presents the number of stocks for which the respective time series of a given information arrival indicator was included in a model, given in Equation (6), describing the changes in daily FV_t of a given stock. In our analysis we consider 83 economic information indicators and 16 media attention indicators (i.e., $M = 99$) and include up to three lags of the corresponding information arrival indicator (i.e., $J = 3$). FV_t is estimated using intraday data. Sample length differs from stock to stock but generally spans from January 29, 2001 to July 31, 2009. CVX, HPQ, INTC, MSFT, T start at the following dates respectively: October 10, 2001; April 9, 2011; May 6, 2002; December 1, 2005; April 9, 2001.

model selection procedure includes it 11 times, while in regressions where we also consider lagged information indicators, it is included 9 times across our sample of 28 stocks.

Table 6 also shows that *Earnings Surprise* is the second most common information arrival indicator across stocks. This series includes earnings announcements deemed by news agencies and the Dow Jones Intelligent Indexing system to convey material change in performance. Newsflow in this category often contains headlines mentioning earnings or profitability measures and such words as *decline*, *disappoint*, *drop*, *loss*, *plunge*, *jump*, *strong*, *beat*, and *boost*. American Express (NYSE: AXP) is one of the companies where *Earnings Surprise* newsflow is included. The news items often refer specifically to changes in the performance of individual business units and end-markets. For example, trends in card issuance and delinquencies associated with American Express's credit card business. Moreover, it is not uncommon to find news items discussing earnings surprises for competitors and companies in the same lines of business.

The results found in Table 6 are in line with Patell and Wolfson (1984) since we find that *Earnings Surprise* is the information arrival indicator most likely to have a lagged relationship with firm specific realized variance. This can be observed by examining the number of lagged indicators for the *Earnings Surprise* category that are used in stock specific regressions. We see that the contemporaneous *Earnings Surprise* indicator is included for nine stocks, while between three and six of the stock specific regressions include lagged effects of this indicator. This pattern suggests that the release of *Earnings Surprise* information is more likely to provide information-based investors with an advantage from information processing activities, since there are information-related effects up to three days after a shock to the *Earnings Surprise* newsflow.

An interesting result in Table 6 is that no category, other than *Management Issues and Earnings Surprises*, is included across more than one third of our 28 stocks. An examination of the underlying newsflow significant for each stock also reveals material differences in the type of issues surfacing within any single newsflow category. For example, we examine the *Corporate Crime/Legal/Judicial* category which is significant for both IBM (NYSE: IBM) and Citigroup (NYSE: C). For IBM, an information technology products and services provider, themes in the newsflow include a series of lawsuits including a lawsuit by Compuware for misappropriation of source code used in mainframe testing, a civil lawsuit by the SCO Group related to licensing rights to the UNIX computer operating system, as well as investigations into IBM's accounting practices by the SEC. For Citigroup, a financial services provider, the same newsflow category covers an investigation by the German financial services regulator into Citigroup's alleged manipulation of European bond markets as well as litigation with respect to underwriting activities for WorldCom, Enron, and Parmalat. These examples illustrate the diversity of the events covered in the news for different companies.

Many of the corporate information events covered by our indicators have previously been studied. In fact, the idea that new economic information drives changes in asset prices is supported by a large body of literature using event study methodologies to explore how different corporate events are related to changes in stock prices. Neuhierl et al. (2013) report significant stock price reaction to news contained in corporate press releases classified as related to corporate strategy, customers and partners, products and services, management changes, and legal developments. A common concern in event studies is the ability to control for confounding events. On average, the stocks in our sample have between 12 and

195 news items each per trading day.[17] In fact, it is seldom that these stocks a have a day where they are not mentioned in the media. Despite the presence of multiple news items, our model selection procedure still identifies certain news subjects as significant. Results in Table 6 confirm the importance of several corporate events studied in prior research while controlling for the vast number of simultaneously occurring news items available for a given company.

5.2 Media Attention

Table 7, like Table 6, presents an overview of the media attention indicators found significant across our sample of stocks. The indicators in this table have all been included in order to simultaneously control for the media attention hypothesis put forth by studies such as Huberman and Regev (2001), Barber and Odean (2008), and Engelberg and Parsons (2011).

Results presented in Table 7 indicate that newswire media attention is important when explaining changes in firm specific realized variance. This can be observed by examining the number of lagged indicators for the *Newswires* category that are used in stock-specific regressions. We see that the contemporaneous *Newswires* indicator is included for 22 stocks, while 9 to 10 of the stock specific regressions include lagged effects of this indicator. Our results appear to be at odds with the results of Engelberg and Parsons (2011). Engelberg and Parsons (2011) compare the behavior of investors with access to different media coverage of the same information event. They find that local media coverage predicts local trading while controlling for the characteristics of earnings announcements, investors, and individual newspapers. They suggest that their results would also apply to large media outlets, such as newspapers with substantial circulation, and that the media attention relationship is more likely to drive trade than unanticipated economic information.

We test the information processing hypothesis while controlling for the media effects. From Tables 6 and 7, we conclude that while several economic information categories are relevant for explaining changes in firm-specific variance, the media attention hypothesis cannot be ruled out. However, the main channel for the media attention effect in our study appears to be the aggregate attention from newswires as opposed to that from leading newspapers with large circulation, such as *USA Today*. We argue that the newspaper media attention effect is more plausible when thought of as a channel for the resolution of asymmetric information between informed and noninformed traders as modeled in Foster and Viswanathan (1990) and as examined in Tetlock (2010). On the other hand, the significance of the newswire media attention indicator may arise from the indicator's ability to proxy for the number of informed traders. For example, Holden and Subrahmanyam (1992) show that price change variance increases with the number of informed traders.

Table 7 confirms the evidence in Tetlock (2010) that the number of newswire items has predictive power for news day price changes. Tetlock (2010) uses 2.2 million articles from the Wall Street Journal and the Dow Jones News Service to examine if public news eliminates information asymmetry from two types of traders for a sample of 13,842 firms from 1979 to 2007. He suggests that public news levels the playing field for other investors by resolving asymmetric information. In addition, he finds that the number of newswire

17 Derived from the ratio of fourth column in Table 4 to last column in Table 3.

Table 7 Parameter overview—media attention indicators

Indicator	Number of stocks for which each information arrival indicator is included the model				
	News only	News and lagged news			
	γ_0	γ_0	γ_1	γ_2	γ_3
Newswires	18	22	9	10	9
Dow Jones newswires	16	13	5	2	2
Reuters newswires	16	11	4	3	2
Newspapers	13	3	2	3	6
Associated press	9	5	4	2	1
Industry publications (DJ)	6	2	8	1	0
Press release newswires	6	5	4	5	2
Wall Street Journal	5	4	1	2	1
Newsletters (DJ)	4	8	3	2	2
Washington Post	4	1	5	1	1
Government and Politics publications (DJ)	4	2	1	5	0
New York Times	3	1	2	2	2
Financial Times	3	2	1	1	4
Press release newswire	3	2	2	3	0
Major business news (DJ)	2	4	4	1	0
Other sources	2	4	4	1	4

This table summarizes which indicators of economic information arrival generally are included in the time series regressions in the cross-section of our sample of 28 large U.S. stocks. The *Number* column presents the number of stocks for which the respective time series of a given information arrival indicator was included in a model, given in Equation (6), describing the changes in daily FV_t of a given stock. In our analysis we consider 83 economic information indicators and 16 media attention indicators (i.e., $M = 99$) and include up to three lags of the corresponding information arrival indicator (i.e., $J = 3$). FV_t is estimated using intraday data. Sample length differs from stock to stock but generally spans from January 29, 2001 to July 31, 2009. CVX, HPQ, INTC, MSFT, T start at the following dates respectively: October 10, 2001; April 9, 2011; May 6, 2002; December 1, 2005; April 9, 2001. DJ indicates that source is the Dow Jones & Company.

messages subsumes the predictive power of news day returns but finds a small positive relationship between contemporaneous public information and absolute returns.

The results in Tables 6 and 7 differ in one key way from Tetlock (2010). We find a significant relationship between contemporaneous information arrival and changes in realized variance. We think that Tetlock's (2010) finding that public information cannot account for news day returns may be due to the omission of sources other than the Dow Jones newswires and the *Wall Street Journal*, in order to measure content materiality better. For example, as presented in Table 7, measures of increases in aggregate newswire attention are included in the time series regressions of 22 out of 28 companies. Moreover, we find that the main Dow Jones and Reuters news wires are often included, 13 and 11 times, respectively. In addition, the *Wall Street Journal* is included between four and five times across stocks. This may reflect the fact that it is focused on in-depth analysis of corporate information events originally disclosed by Dow Jones, the *Wall Street Journal*'s corporate parent,

family of newswires. In other words, we find support for the resolution of information asymmetry due to arrival of public information (Foster and Viswanathan, 1990, Tetlock, 2010) as a result of inclusion of several media attention indicators for leading newspapers; however, we suggest that, by exploiting the full cross-section of news sources, our empirical approach more accurately measures firm-specific news arrival.

Our view, that lagged information effects are related to the processing of public information, is further supported by a study by Engelberg, Reed, and Ringgenberg (2012). Engelberg et al. (2012) examine the trading pattern of short sellers when news from the Dow Jones newswires and the *Wall Street Journal* arrives. Using a sample covering 2005–2007, they find a significant increase in short selling after news events and that the most informative trades appear to be on days following news arrival. In addition, the most profitable trades are made on days where trades arrive later than those of other investors. They suggest that informed traders do not predict information arrival but rather gain their information advantage from processing publicly available information. They conclude that public news arrival creates trading opportunities for skilled information processors.

6 Robustness Checks

In this section, we perform a series of robustness tests. First, we check the robustness of our results to the alternative dependent variables, which are related to information arrival. Second, we investigate the impact of mixing the firm specific newsflow across the 28 stocks. Finally, we check the sensitivity of our results to data period used by splitting the sample into nonoverlapping periods of three sub-samples.

6.1 Alternative-dependent Variables

First, we use dependent variables that differ from realized FV: TV, squared returns (SQs), and IV from 1 month at-the-money call options (IV). When using TV, squared close-to-close returns, and IV, we aim to check the relationship of news with other measures, linked to the firm's common stock that we expect are related to the arrival of new information. Several empirical studies have shown that TV, SQ, and IV are related to information arrival.

Berry and Howe (1994) and Mitchell and Mulherin (1994) use aggregated news measures and document a positive relationship between broad market activity and the number of news items from the Reuters and Dow Jones news wires, respectively. Tetlock (2010) finds that the contemporaneous cross-correlation between absolute returns and abnormal TV is temporarily higher by 3.5% on days with Dow Jones or Wall Street Journal news items. At high-frequency, Groß-Klußmann and Hautsch (2011) find that volatility and TV are most sensitive to news arrival. With the media attention hypothesis in mind, Engelberg and Parsons (2011) are able to explain roughly 30% of the variation in log-transformed TV by using proxies for local media attention and indicators of earnings announcements. These studies suggest that reasonable proxies for information arrival and media attention should also be related to excess TV and excess volatility in close-to-close returns. Table 8 provides evidence that our indicators of information arrival and changes in media attention are able to explain significant amounts of changes in these alternative measures.

TV and SQ are directly related to realized variance through the trading process, this is not the case for IV from options on a given common stock. To add further flavor to our

Table 8 Estimation results with alternative dependent variables

Ticker	TV			SQ			IV		
	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$
AA	86	34	3	24	9	4	98	5	0.2
AXP	79	23	2	40	8	4	99	3	0.1
BA	62	26	8	20	8	4	98	3	0.1
BAC	96	64	1	45	16	4	99	11	0.1
C	94	58	2	45	14	4	99	10	0.1
CAT	84	29	5	25	10	6	98	3	0.2
CVX	89	29	1	17	7	2	96	16	0.3
DD	69	21	3	25	5	1	98	1	0.0
DIS	66	28	7	26	6	3	98	6	0.1
GE	86	49	5	34	10	3	99	3	0.1
HD	76	19	5	26	10	5	98	3	0.2
HPQ	61	37	17	23	8	5	96	18	0.3
IBM	62	24	8	26	7	3	98	11	0.3
INTC	51	34	20	26	9	4	97	14	0.3
JNJ	68	29	8	17	8	3	97	15	0.2
JPM	89	50	3	42	11	5	99	2	0.1
KO	70	25	4	20	7	4	98	5	0.1
MCD	62	26	11	13	6	5	96	13	0.2
MMM	69	27	8	23	5	3	97	2	0.2
MRK	68	24	2	19	8	3	96	4	0.1
MSFT	52	31	7	28	11	5	98	5	0.1
PG	85	34	3	13	4	2	97	3	0.1
PFE	83	50	6	21	10	5	96	11	0.4
T	65	15	6	23	7	2	98	3	0.2
VZ	73	39	5	23	6	2	98	5	0.1
UTX	75	37	3	24	6	2	98	2	0.1
WMT	75	36	6	18	8	3	97	4	0.1
XOM	83	43	1	19	9	4	97	19	0.2

Notes: TV and squared close-to-close returns (SQ) are obtained from the CRSP database, while IV is obtained from Bloomberg.

This table presents R^2s from estimating Equation (6) using alternative dependent variables: TV, squared close-to-close returns (SQ), and IV calculated using 1-month at-the-money call options (IV):

$$\ln(\text{TV}_t) = \omega + \alpha \ln(\text{SQ}_{\text{SP500},t}) + \sum_{k=1}^{K} \rho_k \ln(\text{TV}_{\text{t-k}}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t$$

$$\ln(\text{SQ}_t) = \omega + \alpha \ln(\text{SQ}_{\text{SP500},t}) + \sum_{k=1}^{K} \rho_k \ln(\text{SQ}_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t$$

$$\ln(\text{IV}_t) = \omega + \alpha \ln(\text{IV}_{\text{SP500},t}) + \sum_{k=1}^{K} \rho_k \ln(\text{IV}_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t$$

where $n_{i,\,t}$ are indicators of new information arrival. R^2_{Total} denotes the R^2 of the model. R^2_{news} denotes the partial R^2 for the variation in the firm-specific variance that is explained with news, and $R^2_{\text{cont.}}$ represents the partial R^2 of the contemporaneous news. Sample length generally spans from January 29, 2001 to July 31, 2009.

robustness check, we examine the relationship between information arrival and changes in IV. Beyond the obvious connection, previous research suggests that the options market reflects the characteristics of the realized volatility process. Using the news tone scores for firm-specific news announcements contained in the Thomson Reuters News Analytics database, between January 2003 and December 2011, Dzielinski and Hasseltoft (2017) find a significant and positive relation between news-tone dispersion and IV. Glasserman and Mamaysky (2019), using news articles in the Thomson Reuters database for the top 50 global banks between January 1996 and December 2014, show that measures of sentiment and unusualness of news stories are able to predict IV.[18] In other words, prior empirical research suggests that IV should be related to measures of firm-specific news.

Table 8 confirms the evidence found in prior research suggesting that TV and IV are related to the arrival of firm-specific information. We argue that this robustness check illustrates the strength of the evidence in Table 5 and supports our proposition that positive changes in newsflow are indicators of the arrival of relevant information.

6.2 Newsflow Mix

Second, we proceed in our robustness checks by using the original-dependent variable, realized firm-specific variance and mix the newsflow across our sample of stocks. In other words, we use firm-specific newsflow for companies other than the company whose firm-specific realized variance we are using as the dependent variable. Figure 2 shows that for the stocks in our sample, using the correctly matched newsflow mostly yield higher R^2's than using the incorrect newsflow, that is, news related to other companies. This is in particular the case for the stock where the newsflow has most explanatory power. These results suggest that our measures of information arrival and media attention are firm-specific. Still, we see for some stocks that new from other stocks have notable explanatory power. This is to be expected as some of the stock belongs to the same sector. Hence, we should, for example, except some news about one of the pharmaceutical companies to affect the stock prices of the other pharmaceutical companies.

6.3 Sub-Sample Analysis

Our third robustness check consists of splitting our sample into different periods and subsequently investigating the sensitivity of our results to this approach. We split our sample into three nonoverlapping sub-samples corresponding to the ranges 2001–2003, 2004–2006, and 2007–2009. Comparing results in Table 9 to those in Table 5 suggests that in most cases the R^2 measures are somewhat higher across sub-samples. While we observe that R^2_{news}, that is, how much of the variation in the firm-specific variance that is explained with news, during the first, second and third 3-year sub-samples are higher for 17, 17, and 23 stocks, respectively, compared to the entire sample period, on average partial R^2's consistently increase over each sub-sample. Except for four stocks in the 2007–2009 sub-sample,

18 Calomiris and Mamaysky (2019) also utilize the TRNA database in their investigation of the context of news articles to predict risk and return in 51 developed and emerging stock markets between 1996 and 2015. They find that the information extracted from news text flow, using measures such as sentiment, frequency, and unusualness, forecasts 1-year ahead returns and risk, measured by rolling 20-day realized volatility as well as 12-month drawdowns.

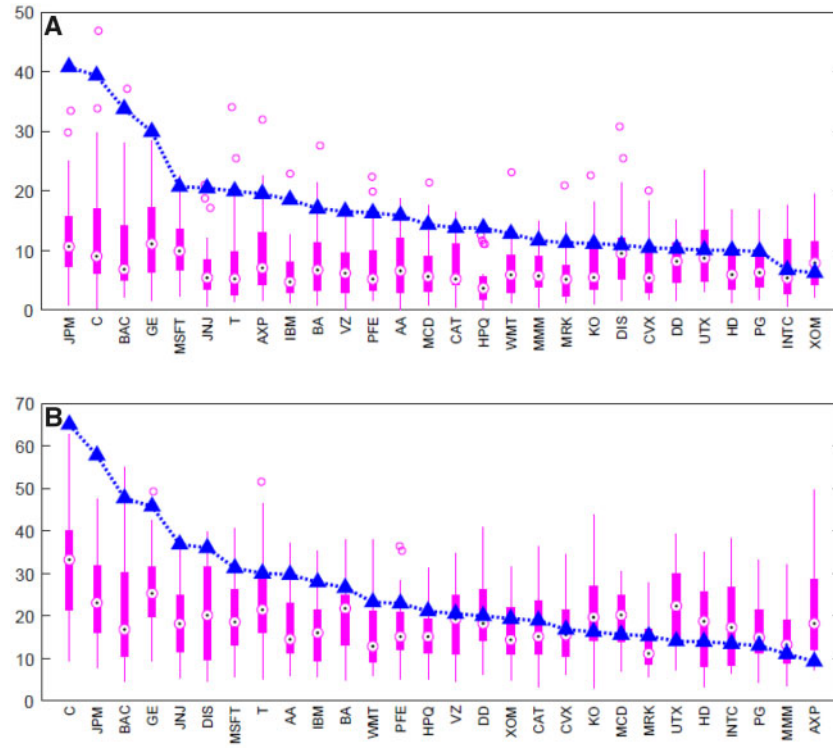

Figure 2 R^2's when mixing newsflow across stocks.

Notes: Blue triangles indicate the R^2 from models where contemporaneous news (top plot) and contemporaneous and lagged news (bottom plot) is included from a model where a company's Firm Specific newsflow is used to explain changes in daily FV_t. The box plot for each stock ticker corresponds to the distribution of R^2 for models using the newsflow related to other companies. Each boxplot consists of 27 model estimations using newsflow for each of the other companies in the sample. AT&T (NYSE: T) is excluded due to its substantially smaller sample.

we find that all partial R^2's for models including the contemporaneous and lagged news (R^2_{news}) are higher.

Results from sub-sample estimations highlight the time-varying relationship between volatility and firm-specific news arrival. Observed increases in partial R^2's over each subsample could be due to enhancements in technology employed by the news media industry over time.

These results could be due to a true data generating process where new information has a time-varying relationship with volatility, for example, the impact of similar corporate events, for example, a new product launch, may be different depending on the state of the macroeconomic environment.

Our investigation also highlights the challenge inherent in correctly measuring economic information arrival once all relevant news items are observed. Boudoukh et al. (2019) also discuss the role of appropriate identification of public information. Improving corporate

Table 9 Results from sub-sample estimation

Ticker	2001–2003			2004–2006			2007–2009		
	AR(5) + contemporaneous news + lagged effects of news			AR(5) + contemporaneous news + lagged effects of news			AR(5) + contemporaneous news + lagged effects of news		
	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$	R^2_{Total}	R^2_{News}	$R^2_{\text{Cont.}}$
AA	58	14	5	47	26	7	75	19	4
AXP	78	31	2	45	23	14	85	10	1
BA	63	27	7	40	21	15	73	37	5
BAC	79	24	1	42	24	6	91	37	2
C	80	50	3	38	15	7	92	74	3
CAT	65	26	4	46	28	11	80	30	4
CVX	76	33	4	64	30	3	70	26	1
DD	72	27	3	37	21	9	73	26	5
DIS	69	34	4	44	29	5	79	27	3
GE	69	27	3	38	19	3	88	50	3
HD	69	17	3	39	21	11	79	30	3
HPQ	83	79	5	45	28	15	78	40	6
IBM	75	36	1	47	33	13	77	52	4
INTC	74	11	2	40	22	7	74	35	5
JNJ	58	19	6	56	38	14	76	31	5
JPM	78	60	4	43	26	11	90	69	4
KO	68	28	6	48	30	11	74	20	2
MCD	53	30	15	41	25	14	72	16	6
MMM	73	16	3	40	29	18	75	36	6
MRK	53	23	2	31	14	2	73	37	9
MSFT	65	16	3	39	23	6	78	50	5
PG	74	20	2	46	34	19	77	25	4
PFE	60	22	4	43	32	20	81	43	5
VZ	66	24	2	41	20	11	75	25	4
UTX	65	21	2	42	21	10	77	19	5
WMT	74	30	3	37	26	10	71	32	6
XOM	72	16	1	60	27	4	75	37	3

This table presents R^2s from models using nonoverlapping sub-samples, in estimating Equation (6), describing the changes in daily FV_t of a given stock. In our analysis, we consider 83 economic information indicators and 16 media attention indicators (i.e., $M = 99$) and include up to three lags of the corresponding information arrival indicator (i.e., $J = 3$). AR(5) corresponds to five lags of the log-level realized FV_t. FV_t is estimated using intraday data. Sample length differs from stock to stock but generally spans from January 29, 2001 to July 31, 2009. CVX, HPQ, INTC, and MSFT start at the following dates, respectively: October 10, 2001; April 9, 2011; May 6, 2002; April 9, 2001. R^2_{Total} denotes the R^2 of the model. R^2_{news} denotes the partial R^2 for the variation in the firm-specific variance that is explained with news, and $R^2_{\text{cont.}}$ represents the partial R^2 of the contemporaneous news.

event and information taxonomies and documenting characteristics of corporate newsflow are clear steps forward in resolving such measurement issues.

Finally, while we have attempted to gather as collectively exhaustive a dataset of firm specific newsflow as possible, the database that we use, while extensive, does not include

all news sources. For example, any news published by Bloomberg's news agency is not completely included in our dataset. Bloomberg news items are only represented when other publications or newswires import content created by Bloomberg's news agency. Like the Dow Jones Factiva database, Bloomberg provides access to a wide range of news sources. The extent and impact of missing news sources had been examined and reveals that there are significant overlaps in the stories reported by leading news agencies. On a more general note, both Bloomberg and Factiva report that they provide access to more than 20,000 sources, suggesting a high degree of overlap between the two databases. As a final remark, we highlight our focus on news items as the source for measuring economic information arrival. For example, filings with the SEC and other corporate disclosure suggest that industry supply and demand for aluminum should affect the economic performance of Alcoa (NYSE: AA). We have not included such sources of economic information unless they are specifically covered by news articles within Factiva database. However, it should be noted that the firm-specific newsflow contains news items describing events related to aluminum and other commodities. Our results control for firm-specific newsflow covering *General & Political*, *Macroeconomic*, and *Financial Market* subject categories to the extent news agencies and the Dow Jones Intelligent Indexing system mark such information as being relevant to the specific company. We believe that any omitted sources of economic information arrival will be independent of the information identified and measured in this investigation.

7 Conclusion

In this article, we develop the information processing hypothesis of return volatility to investigate the relation between firm-specific news and volatility. We propose a novel dynamic econometric specification and test it using time series regressions employing a machine learning model selection procedure. Our primary contribution originates from the unique hypothesis we develop to directly test the impact of firm specific news on firm specific volatility for which we present economically significant results.

We identify a collectively exhaustive measure of firm-specific newsflow by gathering all firm-specific news in the cross-section of approximately 30,000 different news sources for a sample of 28 large U.S. companies from more than 3 million news items labeled with 83 subject categories.

Our proposed econometric specification for firm-specific return volatility is a simple mixture model with two components: public information and private processing of public information. The public information processing component is defined by the contemporaneous relation with public information and volatility, while the private processing of public information component is specified as a general autoregressive process corresponding to the sequential price discovery mechanism of investors as additional information, previously not publicly available, is generated and incorporated into prices.

Our results show that changes in return volatility are related to public information arrival. For all 28 stocks in our sample, adding contemporaneous and lagged firm specific news explains a significant proportion of changes in firm-specific return volatility. Firm-specific volatility generally accounts for 58% to 77% of all variation in 5-min returns within the trading day. Including contemporaneous and lagged indicators explains on average 26% (between 9% and 65%) of changes in firm-specific volatility.

Robustness checks confirm that our measures of information arrival are indeed firm-specific and capture relevant information related to the firm. Furthermore, robustness checks underscore the time-varying relationship between volatility and firm-specific news arrival.

Although early finance literature found little evidence for the impact of public information on volatility, results from testing our information processing hypothesis of return volatility further reinforce the notion that return volatility is related to public information arrival, which is also supported by few other recent investigations, as news arrival triggers some form of information processing. Furthermore, combined with the findings of Jeon, McCurdy, and Zhao (2019) who focus on news and jump diffusion, our information processing hypothesis of return volatility could enrich certain option pricing models, such as the one developed by Christoffersen, Jacobs, and Ornthanalai (2012) who incorporate both dynamic volatility and dynamic jump intensities.

References

Admati, A., and P. Pfleiderer. 1988. A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies* 1: 3–40.

Andersen, T. G. 1996. Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility. *The Journal of Finance* 51: 169–204.

Andersen, T. G., and T. Bollerslev. 1998. Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *The Journal of Finance* 53: 219–265.

Andersen, T. G., T. Bollerslev, F. Diebold, and H. Ebens. 2001. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics* 61: 43–76.

Antoniou, C., J. A. Doukas, and A. Subrahmanyam. 2013. Cognitive Dissonance, Sentiment, and Momentum. *Journal of Financial and Quantitative Analysis* 48: 245–275.

Bali, T. G., and R. F. Engle. 2010. The Intertemporal Capital Asset Pricing Model with Dynamic Conditional Correlations. *Journal of Monetary Economics* 57: 377–390.

Barber, B., and T. Odean. 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies* 21: 785–818.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2009. Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal* 12: C1–C32.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2011. Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. *Journal of Econometrics* 162: 149–169.

Berry, T. D., and K. M. Howe. 1994. Public Information Arrival. *The Journal of Finance* 49: 1331–1346.

Black, F. 1976. "Studies of Stock Price Volatility Changes." *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, American Statistical Association, Washington DC, 177–181.

Black, F. 1986. Noise. The *Journal of Finance* 41: 529–543.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson. 2019. Information, Trading, and Volatility: Evidence from Firm-Specific News. *Review of Financial Studies* 32: 992–1033.

Boudoukh, J., M. Richardson, Y. Shen, and R. Whitelaw. 2007. Do Asset Prices Reflect Fundamentals? Freshly Squeezed Evidence from the OJ Market. *Journal of Financial Economics* 83: 397–412.

Brogaard, J., T., Hendershott, and R., Riordan. 2014. High Frequency Trading and Price Discovery. *Review of Financial Studies* 27: 2267–2306.

Calomiris, C. W., and H. Mamaysky. 2019. How News and Its Context Drive Risk and Returns around the World. *Journal of Financial Economics* 133: 299–336.

Campos, J., N. R. Ericsson, and D. F. Hendry. 2005. General-to-Specific Modeling: An Overview and Selected Bibliography. *International Finance Discussion Paper.*

Castle, J. L., X. Qin, and W. R. Reed. 2013. Using Model Selection Algorithms to Obtain Reliable Coefficient Estimates. *Journal of Economic Surveys* 27: 269–296.

Christie, A. 1982. The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects. *Journal of Financial Economics* 10: 407–432.

Christoffersen, P., K. Jacobs, and C. Ornthanalai. 2012. Dynamic Jump Intensities and Risk Premiums: Evidence from S&P500 Returns and Options. *Journal of Financial Economics* 106: 447–472.

Chua, C. L., and S. Tsiaplias. 2019. Information Flows and Stock Market Volatility. *Journal of Applied Econometrics* 34: 129–148.

Clark, P. K. 1973. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41: 135–156.

DeGennaro, R., and R. Shrieves. 1997. Public Information Releases, Private Information Arrival and Volatility in the Foreign Exchange Market. *Journal of Empirical Finance* 4: 295–315.

Doornik, J.2009. Autometrics. *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry, Castle, J. L. and Shephard, N., editors*. Chapter 4: 88–122. Oxford UK: Oxford Scholarship Online Monographs, Oxford University Press.

Dougal, C., J. Engelberg, D. Garcia, and C. A. Parsons. 2012. Journalists and the Stock Market. *Review of Financial Studies* 25: 639–679.

Dzielinski, M., and H. Hasseltoft. 2017. "News Tone Dispersion and Investor Disagreement." *Unpublished Working Paper*. Available at https://ssrn.com/abstract=2192532. (last accessed October 30, 2020)

Ederington, L. H., and J. H. Lee. 1993. How Markets Process Information: News Releases and Volatility. *The Journal of Finance* 48: 1161–1191.

Engelberg, J. 2008. 'Costly information processing: Evidence from earnings announcements." *Unpublished Working Paper, AFA 2009 Meetings Paper*. Available at SSRN: https://ssrn.com/abstract=1107998. (last accessed October 30, 2020)

Engelberg, J., R. D. Mclean, and J. Pontiff. 2018. Anomalies and News. *The Journal of Finance* 73: 1971–2001.

Engelberg, J., and C. Parsons. 2011. The Causal Impact of Media in Financial Markets. *The Journal of Finance* 66: 67–97.

Engelberg, J., A. Reed, and M. Ringgenberg. 2012. How Are Shorts Informed? Short Sellers, News, and Information Processing. *Journal of Financial Economics* 105: 260–278.

Fang, L., and J. Peress. 2009. Media Coverage and the Cross-Section of Stock Returns. *The Journal of Finance* 64: 2023–2052.

Foster, F., and S. Viswanathan. 1990. A Theory of the Interday Variations in Volume, Variance, and Trading Costs in Securities Markets. *Review of Financial Studies* 3: 593–624.

Foucault, T., J., Hombert, and I., Roşu. 2016. News Trading and Speed. The *Journal of Finance* 71: 335–382.

French, K., and R. Roll. 1986. Stock Return Variances: The Arrival of Information and the Reaction of Traders. *Journal of Financial Economics* 17: 5–26.

Glasserman, P., and H. Mamaysky. 2019. Does Unusual News Forecast Market Stress? *Journal of Financial and Quantitative Analysis* 54: 1937–1974.

Glosten, L., and P. Milgrom. 1985. Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics* 14: 71–100.

Green, T. C. 2004. Economic News and the Impact of Trading on Bond Prices. *The Journal of Finance* 59: 1201–1233.

Groß-Klußmann, A., and N. Hautsch. 2011. When Machines Read the News: Using Automated Text Analytics to Quantify High Frequency News-Implied Market Reactions. *Journal of Empirical Finance* 18: 321–340.

Grossman, S. J. 1986. An Analysis of the Role of 'Insider Trading' on Futures Markets. *The Journal of Business* 59: S129–S146.

Hansen, M. 2012. "A Look into the Ant Farm: The Market for Firm Specific News." Unpublished Working Paper. CREATES - the Center for Research in Econometric Analysis of Time Series, Aarhus University.

Hansen, P. R., A. Lunde, and V. Voev. 2014. Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility. *Journal of Applied Econometrics* 29: 774–799.

Heston, S. L., and N. R. Sinha. 2017. News Versus Sentiment: Predicting Stock Returns from News Stories. *Financial Analysts Journal* 73: 67–83.

Hillert, A., H. Jacobs, and S. Müller. 2014. Media Makes Momentum. *Review of Financial Studies* 27: 3467–3501.

Holden, C., and A. Subrahmanyam. 1992. Long-Lived Private Information and Imperfect Competition. *The Journal of Finance* 47: 247–270.

Hong, H., and J. C. Stein. 1999. A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets. *The Journal of Finance* 54: 2143–2184.

Hong, H., T. Lim, and J. C. Stein. 2000. Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies. *The Journal of Finance* 55: 265–295.

Huberman, G., and T. Regev. 2001. Contagious Speculation and a Cure for Cancer: A Nonevent That Made Stock Prices Soar. *The Journal of Finance* 56: 387–396.

Jeon, Y., T. H. McCurdy, and X. Zhao. 2019. *News as Sources of Jumps in Stock Returns: Evidence From 21 Million News Articles for 9000 Companies*. University of Toronto Working Paper. Available at SSRN: https://ssrn.com/abstract=3318517. (last accessed October 30, 2020)

Kalev, P. S., W.-M. Liu, P. K. Pham, and E. Jarnecic. 2004. Public Information Arrival and Volatility of Intraday Stock Returns. *Journal of Banking & Finance* 28: 1441–1467.

Kandel, E., and N. Pearson. 1995. Differential Interpretation of Public Signals and Trade in Speculative Markets. *Journal of Political Economy* 103: 831–872.

Kim, O., and R. E. Verrecchia. 1991. Market Reactions to Anticipated Announcements. *Journal of Financial Economics* 30: 273–309.

Kim, O., and R. E. Verrecchia. 1994. Market Liquidity and Volume around Earnings Announcements. *Journal of Accounting and Economics* 17: 41–67.

Lumsdaine, R. 2010. *What the Market Watched: Bloomberg News Stories and Bank Returns as the Financial Crisis Unfolded*. Unpublished Working Paper. Available at SSRN: https://ssrn.com/abstract=1482019. (last accessed October 30, 2020)

Maheu, J. M., and T. H. McCurdy. 2004. News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns. *The Journal of Finance* 59: 755–793.

Merton, R. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. The *Journal of Finance* 29: 449–470.

Mitchell, M. L., and J. H. Mulherin. 1994. The Impact of Public Information on the Stock Market. *The Journal of Finance* 49: 923–950.

Neuhierl, A., A. Scherbina, and B. Schlusche. 2013. Market Reaction to Corporate Press Releases. *Journal of Financial and Quantitative Analysis* 48: 1207–1240.

Patell, J., and M. Wolfson. 1984. The Intraday Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements. *Journal of Financial Economics* 13: 223–252.

Patton, A., and M. Verardo. 2012. Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability. *Review of Financial Studies* 25: 2789–2839.

Roll, R. 1984. Orange Juice and Weather. *American Economic Review* 74: 861–880.

Roll, R. 1988. R2 (R Squared). The *Journal of Finance* 43: 541–566.

Rubinstein, A. 1993. On Price Recognition and Computational Complexity in a Monopolistic Model. *Journal of Political Economy* 101: 473–484.

Shiller, R. 1981. Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends? *American Economic Review* 71: 421–436.

Tetlock, P. C. 2007. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance* 62: 1139–1168.

Tetlock, P. C. 2010. Does Public Financial News Resolve Asymmetric Information? *Review of Financial Studies* 23: 3520–3557.

Tetlock, P. C. 2011. All the News That's Fit to Reprint: Do Investors React to Stale Information? *Review of Financial Studies* 24: 1482–1512.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy. 2008. More than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance* 63: 1437–1467.

Theil, H., and J. C. G. Boot. 1962. The Final Form of Econometric Equation Systems. *Revue de L'institut International de Statistique / Review of the International Statistical Institute* 30: 136–152.

Thompson, R. B., C. Olsen, and J. R. Dietrich. 1987. Attributes of News about Firms: An Analysis of Firm-Specific News Reported in the Wall Street Journal Index. *Journal of Accounting Research* 25: 245–274.

Veldkamp, L. L. 2006. Media Frenzies in Markets for Financial Information. *American Economic Review* 96: 577–601.

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 39–52
doi: 10.1093/jjfinec/nbaa033
Advance Access Publication Date: 31 January 2021
Article

On the Autocorrelation of the Stock Market*

Ian Martin
London School of Economics and Political Science

Address correspondence to Ian Martin, Department of Finance, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, or e-mail: i.w.martin@lse.ac.uk.

Received March 5, 2019; revised August 19, 2020; editorial decision August 21, 2020; accepted August 24, 2020

Abstract

I introduce an index of market return autocorrelation based on the prices of index options and of forward-start index options and implement it at a six-month horizon. The results suggest that the autocorrelation of the S&P 500 index was close to zero before the subprime crisis but was negative in its aftermath, attaining values around –20% to –30%. I speculate that this may reflect market perceptions about the likely reaction, via quantitative easing, of policymakers to future market moves.

Do past returns on the market forecast future returns? Is the return on the market autocorrelated? It is well known that any asset return has zero risk-neutral autocorrelation (Samuelson, 1965).[1] But true autocorrelation may diverge significantly from zero—a point first made by LeRoy (1973)—and fluctuate over time. It is not clear whether one should expect positive or negative autocorrelation; indeed, both might be present simultaneously at different horizons. The former might be attributed to the influence of return-chasing investors in the investor population, as in the models of Hong and Stein (1999) and Vayanos and Woolley (2013), or to sluggish response to information; and the latter to bid-ask bounce, to overreaction as in the model of Barberis et al. (2015), or to the response of monetary authorities to fluctuations in asset prices.[2]

* I thank Anthony Neuberger, Anna Cieslak, Leifu Zhang, Fabio Trojani (the editor) and two referees for their comments; Can Gao for research assistance; and Dimitri Vayanos for posing the question that prompted this article. This work was carried out with the support of the Paul Woolley Centre and the ERC (Starting Grant 639744).

1 This statement is precisely true only if interest rates are deterministic; see below.

2 This is a highly incomplete, and somewhat arbitrary, list. Many other authors have studied the properties of autocorrelation; see, for example, Roll (1984), Grossman and Miller (1988), and Campbell, Grossman, and Wang (1993).

Several authors, including Fama and French (1988), Lo and MacKinlay (1988), Poterba and Summers (1988), and Moskowitz, Ooi, and Pedersen (2012), have studied the properties of realized autocorrelation of the market return, with results that vary depending on the horizon studied and on the sample period (on the latter, see Campbell, 2018, pp. 125–127). But how can we infer the forward-looking autocorrelation perceived by sophisticated investors? One straightforward approach is simply to ask investors what they think, following Shiller (1987) and others. But the expectations reported in such surveys appear to be far from rational: for example, Greenwood and Shleifer (2014) argue that times when surveyed investors are optimistic about future returns are in fact associated with low, not high, subsequent returns.

I therefore take a different approach and ask what autocorrelation must be perceived by a rational, risk-averse investor—specifically, by an unconstrained rational investor with log utility who chooses to invest his or her wealth fully in the market. It turns out to be possible to give a precise answer to this question in terms of the prices of various types of options. The fact that the autocorrelation index is computed directly from forward-looking asset prices, rather than from historical measures, is the major innovation of the article. The price to pay is that one has to accept the log investor's perspective as being a reasonable one to adopt. Nonetheless, related approaches have proved fruitful in forecasting returns on the stock market (Martin, 2017), on individual stocks (Martin and Wagner, 2019), and on currencies (Kremens and Martin, 2019); and the approach has the obvious advantage of bringing a novel type of evidence to bear on a classic question. Moreover, as in these earlier papers, we have the benefit of not requiring statistical assumptions on the underlying process (e.g., that it is stationary or ergodic). Such assumptions are widely made in the empirical literature, but they are not uncontroversial.

The theoretical results are derived in Section 1. They show that the autocorrelation index can be calculated from the prices of European index options and of forward-start index options. The latter options are relatively exotic, but I have been able to obtain indicative price quotes from a major investment bank for a small number of days between June 2007 and December 2013. Section 2 uses these prices to calculate the autocorrelation index and compares the implied forward-looking autocorrelations that emerge to the corresponding realized autocorrelations. Section 3 concludes.

1 Measuring Autocorrelation

Today is time t; the price of the underlying asset at time t is S_t. The goal is to measure the correlation between the return on the asset over the next period, $R_{t\to t+1}$ and the return over the following period, $R_{t+1\to t+2}$. I assume that the underlying asset does not pay dividends, so $R_{t\to t+1} = S_{t+1}/S_t$. I write the rate at which money can be risklessly invested from time u to time v as $R_{f,u\to v}$, so $R_{f,t\to t+1}$ and $R_{f,t\to t+2}$ are one- and two-period spot rates.

I write $\mathbb{E}_t^*$ for the risk-neutral expectation operator whose defining property is that the time-t price of a time-$(t+2)$ payoff X_{t+2} is $\frac{1}{R_{f,t\to t+2}}\mathbb{E}_t^* X_{t+2}$; and cov_t^* and corr_t^* for the corresponding risk-neutral covariance and risk-neutral correlation operators.

When seeking a measure of autocorrelation that can be computed directly from asset prices, the natural first thought is to consider *risk-neutral* autocorrelation. Unfortunately, we have the following well-known result.

Result 1. *Suppose that interest rates are deterministic. Then the return on any asset has zero risk-neutral autocorrelation:* $\text{corr}_t^*(R_{t\to t+1}, R_{t+1\to t+2}) = 0$.

Proof. By the defining property of the risk-neutral expectation operator, we have $\mathbb{E}_t^* R_{t\to t+1} = R_{f,t\to t+1}$ and $\mathbb{E}_{t+1}^* R_{t+1\to t+2} = R_{f,t+1\to t+2}$. As interest rates are deterministic, the second equality implies that $\mathbb{E}_t^* R_{t+1\to t+2} = R_{f,t+1\to t+2}$ by the law of iterated expectations. So we can write

$$\begin{aligned}\text{cov}_t^*(R_{t\to t+1}, R_{t+1\to t+2}) &= \mathbb{E}_t^*[(R_{t\to t+1} - R_{f,t\to t+1})(R_{t+1\to t+2} - R_{f,t+1\to t+2})] \\ &= \mathbb{E}_t^*[(R_{t\to t+1} - R_{f,t\to t+1})\mathbb{E}_{t+1}^*(R_{t+1\to t+2} - R_{f,t+1\to t+2})] \\ &= 0,\end{aligned}$$

using the law of iterated expectations again for the second equality. □

Although interest rates are not deterministic, they are typically extremely stable by comparison with returns on stock indices, so Result 1 rules out the autocorrelation perceived by a risk-neutral investor as a useful measure. Moreover, it is easy to adapt the proof above to show that the risk-neutral autocorrelation of *excess* returns is zero even if interest rates are stochastic.

How, then, can we define a non-trivial measure of autocorrelation? This article introduces an index that can be interpreted as the autocorrelation perceived by a rational, unconstrained investor with log utility whose wealth is fully invested in the market. The next result, which is also exploited by Martin (2017), provides the key to calculating this quantity.

Result 2. *Let* X_T *be some random variable of interest whose value becomes known at time T, and suppose that we can price a claim to* $X_T R_{t\to T}$ *delivered at time T. Then we can compute the expected value of* X_T *from the perspective of an investor with log utility whose wealth is invested in the market:*

$$\mathbb{E}_t X_T = \text{time } t \text{ price of a claim to the time } T \text{ payoff } X_T R_{t\to T}. \tag{1}$$

Proof. An investor with log utility who chooses to hold the market must perceive that the return on the market is growth-optimal. As the reciprocal of the growth-optimal return is a stochastic discount factor (SDF), the right-hand side of Equation (1) equals $\mathbb{E}_t\left[\frac{1}{R_{t\to T}} X_T R_{t\to T}\right]$, and the result follows. □

This result provides a general strategy for inferring the *true* expectation of the log investor from traded asset prices. If we can price the claim $X_T R_{t\to T}$, then we can infer the investor's expectation of X_T, even if X_T is not itself a tradable payoff. In particular, Result 2 will allow us to calculate $\text{corr}_t(R_{t\to t+1}, R_{t+1\to t+2})$.

To that end, we wish to compute

$$\text{cov}_t(R_{t\to t+1}, R_{t+1\to t+2}) = \mathbb{E}_t R_{t\to t+2} - \mathbb{E}_t R_{t\to t+1}\mathbb{E}_t R_{t+1\to t+2}. \tag{2}$$

By Result 2, $\mathbb{E}_t R_{t\to T}$ is equal to the price of a claim to the square of the return on the market, $R^2_{t\to T}$. This price can be calculated by a replication argument, as in Martin (2017), using the fact that

$$R^2_{t\to T} = \left(\frac{S_T}{S_t}\right)^2 = \frac{2}{S_t^2}\int_0^\infty \max\{0, S_T - K\}dK.$$

This equation expresses the desired payoff—the squared return—as the payoff on a portfolio that holds equal quantities of calls of all strikes. Thus

$$\mathbb{E}_t R_{t\to T} = \frac{2}{S_t^2}\int_0^\infty \text{call}_{t,T}(K)dK, \tag{3}$$

and setting $T = t+1$ and $T = t+2$ in this expression delivers the first two expectations on the right-hand side of Equation (2).

It is more difficult to compute $\mathbb{E}_t R_{t+1\to t+2}$, and doing so is the main innovation of the article. In view of Result 2, to calculate this quantity we need the time-t price of a claim to $R_{t+1\to t+2}R_{t\to t+2}$ delivered at time $t+2$. That is, we must price a claim to $S^2_{t+2}/(S_tS_{t+1})$.

It will turn out that we can replicate this claim using *forward-start options*. A forward-start call option that is initiated at time t, for settlement at time $t+2$, has the payoff

$$\max\{0, S_{t+2} - KS_{t+1}/S_t\}$$

for some fixed K. The unusual feature of a forward-start option is that its strike price, KS_{t+1}/S_t, is not determined until the intervening time $t+1$. (The introduction of S_t, a known constant from the perspective of time t, is simply a convenient normalization.) In contrast, the strike price of a conventional option is determined at the initiation of the trade. I write $\text{FScall}_t(K)$ for the time-t price of the above payoff, and $\text{FSput}_t(K)$ for the price of the corresponding put payoff, $\max\{0, KS_{t+1}/S_t - S_{t+2}\}$.

If we hold a portfolio consisting of $2/S_t^2 dK$ forward-start calls for each K, the portfolio payoff is

$$\frac{2}{S_t^2}\int_0^\infty \max\{0, S_{t+2} - KS_{t+1}/S_t\}dK = \frac{S^2_{t+2}}{S_tS_{t+1}}. \tag{4}$$

Since the payoff on the portfolio of forward-start calls replicates the desired payoff, the price of the payoff $S^2_{t+2}/(S_tS_{t+1})$ is the price of the portfolio of forward-start calls, and hence

$$\mathbb{E}_t R_{t+1\to t+2} = \frac{2}{S_t^2}\int_0^\infty \text{FScall}_t(K)dK. \tag{5}$$

Before using Equations (3) and (5) to compute the covariance $\text{cov}_t(R_{t\to t+1}, R_{t+1\to t+2})$, it will be convenient to rearrange them by replacing in-the-money calls and forward-start calls with out-of-the-money puts and forward-start puts. For vanilla options, we can do so by exploiting put-call parity, which in our context states that

$$\text{call}_{t,T}(K) - \text{put}_{t,T}(K) = S_t - \frac{K}{R_{f,t\to T}}.$$

The next result provides the corresponding relation for forward-start options.

Result 3 (Put-call parity for forward-start options)**.** *Let G_t be defined by the equation* $\text{FScall}_t(G_tS_t) = \text{FSput}_t(G_tS_t)$ *(so G_t is observable at time t, assuming the prices of forward-start options of all strikes are available). Then*

$$\text{FScall}_t(K) - \text{FSput}_t(K) = S_t - \frac{K}{G_t}. \tag{6}$$

If interest rates are deterministic, then G_t equals the forward (gross) interest rate for investment from time $t+1$ to $t+2$.

Proof. The time-$(t+2)$ payoff on a portfolio that is long a forward-start call and short a forward-start put, each with strike K, is $S_{t+2} - KS_{t+1}/S_t$. It follows that

$$\text{FScall}_t(K) - \text{FSput}_t(K) = S_t - \frac{1}{R_{f,t\to t+2}}\mathbb{E}_t^*\left[\frac{KS_{t+1}}{S_t}\right] = S_t - \lambda K,$$

where λ is the time-t price of a claim to S_{t+1}/S_t delivered at time $t+2$. We can pin down λ by applying the equation immediately above in the case $K = G_tS_t$ to conclude that $\lambda = 1/G_t$. This gives the result of Equation (6).
If interest rates are deterministic, $\lambda = 1/R_{f,t+1\to t+2}$. For we can replicate the payoff S_{t+1}/S_t, paid at time $t+2$, by investing $1/R_{f,t+1\to t+2}$ in the market from time t to $t+1$, and then at the riskless rate from time $t+1$ to $t+2$. □

Martin (2017) defined the volatility index $\text{SVIX}^2_{t,T} = \frac{1}{T-t}\text{var}_t^*(R_{t\to T}/R_{f,t\to T})$:

$$\text{SVIX}^2_{t,T} = \frac{2}{(T-t)R_{f,t\to T}S_t^2}\left[\int_0^{S_tR_{f,t\to T}} \text{put}_{t,T}(K)dK + \int_{S_tR_{f,t\to T}}^{\infty} \text{call}_{t,T}(K)dK\right].$$

We can define a forward volatility index FSVIX_t that is new to this article:

$$\text{FSVIX}^2_t = \frac{2}{G_tS_t^2}\left[\int_0^{S_tG_t} \text{FSput}_t(K)dK + \int_{S_tG_t}^{\infty} \text{FScall}_t(K)dK\right].$$

Using the put-call parity relations to substitute out calls and forward-start calls that have low strikes (i.e., are in-the-money), and then introducing these definitions, Equations (3) and (5) can be rewritten as

$$\mathbb{E}_tR_{t\to T} = R_{f,t\to T}\left(1 + (T-t)\text{SVIX}^2_{t,T}\right) \tag{7}$$

$$\mathbb{E}_tR_{t+1\to t+2} = G_t(1 + \text{FSVIX}^2_t). \tag{8}$$

These definitions lead to the following characterization.

Result 4. *The forward-looking autocovariance of returns, as perceived by the log investor, can be expressed in terms of spot and forward volatility indices as*

$$\begin{aligned}\operatorname{cov}_t(R_{t\to t+1}, R_{t+1\to t+2}) = R_{f,t\to t+2}(1 + 2\mathrm{SVIX}^2_{t,t+2}) \\ - R_{f,t\to t+1}G_t(1 + \mathrm{SVIX}^2_{t,t+1})(1 + \mathrm{FSVIX}^2_t).\end{aligned} \tag{9}$$

This expression simplifies if interest rates are deterministic:

$$\operatorname{cov}_t(R_{t\to t+1}, R_{t+1\to t+2}) = R_{f,t\to t+2}(2\mathrm{SVIX}^2_{t,t+2} - \mathrm{SVIX}^2_{t,t+1} - \mathrm{FSVIX}^2_t - \mathrm{SVIX}^2_{t,t+1}\mathrm{FSVIX}^2_t). \tag{10}$$

Proof. Equation (9) follows on substituting Equations (7) and (8) into the definition (2) of autocovariance. If interest rates are deterministic, then $R_{f,t\to t+1}G_t = R_{f,t\to t+2}$ (because, as shown in Result 3, G_t is then equal to the forward rate from $t+1$ to $t+2$); Equation (10) follows. □

Result 4 has the intuitive implication that forward-looking autocorrelation is positive if long-dated options (whose prices are embedded in $\mathrm{SVIX}_{t,t+2}$) are sufficiently expensive relative to short-dated and forward-start options (whose prices are embedded in $\mathrm{SVIX}_{t,t+1}$ and FSVIX_t).

The remaining task is to compute $\operatorname{var}_t R_{t\to t+1}$ and $\operatorname{var}_t R_{t+1\to t+2}$. As one might by now expect, the former can be computed from vanilla options and the latter from forward-start options. We have already calculated $\mathbb{E}_t R_{t\to t+1}$ and $\mathbb{E}_t R_{t+1\to t+2}$, so it only remains to find $\mathbb{E}_t R^2_{t\to t+1}$ and $\mathbb{E}_t R^2_{t+1\to t+2}$. By Result 2, the first of these is equal to the time-t price of a claim to $R^3_{t\to t+1}$ paid at time $t+1$, and since

$$\left(\frac{S_{t+1}}{S_t}\right)^3 = \frac{6}{S_t^3}\int_0^\infty K \max\{0, S_{t+1} - K\}\, dK,$$

the desired quantity is

$$\mathbb{E}_t R^2_{t\to t+1} = \frac{6}{S_t^3}\int_0^\infty K\ \mathrm{call}_{t,t+1}(K) dK. \tag{11}$$

The remaining term, $\mathbb{E}_t R^2_{t+1\to t+2}$, is equal to the price of a claim to $R^2_{t+1\to t+2}R_{t\to t+2}$ at time $t+2$. Since

$$R^2_{t+1\to t+2}R_{t\to t+2} = \frac{S^3_{t+2}}{S_t S^2_{t+1}} = \frac{6}{S_t^3}\int_0^\infty K \max\{0, S_{t+2} - KS_{t+1}/S_t\}\, dK,$$

we have

$$\mathbb{E}_t R^2_{t+1\to t+2} = \frac{6}{S_t^3}\int_0^\infty K\ \mathrm{FScall}_t(K)\ dK. \tag{12}$$

Using the put-call parity relations to replace in-the-money calls with out-of-the-money puts, Equations (11) and (12) become

$$\mathbb{E}_t R^2_{t\to t+1} - R^2_{f,t\to t+1} = \frac{6}{S_t^3}\left[\int_0^{S_t R_{f,t\to t+1}} K\ \mathrm{put}_{t,t+1}(K)\ dK + \int_{S_t R_{f,t\to t+1}}^\infty K\ \mathrm{call}_{t,t+1}(K)\ dK\right] \tag{13}$$

and

$$\mathbb{E}_t R_{t+1\to t+2}^2 - G_t^2 = \frac{6}{S_t^3}\left[\int_0^{S_t G_t} K\ \mathrm{FSput}_t(K)dK + \int_{S_t G_t}^{\infty} K\ \mathrm{FScall}_t(K)dK\right]. \tag{14}$$

Equations (7), (8), (13), and (14) provide the ingredients needed to calculate the autocorrelation index

$$\mathrm{corr}_t(R_{t\to t+1}, R_{t+1\to t+2}) = \frac{\mathbb{E}_t R_{t\to t+2} - \mathbb{E}_t R_{t\to t+1}\mathbb{E}_t R_{t+1\to t+2}}{\sqrt{\mathrm{var}_t R_{t\to t+1}\mathrm{var}_t R_{t+1\to t+2}}}. \tag{15}$$

1.1 The Autocorrelation Index in Homogeneous Models

In many familiar theoretical models, the autocorrelation index is exactly zero. (As we will see in the next section, this is counterfactual.) As an illustration, consider the model of Black and Scholes (1973). At time $t+1$, a forward-start call becomes identical to a vanilla call with strike KS_{t+1}/S_t, so by the Black–Scholes formula (with volatility σ and a continuously compounded riskless rate of r), the forward-start call is worth

$$S_{t+1}\Phi\left(\frac{\log\frac{S_t}{K} + r + \frac{1}{2}\sigma^2}{\sigma}\right) - K\frac{S_{t+1}}{S_t}e^{-r}\Phi\left(\frac{\log\frac{S_t}{K} + r - \frac{1}{2}\sigma^2}{\sigma}\right)$$

at time $t+1$. As a claim to S_{t+1} at time $t+1$ is worth S_t at time t, the above expression implies that at time t, the forward-start call is worth the same as a one-period vanilla call: $\mathrm{FScall}_t(K) = \mathrm{call}_{t,t+1}(K)$. It follows by put-call parity that $\mathrm{FSput}_t(K) = \mathrm{put}_{t,t+1}(K)$, and hence also that $\mathrm{FSVIX}_t^2 = \mathrm{SVIX}_{t,t+1}^2$. The autocorrelation index therefore takes a particularly simple form: as we have $\mathrm{SVIX}_{t,T}^2 = \frac{1}{T-t}\left(e^{\sigma^2(T-t)} - 1\right)$,

$$\mathrm{cov}_t(R_{t\to t+1}, R_{t+1\to t+2}) = e^{2r}\left(e^{2\sigma^2} - 1 - 2(e^{\sigma^2} - 1) - (e^{\sigma^2} - 1)^2\right) = 0.$$

That is, the autocorrelation index is zero in the Black–Scholes model. Another way to make the same point is that with constant risk aversion (through log utility) and constant volatility, the risk premium is constant, so there is no room for autocorrelation to arise through the drift term. As volatility is also constant, there is no autocorrelation at all.

More generally, let us say that a model is *homogeneous* if interest rates are constant and call prices have the property that $\mathrm{call}_{t,T}(K) = Kg(S_t/K, T-t)$ for some function g. (Many option-pricing models have this property, including the model of Black and Scholes (1973), the jump-diffusion model of Merton (1976), the variance-gamma model of Madan, Carr, and Chang (1998), and the model of Heston (1993) among others; the local volatility framework of Dupire (1994) is an example of a setting in which the homogeneity property does *not* hold.) We have the following result.

Result 5. *In a homogeneous model, the relationship between a forward-start option and a vanilla European option is trivial. That is, we have* $\mathrm{FScall}_t(K) = \mathrm{call}_{t,t+1}(K)$ *and* $\mathrm{FSput}_t(K) = \mathrm{put}_{t,t+1}(K)$. *It follows that* $\mathrm{FSVIX}_t^2 = \mathrm{SVIX}_{t,t+1}^2$ *in homogeneous models.*

Proof. A forward-start call with strike K, initiated at time t for final settlement at time $t+2$, has the payoff $\max\{0, S_{t+2} - KS_{t+1}/S_t\}$ at time $t+2$. From the perspective of time $t+1$, this is equivalent to the payoff on a vanilla call with strike KS_{t+1}/S_t. By the

homogeneity assumption, at time $t+1$, the vanilla call (and hence also the forward-start call) is worth

$$\frac{KS_{t+1}}{S_t} g\left(\frac{S_{t+1}}{KS_{t+1}/S_t}, 1\right) = \frac{KS_{t+1}}{S_t} g(S_t/K, 1).$$

The forward-start call is therefore worth $Kg(S_t/K, 1)$ at time t. In other words, by the homogeneity property, $\text{FScall}_t(K) = \text{call}_{t,t+1}(K)$. Hence $\text{FSput}_t(K) = \text{put}_{t,t+1}(K)$, by the put-call parity relations for vanilla and forward-start options. It follows that $\text{FSVIX}_t^2 = \text{SVIX}_{t,t+1}^2$ as an immediate corollary. □

1.2 Beyond Log Utility

Our measure of implied autocorrelation exploits asset price data alone, without reference to, say, survey forecasts, or accounting or macroeconomic data. Several recent papers, including Martin (2017), Kadan and Tang (2019), Kremens and Martin (2019), Martin and Wagner (2019), and Schneider and Trojani (2019), have adopted similar approaches. But we have made a stronger structural assumption on the form of the stochastic discount factor than these papers do, so it is natural to wonder whether the approach of this article can be generalized to allow for, say, power utility rather than log utility.

Unfortunately, it cannot. To be concrete, suppose we wish to compute the autocorrelation perceived by a hypothetical investor who has power utility over wealth at time $t+2$ and who chooses to invest fully in the market. As shown by Martin (2017, online appendix), the "easy" terms that appear in Equation (2)—namely, $\mathbb{E}_t R_{t\to t+1}$ and $\mathbb{E}_t R_{t\to t+2}$—can be calculated in this more general setting.

The difficulty lies in the term $\mathbb{E}_t R_{t+1\to t+2}$. To compute this quantity[3] with power utility (i.e., with an SDF proportional to $R_{t\to t+2}^{-\gamma}$), we would have to replicate (and hence price) the payoff $S_{t+2}^{1+\gamma}/(S_t^{\gamma} S_{t+1})$. To do so by holding a portfolio of $f(K)dK$ forward-start calls for each K—where $f(K)$ is some function that we can choose freely—we would need to have

$$\int_0^{\infty} f(K)\max\{0, S_{t+2} - KS_{t+1}/S_t\}dK = \frac{S_{t+2}^{1+\gamma}}{S_t^{\gamma} S_{t+1}}.$$

In the log utility case $\gamma = 1$, Equation (4) shows that $f(K)$ can be taken to be a constant known at time t. More generally, dividing through by S_{t+1}, we require that

$$\int_0^{\infty} f(K)\max\{0, S_{t+2}/S_{t+1} - K/S_t\}dK = \frac{S_{t+2}^{1+\gamma}}{S_t^{\gamma} S_{t+1}^2}. \tag{16}$$

The left-hand side of Equation (16) is a function of S_{t+2}/S_{t+1} (as S_t is known at time t). Therefore the right-hand side must also be a function of S_{t+2}/S_{t+1}; but this forces $\gamma = 1$. Thus, log utility is the only case in which this article's approach works.

3 Given a random variable X and SDF M, one can compute $\mathbb{E}X$ if the payoff X/M can be priced, as $\mathbb{E}X = \mathbb{E}\left(M\frac{X}{M}\right)$, and the latter is the price of the payoff X/M.

2 Empirical Results

To show how the theoretical results of Section 1 might be applied in practice, I obtained indicative price quotes for 6-month and 12-month vanilla call and put options on the S&P 500 index, together with 6-month-into-6-month forward-start options, from a major investment bank. All prices were supplied for a range of dates—June 15, 2007; June 20, 2008; November 21, 2008; February 20, 2009; December 17, 2010; July 15, 2011; December 20, 2012; and December 20, 2013—and for at-the-money, 5% and 10% out-of-the-money strikes for puts and for calls, together with the level of S&P 500 spot and the bank's internally marked 6-month and 12-month interest rate. I also obtained daily updated prices of vanilla European call and put options on the S&P 500 index from OptionMetrics in order to plot the daily time-series shown in Figure 1.

It should be emphasized at the outset that this exercise is a first step, given the limited data I have been able to obtain. A serious empirical exploration of the theoretical results of the previous section would require better data in the form both of a longer time series and of a more extensive range of strikes at each point in time.

I calculate the autocorrelation index (15) using expressions (7), (8), (13), and (14), interpolating linearly between option prices inside the range of observed strikes. I report results using two alternative methods to extrapolate option prices outside the observed range of strikes. In the first, I assume a flat volatility smile outside the range of observed strikes, following the approach of Carr and Wu (2009): in other words, for out-of-the-money puts with moneyness below the lowest observed strike, I use the Black–Scholes implied volatility at the lowest observed strike price, and for out-of-the-money calls, I use the Black–Scholes implied volatility at the highest observed strike price. In the second, I extrapolate implied volatilities linearly outside the range of observed strikes.[4] Lastly, I set G_t equal to the forward rate from $t+1$ to $t+2$.

Figure 1 shows the six-monthly autocorrelation of the S&P 500 index—that is, $\mathrm{corr}_t(R_{t\to t+6\mathrm{mo}}, R_{t+6\mathrm{mo}\to t+12\mathrm{mo}})$—on a sample of dates. The solid line uses a flat volatility smile for options with strikes outside the observed range; the dashed line extrapolates implied volatility linearly outside the observed range, as described in the previous paragraph. Both methods deliver similar conclusions: autocorrelation was close to zero at the beginning of the sample period, and declined sharply following the subprime crisis.

For comparison, Figure 2 plots the realized autocorrelation of six-monthly price changes, $P_{t+6\mathrm{mo}}/P_t$, of the S&P 500 index over various time periods. I take the end-of-month level of the S&P 500 index from CRSP, over the period January 1950 to September 2019. The sample autocorrelation depends on which start month is chosen, so in each panel, I show every possible choice. For example, the autocorrelation of January–July and July–January price changes was around 0.08 over the full sample period and around –0.2 over the most recent decade, whereas the autocorrelations of the corresponding March–September and September–March price changes were around –0.1 and –0.6, respectively.

4 I introduce a floor at zero volatility in cases where a downward-sloping smile would lead to negative implied volatilities at very high strikes; the results are not sensitive to where the floor occurs because the associated prices of deep-out-of-the-money calls are essentially zero for any reasonable level of volatility.

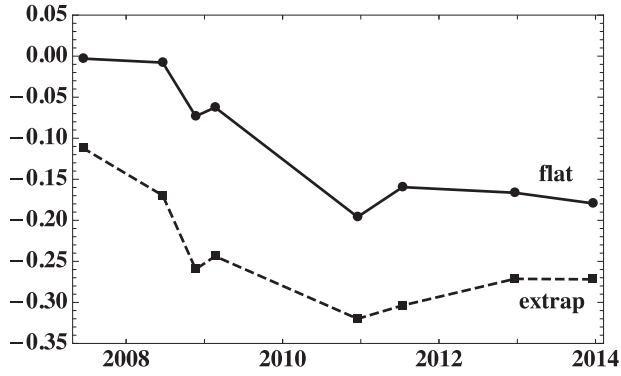

Figure 1 The autocorrelation of the S&P 500, $\text{corr}_t(R_{t\to t+6\text{mo}}, R_{t+6\text{mo}\to t+12\text{mo}})$. The solid line imposes a flat volatility smile for strikes outside the observed range of strikes. The dashed line extrapolates implied volatility linearly outside the observed range of strikes.

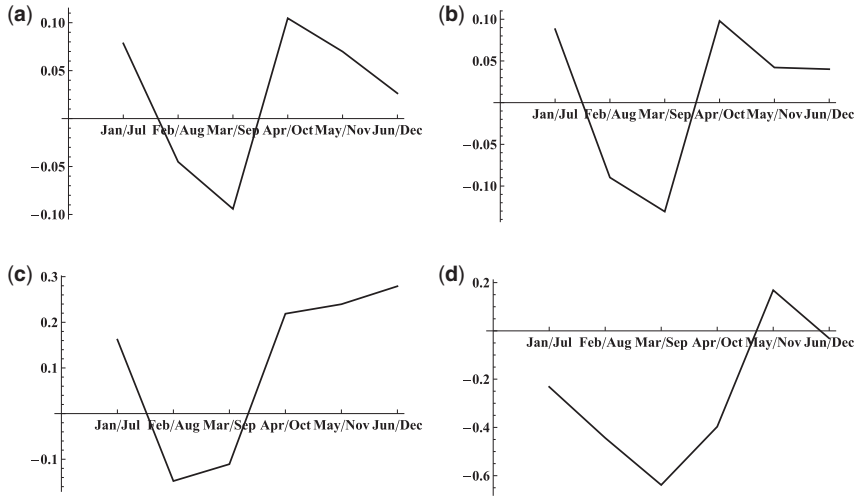

Figure 2 Realized autocorrelation in six-monthly price changes of the S&P 500 index over various time periods. (a) 1950–2019. (b) 1970–2019. (c) 1990–2019. (d) 2010–2019.

In the context of Figure 2, the magnitude of the autocorrelation index shown in Figure 1 appears reasonable.

Even so, it might seem, given Figure 1, that strategies designed to exploit reversals should earn Sharpe ratios that are too good to be true. In order to assess this possibility, the next result shows how to use vanilla option prices to calculate the maximum attainable Sharpe ratio perceived by the log investor.

Result 6. *The maximal Sharpe ratio over the period from t to $t+n$, as perceived by the log investor, satisfies*

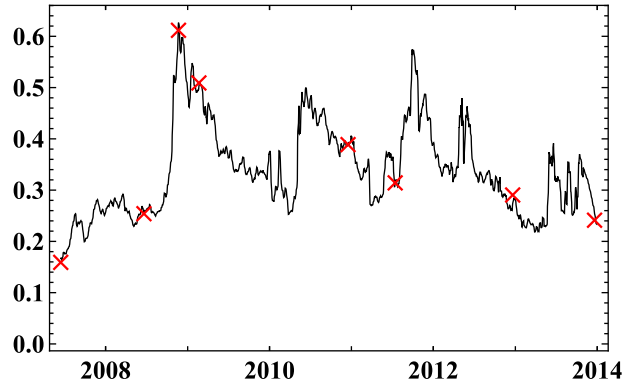

Figure 3 The maximal Sharpe ratio (one-year horizon). Crosses indicate the dates on which the autocorrelation index is computed in Figure 1.

$$\text{max Sharpe ratio} \leq R_{f,t\to t+n}\sqrt{\int_0^\infty \frac{2S_t}{K^3}\Omega_{t,t+n}(K)dK}, \tag{17}$$

where $\Omega_{t,t+n}(K)$ is the time t price of an out-of-the-money European option with strike K expiring at time $t+n$:

$$\Omega_{t,t+n}(K) = \begin{cases} \text{put}_{t,t+n}(K) & \text{if } K \leq S_t R_{f,t\to t+n} \\ \text{call}_{t,t+n}(K) & \text{if } K > S_t R_{f,t\to t+n} \end{cases}$$

Proof. Using the result of Hansen and Jagannathan (1991) and the fact that $M_{t\to t+n} = 1/R_{t\to t+n}$, we have

$$\begin{aligned} \text{max Sharpe ratio} &\leq R_{f,t\to t+n}\sqrt{\text{var}_t \frac{1}{R_{t\to t+n}}} \\ &= \sqrt{R^2_{f,t\to t+n}\mathbb{E}_t \frac{1}{R^2_{t\to t+n}} - 1}. \end{aligned} \tag{18}$$

By the result of Breeden and Litzenberger (1978), as rewritten by Carr and Madan (1998), the time t price of a claim to $f(S_{t+n})$ paid at time $t+n$ is

$$\mathbb{E}_t \frac{f(S_{t+n})}{R_{t\to t+n}} = \frac{f(S_t R_{f,t\to t+n})}{R_{f,t\to t+n}} + \int_0^\infty f''(K)\Omega_{t,t+n}(K)dK.$$

Setting $f(K) = S_t/K$, this implies that

$$\mathbb{E}_t \frac{1}{R^2_{t\to t+n}} = \frac{1}{R^2_{f,t\to t+n}} + \int_0^\infty \frac{2S_t}{K^3}\Omega_{t,t+n}(K)dK. \tag{19}$$

The result follows on substituting Equation (19) into Equation (18). □

Figure 3 plots the time series of the right-hand side of inequality (17) of Result 6, which provides an upper bound on the maximal Sharpe ratio at a one-year horizon. The dates on which the autocorrelation index is calculated in Figure 1 are marked with crosses. While the maximum attainable Sharpe ratio (as perceived by the log investor) spiked in late 2008, it was not implausibly high. Thus, although there are several potential ways reversal strategies might be implemented in practice, none of them has an unreasonably high Sharpe ratio from the perspective of the log investor.

3 Discussion

This article has introduced a new index of autocorrelation and constructed it at the six-month horizon using indicative prices obtained from a major investment bank on various days between mid-2007 and late 2013. Implied autocorrelation was close to zero at the beginning of the sample period but turned negative, in the range of –0.2 to –0.3, following the subprime crisis.

Negative autocorrelation during this period may have been driven by market participants' expectations about the behavior of policymakers. The Federal Open Market Committee (FOMC) statement of September 21, 2010 contains the following paragraph,[5] which heralded a second round of quantitative easing (QE2):

> The Committee will continue to monitor the economic outlook and financial developments and is prepared to provide additional accommodation if needed to support the economic recovery and to return inflation, over time, to levels consistent with its mandate.

Based on this statement, it would have been reasonable to conclude that policy would be more expansive conditional on further declines in the market and relatively more contractionary conditional on further rises and, hence, to anticipate a decline in market autocorrelation. Indeed, Cieslak, Morse, and Vissing-Jorgensen (2018) argue that the behavior of stock returns over the "FOMC cycle" is consistent with this view (though they focus on shorter horizons and emphasize the importance of timing within the cycle). Consistent with this interpretation, the low point of the autocorrelation measure occurs in December 2010.

As the autocorrelation index depends only on asset prices, it has the great advantage of being computable, in principle, in real time. The central novel feature of the index is that it is based on the prices of forward-start index options. As shown by Hobson and Neuberger (2012), the prices of forward-start options are not tightly constrained by the prices of ostensibly closely related vanilla options. This fact is precisely what makes them interesting; nonetheless, it should be emphasized that they are exotic derivatives, with all the caveats that entails—most notably, that the forward-start option market is not nearly as liquid as the vanilla option market. A full empirical investigation of the theoretical results of this article would require considerably more data than I have been able to obtain.

5 The precise phrasing of the paragraph was discussed extensively during the meeting: see pages 78, 98, 101, 113, and 124–126 of the Transcript of the Federal Open Market Committee Meeting on September 21, 2010, which is available at https://www.federalreserve.gov/monetarypolicy/files/FOMC20100921meeting.pdf (last accessed on September 4, 2020). One consequence of the discussion was that the phrase "as needed" was replaced with "if needed," which was felt to emphasize the conditionality of any potential Fed action more clearly.

A further contribution of the article, however, is to point out that such options—variants of the more familiar "plain vanilla" European call and put options—have a natural economic application. It is sometimes tempting, when confronted with a cliquet, a lookback, a Napoleon, Himalayan, Bermudan, Asian, best-of, worst-of, or rainbow option, or with any other member of the bewildering menagerie of exotic derivatives, to conclude that such contracts play no more significant a role than to transfer resources between groups of quants. Precisely because there is an element of truth in this caricature, financial economists have a role to play in pointing out when some seemingly obscure derivative contract is in fact of economic interest.

References

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer. 2015. X-CAPM: An Extrapolative Capital Asset Pricing Model. *Journal of Financial Economics* 115: 1–24.

Black, F., and M. Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81: 637–659.

Breeden, D. T, and Litzenberger R.H.1978. Prices of State-Contingent Claims Implicit in Option Prices. Journal of Business 51: 621–651.

Campbell, J. Y. 2018. *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton, NJ: Princeton University Press.

Campbell, J. Y., S. J. Grossman, and J. Wang. 1993. Trading Volume and Serial Correlation in Stock Returns. *The Quarterly Journal of Economics* 108: 905–939.

Carr, P., and D. Madan. 1998. "Towards a Theory of Volatility Trading." In R. Jarrow (ed.), *Volatility: New Estimation Techniques for Pricing Derivatives*. London: Risk Books, pp. 417–427.

Carr, P., and L. Wu. 2009. Variance Risk Premiums. *Review of Financial Studies* 22: 1311–1341.

Cieslak, A., A. Morse, and A. Vissing-Jorgensen. 2018. Stock Returns over the FOMC Cycle. *Journal of Finance*, forthcoming.

Dupire, B. 1994. Pricing with a Smile. *Risk* 7: 18–20.

Fama, E. F., and K. R. French. 1988. Permanent and Temporary Components of Stock Prices. *Journal of Political Economy* 96: 246–273.

Greenwood, R., and A. Shleifer. 2014. Expectations of Returns and Expected Returns. *Review of Financial Studies* 27: 714–746.

Grossman, S. J., and M. H. Miller. 1988. Liquidity and Market Structure. *The Journal of Finance* 43: 617–633.

Hansen, L. P., and R. Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99: 225–262.

Heston, S. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6: 327–343.

Hobson, D., and A. Neuberger. 2012. Robust Bounds for Forward Start Options. *Mathematical Finance* 22: 31–56.

Hong, H., and J. Stein. 1999. A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *The Journal of Finance* 54: 2143–2184.

Kadan, O., and X. Tang. 2019. A Bound on Expected Stock Returns. *The Review of Financial Studies* 33: 1565–1617.

Kremens, L., and I. W. R. Martin. 2019. The Quanto Theory of Exchange Rates. *American Economic Review* 109: 810–843.

LeRoy, S. F. 1973. Risk Aversion and the Martingale Property of Stock Prices. *International Economic Review* 14: 436–446.

Lo, A. W., and A. C. MacKinlay. 1988. Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test. *Review of Financial Studies* 1: 41–66.

Madan, D. B., P. P. Carr, and E. C. Chang. 1998. The Variance Gamma Process and Option Pricing. *European Finance Review* 2: 79–105.

Martin, I. W. R. 2017. What is the Expected Return on the Market? *The Quarterly Journal of Economics* 132: 367–433.

Martin, I. W. R., and C. Wagner. 2019. What Is the Expected Return on a Stock? *The Journal of Finance* 74: 1887–1929.

Merton, R. C. 1976. Option Pricing When Underlying Stock Price Returns Are Discontinuous. *Journal of Financial Economics* 3: 125–144.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen. 2012. Time Series Momentum. *Journal of Financial Economics* 104: 228–250.

Poterba, J. M., and L. H. Summers. 1988. Mean Reversion in Stock Prices: Evidence and Implications. *Journal of Financial Economics* 22: 27–59.

Roll, R. 1984. A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance* 39: 1127–1139.

Samuelson, P. A. 1965. Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review* 6: 41–49.

Schneider, P., and F. Trojani. 2019. (Almost) Model-Free Recovery. *The Journal of Finance* 74: 323–370.

Shiller, R. J. 1987. "Investor Behavior in the October 1987 Stock Market Crash: Survey Evidence." NBER Working Paper 2446.

Vayanos, D., and P. Woolley. 2013. An Institutional Theory of Momentum and Reversal. *Review of Financial Studies* 26: 1087–1145.

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 53–96
doi: 10.1093/jjfinec/nbaa029
Article

OXFORD

Regulatory Capital and Incentives for Risk Model Choice under Basel 3*

Fred Liu and Lars Stentoft
University of Western Ontario
Address correspondence to Lars Stentoft, Department of Economics, University of Western Ontario, Ontario, Canada, or e-mail: lars.stentoft@uwo.ca.

Received June 30, 2019; revised July 23, 2020; editorial decision July 27, 2020; accepted July 27, 2020

Abstract

In response to the Subprime mortgage crisis, the Basel Committee on Banking Supervision (BCBS) has spent the previous decade overhauling the regulatory framework that governs how banks calculate minimum capital requirements. In 2019, the BCBS finalized the Basel 3 regulatory regime, which changes the regulatory measure of market risk and adds new complex calculations based on liquidity and risk factors. This article is motivated by these changes and seeks to answer the question of how regulation affects banks' choice of risk-management models, whether it incentivizes them to use correctly specified models, and if it results in more stable capital requirements. Our results show that, although the models that minimize regulatory capital for a representative bank portfolio also result in the most stable requirements, these models are generally rejected as being correctly specified and tend to produce inferior forecasts of the regulatory risk measures.

Key words Basel 3, expected shortfall backtesting, regulatory capital requirements

JEL classification C14, C22, C53, G32

The regulatory environment that governs how banks calculate minimum capital requirements has changed dramatically in recent years. First, in response to the 2008 Subprime Mortgage Crisis, the Basel Committee on Banking Supervision (BCBS) adopted the Stressed Value at Risk (VaR) measure. VaR is the conditional quantile of the loss distribution at a given confidence level, and Stressed VaR is defined as the VaR on a 1-year historical dataset with significant financial stress (BCBS 2011). This change significantly increased the capital

* The authors thank participants at the 2016 University of Waterloo PhD Conference, the 2017 meeting in the Canadian Economic Association, the 2017 Bank of Canada Graduate Student Paper Award, the 2017 meeting in the Canadian Econometric Study Group, and seminar participants at the University of Western Ontario for helpful comments on previous versions of this article. The article benefited from excellent comments from two anonymous referees and the editor. All remaining errors are our responsibility.

charges banks faced. Second, in 2019 the BCBS finalized the Fundamental Review of the Trading Book (FRTB) regulatory regime as part of Basel 3. This will change the regulatory measure of market risk from VaR at the 99% confidence level to Expected Shortfall (ES) at the 97.5% confidence level (BCBS 2019). ES is defined as the expected loss conditional on VaR being exceeded and is widely perceived to be a more appropriate risk measure as it captures both the size and likelihood of losses (BCBS 2013). Additionally, in Basel 3 Stressed ES at the 97.5% confidence level, defined as the ES during the most severe 1-year period of losses available, replaces Stressed VaR as the key risk measure.

The new capital requirements for market risk present an interesting trade-off for banks. Assuming bank capital is costly and hence banks minimize their capital requirements, Basel 3 incentivizes them to use models producing low Stressed ES. At the same time, Basel 3 penalizes banks by increasing their capital requirements when their model generates too many VaR exceedances, which incentivizes banks to use more conservative models. It is not obvious how different models for VaR and Stressed ES perform in terms of this trade-off and which model is therefore privately optimal in the sense that it minimizes banks' capital requirements. Moreover, it is likely even more costly, and sometimes impossible, for banks to raise capital during times of high volatility, so prudence might suggest building a buffer stock of capital during good times that can be drawn down during times of stress. Basel 3 aims at creating more stable capital requirements by focusing on stressed risk measures. It is also not obvious how different models for VaR and Stressed ES perform in terms of minimizing measures of capital requirement stability.

Our article is motivated by these changes and seeks to answer the question of how regulation affects banks choice of risk-management models and whether it incentivizes them to use correctly specified models. We also analyze whether the changes made to regulation incentivize banks to use models that lead to more stable (through-the-cycle) capital requirements rather than more dynamic (point-in-time), and therefore potentially more systemically risky, capital requirements. To preview our results, we find that, although the proposed regulation under Basel 3 seems to incentivize banks to use models that have more stable capital requirements through time, the models that minimize average capital requirements appear misspecified, in the sense that they are rejected using standard backtests, and produce inferior forecasts of the regulatory risk measures.

To answer these questions we construct a portfolio of diverse assets, with different risk factors and subject to different levels of liquidity risk, that a bank might realistically hold, and we consider three classes of models for the dynamics of the losses of this portfolio: (i) *ad hoc* methods like Historical Simulation (HS) and RiskMetrics (RM) that involve no parameter estimation, (ii) models within the classical GARCH framework of Engle (1983) and Bollerslev (1986) that are estimated by fitting model parameters to the entire dynamic distribution of losses, and (iii) a new class of models developed by Patton, Ziegel, and Chen (2019) that are estimated by fitting model parameters directly to the dynamics of the risk measures instead, which we refer to as "FZ" models. For the dynamic models, and the GARCH models, in particular, we consider several types of conditional distributions that can accommodate stylized facts like heavy tails and skewness toward losses as documented by Hansen (1994) among others.[1]

1 Using a rolling window estimation setup we also accommodate the recent observation that the conditional distribution and tails in particular of financial returns vary significantly over time and this particularly so during times of crisis (Bollerslev and Todorov 2014 and Kelly and Jiang 2014).

We first examine which of the models are correctly specified using both traditional VaR backtests and a battery of recent joint VaR and ES backtests, and we examine which models produce optimal forecasts of the risk measures.[2] As expected our results show that the *ad hoc* models and the HS model, in particular, fail the backtests and provide inferior forecasts of the risk measures. More interestingly though, our results also show that the FZ models fail some backtests and provide inferior forecasts of the VaR and ES compared to the subset of GARCH models that allow for skewed conditional distributions and are not rejected by the backtests.

Next, we carefully calculate the capital requirements for each model using the Basel 3 formulas and compare the results to those using previous regulatory regimes. Our results show that the HS model has the lowest average Basel 3 capital of 17.21% whereas the best model in the FZ class has average capital requirements of 17.42%. However, both of these models, along with several other models that have low capital requirements, are rejected as being correctly specified or shown to produce inferior forecasts of the regulatory risk measures. The best performing model which is not rejected by the backtests, a skewed GARCH model, provides superior risk measure forecasts but requires nearly 1.55 times higher capital than the HS model. The Filtered Historical Simulation (FHS) model, which is also not rejected and in the set of models that provide superior risk measure forecasts, has capital requirements that are roughly 1.78 times larger than the incorrectly specified HS model. Under Basel 3, there is therefore little incentive for a capital requirement minimizing bank to choose correctly specified models. Compared across the previous regimes our results show that correctly specified models are in fact never the models that minimize capital requirements.

Finally, we measure the variability of the minimum capital requirements under Basel 2, 2.5, and 3 to determine whether the new regulation is successful at increasing the stability of capital requirements. We consider several measures of the volatility of regulatory capital and we also measure peak-to-trough variation as the maximum difference in capital requirements. The BCBS has also focused on the procyclicality of regulation, often measured as peak-to-trough variation in minimum capital requirements (Gordy and Howells 2006; Heid 2007; Shim 2013). Our results show that capital requirements became significantly more stable from Basel 2 to Basel 2.5 due to the introduction of Stressed VaR. Basel 3 will further increase the stability of capital requirements by decreasing nonstandardized and standardized volatility across most models and could also reduce the procyclicality of capital requirements as evidenced by the lower peak-to-trough variation across most models under this regime compared to previous regimes. However, the results also show that the models that result in the most stable capital requirements across regulatory regimes and across variability metrics are generally not the correctly specified models.

2 The methods we use to backtest predicted VaR are standard and can be found in, for example, Christoffersen (2009). In terms of joint VaR and ES backtests, we consider Mcneil and Frey (2000)'s residual test (ER), Bayer and Dimitriadis (2018)'s strict, auxiliary, and one-sided regression tests (ESR), Nolde and Ziegel (2017)'s conditional calibration test, and Gordy and Mcneil (2020)'s spectral backtests. To assess the optimality of a model's forecasts, we consider the Model Confidence Set of Hansen, Lunde, and Nason (2011), the pairwise forecast performance tests from Diebold and Mariano (2002), and Ziegel et al. (2017)'s Murphy Diagrams.

Our findings have important implications for current regulation. In particular, our results show that Basel 3 regulation strongly disincentivizes banks from using correctly specified models. In fact, banks can minimize both the mean and volatility of Basel 3 capital by using the hybrid FZ model of Patton, Ziegel, and Chen (2019) which is not only rejected by some of the backtests but also provides inferior risk measure forecasts for our portfolio. Although the same holds for previous regulatory regimes, the changes suggested in Basel 3 make the relative differences even larger. For example, under Basel 2.5 FHS is only marginally worse than HS and would require only 1.09, instead of 1.78 under Basel 3, times the capital. We identify two possible reasons for this. First, and this is somewhat subtle, it appears that the requirement under Basel 3 to penalize low liquidity assets, that is, assets with long liquidity horizons, additionally, in fact disincentivizes banks from using the correctly specified skewed GARCH models due to their consistently high Stressed ES across liquidity horizons. Second, and this is more obvious, given the low level of the Basel 3 multiplier banks have little incentive to choose conservative and correctly specified models. Thus, if the regulator's objective is to incentivize the use of correctly specified models they would have to reconsider the effect of these changes.[3]

Our article is related to at least three strands of existing literature. First of all, there is a large literature on empirically backtesting VaR (see e.g., Christoffersen, Hahn, and Inoue 2001; Gencay and Selcuk 2004) and a growing literature on backtesting VaR and ES jointly. Our article complements this literature in several ways. First, while most backtesting papers focus on one asset class (most commonly equities), we consider a large and diverse set of assets spanning the multiple risk factors and liquidity horizons that banks may be exposed to. Second, we use cutting-edge ES backtesting techniques recently developed by academics to determine if models are correctly specified and we compare these models' forecasts of VaR and ES. Finally, we go beyond simple backtesting by calculating Basel 3 capital requirements for the representative bank portfolio and by evaluating the trade-off between correctly specified models and models that minimize not only the level but also the variability of the capital requirements.

Next, our article is related to studies of backtesting that use actual bank P&L or VaR data. For example, Berkowitz and O'brien (2002) show that U.S. banks had conservative VaR estimates during the 1998 Asian crisis with few exceedances, but the exceedances were clustered indicating bank models did not adapt to dynamic volatility. Pérignon, Deng, and Wang (2008) find that Canadian banks also have conservative VaR forecasts and Pérignon and Smith (2008) extend the results to international banks. O'brien and Szerszeń (2017), on the contrary, show that these early results were driven by a calm sample period and that U.S. banks had excessive exceedances and clustering during the 2008 financial crisis. Their findings suggest that the banks used misspecified models that do not adapt to time-varying volatility. Berkowitz, Christoffersen, and Pelletier (2011) compare the accuracy of VaR forecasts for trading desks in a commercial bank using a novel spectral backtest. Gordy and Mcneil (2020) study model-implied probabilities associated with bank-reported P&L and reject the hypothesis of correct specification for many U.S. bank models. While these

3 Another possibility is that regulators could attempt to incentivize the use of correctly specified models by penalizing models with high standardized volatility, the only capital stability measure for which correctly specified models, in this case models based on Extreme Value Theory, perform the best.

studies use actual bank data, they are limited in that they do not observe the bank's internal model or portfolio composition. They also do not calculate capital requirements under Basel 3. Our study complements this literature by using a transparent representative bank portfolio and well-known models.

Finally, our study is similar in spirit to the annual BCBS monitoring exercise that provides a hypothetical portfolio for banks to calculate actual risk measures and capital requirements, see, for example, BCBS (2014). However, in these exercises, the banks' internal models are confidential and the BCBS only reports aggregate capital results. Also, formal backtests are not performed to determine if the internal models used by banks are correctly specified as a part of these exercises. Our research complements the BCBS exercises by examining model VaR and ES forecasts on a hypothetical portfolio. We are able to backtest if the models are correctly specified and compare the mean and variability of their capital requirements. Hence, a bank's regulatory supervisor could assess the trade-off between correct specification, capital requirements, and capital stability for the bank's internal model.

The article is organized as follows: Section 1 describes how the regulatory environment has changed over time and outlines the requirements and objectives of the current regulation. Section 2 presents the data used, reviews the different classes of dynamic models considered, and explains how multiperiod risk measures can be calculated with these models. Section 3 contains extensive backtests for the models considered, analyzing which of them are correctly specified and which produce the best risk measure forecasts. Section 4 calculates the regulatory capital under Basel 3 and the previous regulatory regimes and assesses the stability through time of different model's capital requirement. Section 5 concludes. The Online Appendix contains further details on the regulatory calculations, on how to select optimal thresholds for Extreme Value Theory (EVT) models, on the backtesting methods used for ES, and some additional results.

1 Regulatory Capital Calculations

The last 25 years have seen significant changes to the regulation faced by banks when it comes to the regulatory capital requirements. In particular, over this period the Basel Accord has had three regimes for calculating market risk capital requirements: Basel 2, Basel 2.5, and the incoming Basel 3. While banks are currently adapting to the new requirements of Basel 3, it remains important to consider how we arrived at this regulation as well as the motivation and implications of this changing regulation.

To set the scene for the rest of the article, this section first provides a brief background on the regulation that led to Basel 3 paying special attention to how this has changed the calculations of minimum capital requirements. We then provide details on the proposed Basel 3 capital requirements and its formulas for calculating market risk capital. Finally, we discuss the intended and expected impact of this regulation on banks and the requirements it imposes on them.

1.1 Background

The Basel 2 market risk requirements were first introduced by the BCBS in the 1996 Amendment to the Basel Accord (BCBS 1996a) and allowed banks to use their own "internal" models to calculate regulatory capital. The internal model-based approach for

setting market risk capital involved calculating a VaR measure with a 10-day time horizon and at a 99% confidence level. Denoting daily losses, or negative returns, on a single asset or portfolio by L_t then formally the VaR measure at time t is defined as the value VaR^p_{t+1} such that

$$Pr(L_{t+1} > VaR^p_{t+1}|\Im_t) = p, \tag{1}$$

where $\Im_t$ is the information at time t and p is the coverage probability. That is, at period $t+1$, losses exceed VaR^p_{t+1} only with probability p, given the available information. The day t capital requirement for a bank with this VaR is then set as

$$CA^{B2}_t = \max(VaR_{t-1}, m_c \times \overline{VaR}_{t-1}), \tag{2}$$

where m_c is a multiplicative factor, VaR_{t-1} is the previous day's VaR, and $\overline{VaR}_{t-1}$ is the average VaR over the previous 60 days. The multiplicative factor m_c has a minimum value of three and adds a scaling factor between zero and one that depends on the model's backtesting performance and penalizes models that backtest poorly, see BCBS (1996b). Basel 2 backtesting compares VaR with a one-day time horizon at the 99% confidence level to realized exceedances (losses above VaR) over the previous 250 days.[4]

A given internal model's multiplication factor is set according to the "traffic light" system of exceedances reproduced in Table 1. Column 3 shows that if the number of exceedances in the previous 250 days is four or fewer, the model is in the Green Zone and the multiplicative factor m_c is three. If the number of exceedances is between five and nine, the model is in the Yellow Zone and m_c is between 3.4 and 3.85. Note that the large discrete jump between four and five exceedances increases the penalty with >10%. If the number of exceedances is >10, the model is in the Red Zone and m_c is set at four. Additionally, during Red Zone periods, regulatory supervisors can disallow the use of a particular internal model, which forces the bank to use the standard model approach. This could potentially increase overall capital requirements significantly.

The 2008 Subprime Mortgage Crisis revealed that the Basel 2 requirements were far too low to capture systemic risks, which resulted in banks holding insufficient capital before the crisis. Additionally, the sudden jump in capital requirements during the crisis caused banks to deleverage by sharply shedding risk exposures (Adrian and Shin 2014), resulting in fire-sales, liquidity spirals, and disinflation (Brunnermeier and Sannikov 2016). The reduction in bank balance sheets and credit at the height of the financial crisis caused by the procyclicality of Basel 2 capital requirements amplified the downturn (Adrian and Shin 2014).

To address the shortcomings of Basel 2, the BCBS introduced the current regulatory regime, Revisions to the Basel 2 Framework (Basel 2.5), with an implementation date of December 31, 2011 (BCBS 2009). In Basel 2.5, the level and stability of capital requirements are increased by introducing the Stressed VaR measure. Stressed VaR is VaR calculated during a 12-month period of significant financial stress. The stress period is identified

4 Backtesting is not done on a 10-day horizon, since the portfolio composition may change within this horizon.

Table 1 Basel backtesting zone boundaries

Backtesting zone	Exceedances	Basel 2 multiplier	Basel 3 multiplier	Cumulative probability (%)
Green	4 or fewer	3	1.5	89.22
Yellow/Amber	5	3.4	1.7	95.88
–	6	3.5	1.76	98.63
–	7	3.65	1.83	99.60
–	8	3.75	1.88	99.89
–	9	3.85	1.92	99.97
Red	10 or more	4	2	99.99

This table defines the green, yellow/amber, and red zones that supervisors use to assess backtesting results in conjunction with the internal models approach to market risk capital requirements under Basel 2 and 3, see BCBS (1996b) and BCBS (2019). The boundaries shown in the table are based on a sample of 250 observations.

as the 12 months in history that maximizes VaR with a 10-day time horizon and a 99% confidence level (EBA 2012). The Basel 2.5 capital charge is set at

$$CA_t^{B2.5} = \max(VaR_{t-1}, m_c \times \overline{VaR}_{t-1}) + \max(SVaR_{t-1}, m_s \times \overline{SVaR}_{t-1}), \tag{3}$$

where the first term is the Basel 2 capital charge, $SVaR_{t-1}$ is the previous day's Stressed VaR, $\overline{SVaR}_{t-1}$ is the average Stressed VaR over the previous 60 days, and m_s is the stressed multiplicative factor set by the regulatory supervisor. Since Stressed VaR is always at least as large as VaR and assuming $m_c = m_s$, the Basel 2.5 capital charge is at least double the Basel 2 charge. Also since the period for Stressed VaR rarely changes Basel 2.5 should result in more stable capital requirements.

1.2 Basel 3

While Basel 2.5 was implemented to temporarily increase the level and stability of capital requirements to address some of the serious flaws in the capital regulation framework exposed by the 2008 crisis, the BCBS has continued to work on implementing more stringent capital calculations. In 2014, the BCBS proposed Basel 3, formally named the FRTB, as a new and comprehensive approach to determining minimum regulatory capital for market risk. The key changes to market risk calculations under Basel 3 include the use of ES instead of VaR, calculations based on liquidity horizons, calibration to periods of significant financial stress, and diversification restrictions. We now summarize each of the changes and explain the Basel 3 market risk formula based on the finalized FRTB documentation (BCBS 2019).

In Basel 3, the measure used to determine capital changes from 10-day VaR at a 99% confidence level to 10-day ES at a 97.5% confidence level. Formally, the ES measure at time t is defined as

$$ES_{t+1}^p = E(L_{t+1} | L_{t+1} > VaR_{t+1}^p, \Im_t). \tag{4}$$

That is, if losses at time $t+1$ exceed VaR_{t+1}^p, then the expected loss is ES_{t+1}^p. The motivation for changing risk measures is that ES reflects tail risk better than VaR, since ES

captures both the size and likelihood of losses in the tail (e.g., BCBS 2013). Moreover, ES is a coherent risk measure (Artzner et al. 1999) since it satisfies the subadditivity property while VaR does not.[5]

Market liquidity played a large role during the 2008 Subprime Mortgage Crisis. At the height of the crisis, investors took large discounts to sell illiquid assets which further lowered asset prices and caused a liquidity spiral (Brunnermeier and Sannikov (2016)). Basel 3 proposes to account for liquidity risk by scaling the 10-day ES based on an asset's liquidity. Assets are assigned to one of five h_j day liquidity horizons, where $h_1 = 10$, $h_2 = 20$, $h_3 = 40$, $h_4 = 60$, and $h_5 = 120$.[6] Define LH h_j as the portfolio of the subset of assets with a liquidity horizon of h_j days or longer. For example, LH 10 is the portfolio of all assets, LH 20 is the portfolio of assets with a liquidity horizon of 20 days or more, and LH 120 is the portfolio of assets with the longest liquidity horizon of 120 days only. Next, define $ES(j)$ as the ES of portfolio LH h_j. The formula for liquidity-adjusted ES is

$$ES = \sqrt{ES(1)^2 + \sum_{j=2}^{5} \left(\sqrt{\frac{h_j - h_{j-1}}{10}} ES(j) \right)^2}, \tag{5}$$

where h_j is the liquidity horizon, $\sqrt{\frac{h_j - h_{j-1}}{10}}$ is the liquidity scaling based on a normality assumption on asset returns, and $ES(j)$ is the ES for assets with an liquidity horizon of h_j or longer. Hence, $ES(1)$ and $ES(2)$ have no additional liquidity scaling, $ES(3)$ and $ES(4)$ are scaled by $\sqrt{2}$, and $ES(5)$ is scaled by $\sqrt{6}$. The high scaling on illiquid assets such as credit derivatives is meant to reect their additional risk of discounts during financial stress.

A key weakness of Basel 2 was that its risk measures were calibrated to current market conditions, which resulted in undercapitalization and procyclical capital requirements during the crisis. Basel 2.5 introduced Stressed VaR to ensure the capital charge includes periods of significant financial stress in addition to current market conditions, which may be unnecessarily duplicative (BCBS 2013). In Basel 3, the capital charge is instead only based on a Stressed ES, which is ES calibrated to a period of significant financial stress. Since historical data may be unavailable for the full set of risk factors, Basel 3 allows Stressed ES calculations to use a reduced set of risk factors, as long as the reduced set explains at least 75% of the variation in the full set. The formula for the bank's internally modeled capital requirement (IMCC) is

$$IMCC(C) = ES_{R,S} \times \frac{ES_{F,C}}{ES_{R,C}}, \tag{6}$$

where $ES_{R,S}$ is the liquidity-adjusted ES of the reduced set of risk factors calibrated to a period of significant financial stress, $ES_{F,C}$ is the liquidity-adjusted ES of the full set of risk factors calibrated to the current market, $ES_{R,C}$ is the liquidity-adjusted ES of the current reduced set of risk factors, and where the ratio $\frac{ES_{F,C}}{ES_{R,C}}$ is floored at one.

The IMCC calculated above naturally benefits from portfolio diversification. However, during a financial crisis, systemic risk often causes correlations to increase and reduces this

5 A risk measure is subadditive if the risk measure for the sum of two portfolios is no greater than the sum of the risk measures for those portfolios. Subadditivity of risk measures means that diversification may help reduce risks.

6 See Table 2 for some examples of this classification.

effect. To account for this risk, Basel 3 requires banks to calculate partial ES capital requirements, denoted $IMCC(C_i)$ for each regulatory risk class. The undiversified IMCC for risk class i is

$$IMCC(C_i) = ES_{R,S,i} \times \frac{ES_{F,C,i}}{ES_{R,C,i}}, \tag{7}$$

where i refers to interest rate risk, equity risk, foreign exchange risk, commodity risk, or credit spread risk. The stress period used to calculate $ES_{R,S,i}$ is the same as the period used to calculate the portfolio-wide $ES_{R,S}$. The bank's aggregate IMCC is then given by

$$IMCC = \rho(IMCC(C)) + (1-\rho)(\sum_{i=1}^{5} IMCC(C_i)), \tag{8}$$

where $\rho = 0.5$.

The capital charge for modellable risk factors under Basel 3 is

$$CA_t^{B3} = max(IMCC_{t-1}, m_c \times \overline{IMCC}_{t-1}), \tag{9}$$

where $IMCC_{t-1}$ is the previous day's aggregate IMCC, $\overline{IMCC}_{t-1}$ is the average IMCC over the previous 60 days, and m_c is the Basel 3 multiplicative factor. Similar to the previous regimes, Basel 3 backtesting compares VaR with a one-day time horizon at the 99% confidence level to realized exceedances over the previous 250 days. However, Basel 3 essentially halves the Basel 2 multiplicative factor with a minimum value of 1.5 for m_c and a scaling factor between 0 and 0.5 that depends on the model's backtesting performance. The model's multiplication factor is set according to the system of exceedances in Column 4 in Table 1.

1.3 Discussion

Whereas Basel 2.5 introduced only one change to Basel 2 when it comes to calculating capital requirements for market risk, by requiring the use of Stressed VaR, Basel 3 involves several additional changes: (i) the use of ES as the key risk metric instead of VaR, (ii) explicit penalties for exposure to assets with liquidity risk, and (iii) reduction in the possible gains from diversification. Although these changes may appear innocuous they in fact severely complicate the calculations required to conduct proper risk management within banks. First of all, under Basel 3 banks face additional requirements on the data needed to calculate the relevant capital requirement. In particular, long samples of data are needed for all the combinations of risk factors and liquidity horizons toward which the bank is exposed. And if some of these risk factors are less liquidly traded or they do not have sufficient amounts of historical data, mimicking portfolios with equivalent characteristics is needed. Moreover, on each day under Basel 3, there are three liquidity-adjusted ES calculations ($ES_{R,S}$, $ES_{F,C}$, $ES_{R,C}$) and twenty-one possible liquidity horizons across the five risk factors, totaling sixty-three daily ES calculations.

Finally, while ES as a risk measure has several benefits compared to VaR, it is a coherent risk measure that captures both the size and likelihood of losses in the tail, a downside is that it is slightly more difficult to estimate and that historically it has been complicated to backtest this measure as it lacks a property called elicitability (Gneiting 2011). A risk measure is elicitable if there exists a loss function such that the risk measure is the solution to minimizing the expected loss and while VaR is elicitable ES is not so individually.

However, ES is jointly elicitable with VaR as shown by Fissler and Ziegel (2016) and using this result it is possible not only to jointly backtest VaR and ES, see Section 3, but also to model these measures jointly, see Section 2.2.3. An added complication of using 10-day ES is that it cannot be scaled from the one-day ES, and although Basel 3 does allow calculating 10-day ES by using 10-day overlapping periods in many cases this method is not applicable and instead simulation is needed to create these multiperiod forecasts.

So why would regulators consider changing the regulatory capital requirements and, in particular, make these much more complicated to calculate for banks? If the objective is to incentivize banks to use models that best predict such risk measures there now exist several tests that can be used to test the backtesting performance. Models that pass these tests are essentially correctly specified. Moreover, if the question is one of finding the "best" forecasting model for risk measures, this can be examined by using an appropriate and consistent loss function for the risk measures together with Diebold and Mariano (2002) type tests to assess significant difference between two sets of forecasts or the Model Confidence Set (MCS) of Hansen, Lunde, and Nason (2011) to examine which models among a set of models provide superior forecasts. We conduct several such tests and examine which models, among a large class of statistical models, are indeed correctly specified and which provide superior risk measure forecasts.

However, for regulators, it is clearly not sufficient that banks hold enough capital to cover their losses 99% of the time and instead theyrequire banks to keep a buffer of capital that is generally larger than the predicted risk measures by maximizing over current risk and historical averages. Moreover, by using backtesting multipliers banks are incentivized to use more conservative models. We show that though capital requirements have increased, historically Basel regulation has failed to incentivize banks to use correctly specified models and Basel 3 is no exception. Moreover, an additional goal of Basel 3 is to ensure the stability of these capital requirements by calibrating the risk measures to periods of significant financial stress. We show that Basel 3 further dampens the cyclicality in capital requirements by focusing on Stressed ES and by removing most point-in-time calculations. However, the models that minimize capital requirements, though misspecified, are also generally the models that generate the most stable capital requirements for banks seeking to minimize Basel capital variability.

2 Data, Models, and Multiperiod Risk Measures

Basel 3 puts additional requirements on the data needed to calculate regulatory capital, requiring long samples of data for all the combinations of risk factors and liquidity horizons toward which the bank is exposed, on the risk measures that an internal model should be able to produce estimates of, which includes Stressed ES measures, and explicitly disallows simple scaling techniques for generating multiperiod forecasts of risk measures, which in many cases are available only using simulation techniques.

In this section, we first provide an overview of the data used in this article. Next, we introduce the various dynamic models that are used to estimate VaR and ES. Finally, we explain how multiperiod VaR and ES forecasts can be generated using, in most cases, simulation techniques. Readers, who are familiar with all these issues, can skip this section and go straight to our backtesting results in Section 3.

2.1 Data

A key feature distinguishing our article from most of the existing literature is that we consider a realistic portfolio of multiple asset classes that banks likely trade in and hold on their trading book instead of a single asset class like, for example, equities. We use the indexes in Table 2 to proxy for the various risk exposures and liquidity horizons that enter into the calculation of capital requirements under Basel 3. These indexes are highly liquid, widely traded, and constitute a diverse sample that spans all risk factors and relevant liquidity horizons. Our sample is from January 1989 to February 2020, and indexes are included when they become available. All data are obtained from Bloomberg.

For interest rate risk, we use the Bloomberg US Treasury and US Treasury Inflation-Linked indexes for exposure to U.S. government bonds. For equity risk, we include the S&P 500 Index for exposure to large-cap firms and the Russell 2000 Index for exposure to small-cap firms. For equity derivatives risk, we include the CBOE Putwrite Index for exposure to a trading strategy that sells one-month at the money S&P 500 put options and invests the proceeds in one- and three-month Treasury bills. For commodity risk, we include the Bloomberg Commodities Index for exposure to energy, grains, metals, softs, and livestock. For commodity derivatives risk, we include the Bloomberg Commodities Volatility Index. For foreign exchange risk, we include the JP Morgan USD trade-weighted index for exposure to the Australian dollar, British pound, Canadian dollar, Euro, Japanese yen, Swedish krona, and Swiss franc. We also include the Bloomberg Dollar Index for exposure to the Australian dollar, British pound, Canadian dollar, Euro, Japanese yen, Swiss franc, Mexican peso, Chinese renminbi, Korean won, and Indian rupee. For credit risk, we include the Bloomberg US Aggregate, Mortgage-Backed Securities, High Yield, Corporate, and Municipal Indexes for exposures to a large variety of U.S. fixed income securities. For credit derivatives risk, we include the Credit Default Swap Investment Grade and High Yield Indexes for exposure to credit default swaps, which played an important role in the 2008 crisis.

The risk factors we use are similar to those considered in Falato, Iercosan, and Zikes (2019) who use proprietary P&L data reported to the Federal Reserve to show that bank trading desks have exposures to these risk factors. We evaluate a representative bank portfolio that takes an equal-weighted long position in each available index, adjusting the weights as new indexes become available. We choose a simple equal-weighted portfolio, since the start date of the indexes roughly coincides with their importance in the market. Hence, we expect the portfolio to be a good representation of a typical U.S. bank's trading portfolio. Since the index returns are given as simple returns and our models use log returns, we first form the representative portfolio by taking an equal-weighted mean, then transforming the portfolio to log returns to form the representative bank portfolio used to calculate VaR and ES. We perform the same log transformation for portfolios grouped by risk factor (see Section A of the Online Appendix for further details).

Table 3 provides summary statistics for the representative portfolio and individual risk factors. The representative portfolio had a mean daily return of 0.025% mainly driven by the equity, credit, and interest risk factors. Equity had the highest mean return, but also the highest daily volatility of 0.968% corresponding to an annualized volatility of 15%. The representative portfolio is highly skewed toward losses and heavy-tailed, demonstrating that models must accommodate these empirical features to accurately measure tail risk.

Table 2 Sample of indexes used

Risk factor	Index	Symbol	Start date	Liquidity Horizon
Interest rate	Bloomberg US Treasury	LUATTRUU	March 1994	10
Interest rate	Bloomberg US Treasury Inflation-Linked	LBUTTRUU	April 1998	10
Equity	S&P 500	SPX	January 1989	10
Equity	Russel 2000	RTY	January 1989	20
Equity	CBOE Putwrite	PUT	January 1989	60
Commodity	Bloomberg Commodities	BCOM	January 1990	20
Commodity	Bloomberg Commodities Volatility	GSVL1027	February 2001	120
Foreign exch.	JP Morgan USD Trade Weighted	JPMQUSD	January 1990	20
Foreign exch.	Bloomberg Dollar	DXY	January 2005	20
Credit	Bloomberg US Aggregate	LBUSTRUU	January 1989	20
Credit	Bloomberg Mortgage-Backed Securities	LUMSTRUU	January 1991	40
Credit	Bloomberg Corporate	LUACTRUU	August 1998	40
Credit	Bloomberg High Yield	LF98TRUU	August 1998	60
Credit	Bloomberg Municipal	LMBITR	December 2000	60
Credit	Credit Default Swap Investment Grade	CDXIG	April 2007	120
Credit	Credit Default Swap High Yield	CDXHY	April 2007	120

This table shows the list of indexes used to form the representative banking portfolio. The sample is from January 1989 to February 2020. Indexes are sorted by risk factor first and then Basel 3 liquidity horizon category. We also provide the symbol and starting month for each index.

Table 3 Summary statistics for portfolios

Portfolios	Representative	Interest	Equity	Commodity	Foreign exchange	Credit
Mean	0.025	0.020	0.037	0.003	0.005	0.025
Standard deviation	0.276	0.290	0.968	0.616	0.332	0.190
Skewness	−1.016	−0.207	−0.677	−0.661	0.215	−0.815
Kurtosis	9.692	3.790	9.603	10.168	6.377	9.674
0.025	−0.567	−0.583	−2.033	−1.291	−0.648	−0.366
0.01	−0.777	−0.759	−2.865	−1.697	−0.840	−0.516
Min	−3.728	−2.170	−9.747	−9.171	−2.476	−2.118

This table shows summary statistics for the representative banking portfolio and by risk factor. The sample is from January 1989 to February 2020. The representative portfolio is formed by taking the equal-weighted mean return of all available indexes (denoted r_{Bank}) in Table 2, then taking the log transformation $x_{Bank} = \log(1 + r_{Bank})$. Risk factor portfolios are analogously formed by taking the equal-weighted mean return of indexes in the risk factor, then taking the log transformation. Summary statistics are calculated based on each portfolio's daily log return and include the sample mean, standard deviation, skewness, kurtosis, 0.025 quantile, 0.01 quantile, and min in percentage terms.

The 0.025 and 0.01 quantiles and minimum return for the representative portfolio are roughly 2, 3, and 13 standard deviations from the mean, highlighting the non-normality of the distribution's loss tail.

2.2 Dynamic Models for Returns

We assume throughout that losses are governed by a dynamic model given by

$$L_t = \mu_t + \sigma_t \epsilon_t, \quad t = 1, \ldots, T, \tag{10}$$

where μ_t is the conditional mean, σ_t is the conditional volatility, and ϵ_t are independent and identically distributed (i.i.d.) innovations with distribution $G(0, 1)$.

Given the dynamic model in Equation (10), VaR can be expressed as

$$VaR^p_{T+1} = \mu_{T+1} + \sigma_{T+1} G^{-1}_{1-p} \equiv \mu_{T+1} + \sigma_{T+1} c_{1,p}, \tag{11}$$

where G^{-1}_{1-p} denotes the $(1-p)$th quantile of G. For example, if G is the standard normal distribution and $p = 0.01$, then $G^{-1}_{0.99} = \Phi^{-1}_{0.99} = 2.33$, where Φ denotes the standard normal distribution function, and hence $VaR^p_{T+1} = \mu_{T+1} + 2.33\sigma_{T+1}$. Similarly, ES can be expressed as

$$ES^p_{T+1} = \mu_{T+1} + \sigma_{T+1} E(\epsilon_{T+1} | \epsilon_{T+1} > G^{-1}_{1-p}) \equiv \mu_{T+1} + \sigma_{T+1} c_{2,p}. \tag{12}$$

For example, if $\epsilon_t \sim N(0,1)$ and $p = 0.01$, it can be shown that $E(\epsilon_{T+1} | \epsilon_{T+1} > \Phi^{-1}_{0.99}) = \phi(\Phi^{-1}_{0.99})/0.01 = 2.67$, where ϕ denotes the standard normal density function, and hence $ES^p_{T+1} = \mu_{T+1} + 2.67\sigma_{T+1}$. When the innovation distribution is non-normal we can still express VaR and ES as in Equations (11) and (12), though the values of $c_{1,p}$ and $c_{2,p}$ will depend on the distribution G.

In this article, we consider several techniques for estimating the upper tail of the innovation distribution to find the tail risk measures. We first introduce two *ad hoc* models that involve no estimation including the most popular model used in banks called HS. Next, we consider a popular class of models in which the dynamics are parameterized using GARCH processes with parameters fitted to the entire conditional distribution. Finally, we describe a new class of models with conditional dynamics based on generalized autoregressive score (GAS) type models where parameters are instead fitted directly to a relevant loss metric for the risk measures considered.

2.2.1 *Ad hoc* models

The simplest and most popular model for estimating VaR and ES is undoubtedly HS due to ease of implementation.[7] This nonparametric and distribution-free model calculates VaR and ES using the empirical distribution of past losses. The HS estimate for VaR^p_{T+1} is

$$\text{HS} - VaR^p_{T+1} = Q_{1-p}(\{L_t\}), \tag{13}$$

7 Pérignon and Smith (2010) report that 73% of international banks use HS and Mehta et al. (2012) reports that 75% of large banks use only HS.

where $Q_{1-p}(\{L_t\})$ denotes the $(1-p)$th empirical quantile of losses $\{L_t\}_{t=1}^{T}$. The HS estimate for ES_{T+1}^{p} is

$$\text{HS} - ES_{T+1}^{p} = \frac{1}{\left(L_t > HS - VaR_{T+1}^{p}\right)}\left(\sum_{L_t > HS - VaR_{T+1}^{p}} L_t\right), \tag{14}$$

where $(L_t > HS - VaR_{T+1}^{p})$ denotes the number of losses $\{L_t\}_{t=1}^{T}$ exceeding $HS - VaR_{T+1}^{p}$. While HS can capture the nonnormality commonly observed in financial returns it cannot account for the conditional dynamics in Equation (10).

The HS model is a "fully" nonparametric model since no assumptions are made about neither the dynamics nor the conditional distribution. Parametric models, on the contrary, use explicit formulas for the dynamics together with a parameterized distribution to calculate the values of $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$. The simplest parametric model is the RM model developed by Morgan (Morgan 1996). RM assumes losses are normally distributed, that $\mu_t = 0$, and that the conditional variance follows

$$\sigma_t^2 = 0.06(L_{t-1})^2 + 0.94\sigma_{t-1}^2, \tag{15}$$

and thus no estimation is required for this model either. The RM estimate for VaR_{T+1}^{p} is

$$\text{RM} - VaR_{T+1}^{p} = \sigma_{T+1} c_{1,p}^{Norm}, \tag{16}$$

where $c_{1,p}^{Norm} = \Phi_{1-p}^{-1}$. The RM estimate for ES_{T+1}^{p} is

$$\text{RM} - ES_{T+1}^{p} = \sigma_{T+1} c_{2,p}^{Norm}, \tag{17}$$

where $c_{2,p}^{Norm} = \phi(c_{1,p}^{Norm})/p$.

2.2.2 Dynamic location-scale models

The most popular approach for specifying the dynamic model in Equation (10) in a flexible manner is to let μ_t follow some Autoregressive–Moving-Average (ARMA)process and to let σ_t^2 follow a GARCH process. In this article, we will assume that the conditional mean is constant ($\mu_t = \mu$) and that the conditional variance follows a GARCH(1,1) model given by

$$\sigma_t^2 = \omega + \alpha(L_{t-1} - \mu_{t-1})^2 + \beta\sigma_{t-1}^2, \tag{18}$$

where $\alpha + \beta < 1$ to ensure stationarity.[8] The RM model is a special case of this framework which sets $\omega = 0$, $\alpha = 0.06$, and $\beta = 0.94$ in Equation (18).[9]

A first model that corrects the shortcomings of the HS model above is the FHS model which computes $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$ from the empirical distribution of centered innovations $\hat{\epsilon}_t - \bar{\hat{\epsilon}}$. Thus, this model uses the conditional dynamics without the need for distributional assumptions on the empirical innovations.[10] FHS was first proposed by Barone-Adesi, Bourgoin,

8 Naturally, our approach generalizes to more complex specifications of the conditional mean and variance.

9 Since $\alpha + \beta = 1$, the RM model follows a IGARCH random walk process and is therefore not stationary.

10 The parameters of ARMA-GARCH type models can be estimated consistently using Quasi-Maximum Likelihood estimation with a Gaussian likelihood even if the underlying distribution is

and Giannopoulos (1998), Hull and White (1998), and Diebold, Schuermann, and Stroughair (2000). The FHS estimate of $c_{1,p}$ is

$$\hat{c}_{1,p}^{FHS} = Q_{1-p}\left(\{\hat{\epsilon}_t - \bar{\hat{\epsilon}}\}\right), \tag{19}$$

and the FHS estimate of $c_{2,p}$ is

$$\hat{c}_{2,p}^{FHS} = \frac{1}{\left(\hat{\epsilon}_t - \bar{\hat{\epsilon}} > c_{1,p}^{FHS}\right)}\left(\sum_{\hat{\epsilon}_t - \bar{\hat{\epsilon}} > c_{1,p}^{FHS}} \left(\hat{\epsilon}_t - \bar{\hat{\epsilon}}\right)\right). \tag{20}$$

The FHS estimates for VaR_{T+1}^{p} and ES_{T+1}^{p} are then obtained by substituting these estimates into Equations (11) and (12), respectively.

2.2.2.1 Parametric models

Whereas FHS makes no assumptions about the conditional distribution, the RM model assumes losses are normally distributed. Other typical choices in the literature include the Student's t-distribution (STD), the Hansen (1994) Skewed Student's t-distribution (SSTD), or the generalized error distribution (GED) which are popular because they can capture heavy tails exhibited by financial returns. We consider the flexible skewed generalized t-distribution (SGT) of Theodossiou (1998), which nests many of the popular parametric distributional assumptions for modeling financial returns. The probability density function of the SGT distribution is given by

$$f(x|\lambda,n,k) = \frac{k}{2\phi}n^{-\frac{1}{k}}B\left(\frac{1}{k},\frac{n}{k}\right)^{-1}\left(1+\frac{1}{n}\frac{|x-m|^k}{(1+sgn(x-m)\lambda)^k\phi^k}\right)^{-\frac{n+1}{k}}, \tag{21}$$

where m is the mode, ϕ is a scaling constant, $-1 < \lambda < 1$ is a skewness parameter, k and n are positive tail parameters, sgn is the sign function, and B is the Beta function. For a standardized SGT random variable with mean zero and unit variance, the mode is $m = -2\lambda G_1\phi$ and the scaling constant is $\phi = \left((1+3\lambda^2)G_2 - 4\lambda^2 G_1^2\right)^{-\frac{1}{2}}$, where G_1 and G_2 are given by $G_j = n^{\frac{j}{k}}B\left(\frac{j+1}{k},\frac{n-j}{k}\right)B\left(\frac{1}{k},\frac{n}{k}\right)^{-1}$ for $j = 1, 2$.

Theodossiou (2018) shows that the closed-form expression for the $(1-p)$th quantile of the SGT distribution is

$$c_{1,p}^{SGT} = m + (1+\lambda)\phi n^{\frac{1}{k}}t_p^{\frac{1}{k}}(1-t_p)^{-\frac{1}{k}}, \tag{22}$$

where $t_p = IB^{-1}\left(\frac{2|(1-p)-(1-\lambda)/2|}{(1+\lambda)};\frac{1}{k},\frac{n}{k}\right)$ and IB^{-1} is the inverse incomplete Beta function ratio, and that

$$c_{2,p}^{SGT} = m + \frac{(1+\lambda)^2}{2p}\left[1 - IB\left(t_p;\frac{2}{k},\frac{n-1}{k}\right)\right]G_1\phi, \tag{23}$$

where IB is the incomplete Beta function ratio. We calculate the values for VaR_{T+1}^{p} and ES_{T+1}^{p} in Equations (11) and (12) under an SGT distributional assumption on the innovations by using estimated values of λ, n, and k to obtain $\hat{c}_{1,p}^{SGT}$ and $\hat{c}_{2,p}^{SGT}$ in Equations (22) and (23), respectively.

non-Gaussian assuming the correct order of the dynamic processes is specified, see, for example, Bollerslev and Wooldridge (1992) and Gourieroux (1997).

The SGT distribution is extremely flexible and nests all the parametric distributions used in this article. For example, the SGT distribution is equivalent to the normal distribution when $\lambda = 0$, $k = 2$, and $n \to \infty$ and an SGT distribution with $\lambda = 0$, $k = 2$, and $n = d$ is equivalent to an STD with d degrees of freedom. When the skewness parameter $-1 < \lambda < 1$, the SGT distribution with $k = 2$, and $n = d$ is equivalent to Hansen's skewed t-distribution with d degrees of freedom and the same skewness parameter λ. Finally, the SGT distribution with $n \to \infty$ is equivalent to the (skewed when $\lambda \neq 0$) GED distribution with shape parameter $\nu = k$.

2.2.2.2 Methods based on extreme value theory

Whereas the parametric model above imposes assumptions on the entire distribution of the innovations, EVT models the behavior of the distribution's tail. Tail values are described by the conditional excess distribution function defined by

$$F_\eta(y) = Pr\{X - \eta \leq y | X > \eta\} = \frac{F(y+\eta) - F(\eta)}{1 - F(\eta)}, \quad y > 0, \tag{24}$$

which is the probability that x exceeds threshold η by at most y given x exceeds the threshold.[11]

2.2.2.2.1 Generalized Pareto distribution estimator

Balkema and de Haan (1974) show that, for a sufficiently high threshold η, the cumulative distribution function in Equation (24) can be approximated by the Generalized Pareto Distribution (GPD) given by

$$G(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-y/\sigma), & \text{if } \xi = 0, \end{cases} \tag{25}$$

where ξ is a shape parameter and $\sigma > 0$ is a scale parameter defined for $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ when $\xi < 0$. When $\xi > 0$, the distribution becomes the heavy-tailed Pareto distribution. The probability density function of the GPD is given by

$$g(z_t; \xi, \sigma) = \begin{cases} \frac{1}{\sigma}\left[1 + \frac{\xi z_t}{\sigma}\right]^{-(1+1/\xi)}, & \text{if } \xi \neq 0 \\ \frac{1}{\sigma}\exp[-\frac{z_t}{\sigma}], & \text{if } \xi = 0, \end{cases} \tag{26}$$

where the exceedances $\{z_1, \ldots, z_{N_\eta}\}$ are defined as $z_t = \epsilon_t - \eta$ for $1 \leq t \leq N_\eta$, and N_η is the number of exceedances above the threshold η. Maximizing the likelihood function given by

$$L(\xi, \sigma) = \prod_{i=1}^{N_\eta} \frac{1}{N_\eta} g(z_t; \xi, \sigma), \tag{27}$$

yields estimates of ξ and σ.

11 See Christoffersen (2011) for a highly accessible introduction to the use of EVT in risk management.

Using the estimated parameters $\hat{\xi}$ and $\hat{\sigma}$, Mcneil and Frey (2000) show that the closed-form expression for the $(1-p)$th quantile of the GPD distribution is

$$\hat{c}_{1,p}^{GPD} = \eta + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{Tp}{N_\eta}\right)^{-\hat{\xi}} - 1\right), \tag{28}$$

where T is the sample size, and that

$$\hat{c}_{2,p}^{GPD} = \hat{c}_{1,p}^{GPD}\left(\frac{1}{1-\hat{\xi}} + \frac{\hat{\sigma} - \hat{\xi}\eta}{(1-\hat{\xi})\hat{c}_{1,p}^{GPD}}\right). \tag{29}$$

Hence, given a GPD assumption on exceedances, the values for VaR_{T+1}^{p} and ES_{T+1}^{p} are obtained by substituting these estimates into Equations (11) and (12), respectively.

2.2.2.2.2 Hill estimator

Hill (1975) provides an alternative estimation method used in EVT. The Hill estimator assumes that $\xi > 0$ and the distribution has heavy tails. Suppose the tail of the conditional distribution of innovations is approximated by the distribution function

$$F(z) = 1 - L(z)z^{-1/\xi} \approx 1 - cz^{-1/\xi}, \tag{30}$$

whenever $\epsilon_t > u$, where u is the threshold, and $L(z)$ is a slowly varying function, which we approximate with a constant c. Let k be the number of observations that exceed u. The Hill estimator $\hat{\xi}$ is the maximum likelihood estimator of ξ assuming innovations are i.i.d. from an unknown distribution given in closed form by

$$\hat{\xi} = \frac{1}{k}\sum_{t=1}^{k} \ln(\hat{\epsilon}_{(T-t+1)}) - \ln(u), \tag{31}$$

where $\hat{\epsilon}_{(t)}$ denotes the t-th order statistic of $\hat{\epsilon}_t$ such that $\hat{\epsilon}_{(t)} \geq \hat{\epsilon}_{(t-1)}$ for $t = 2, \ldots, T$.

Huisman et al. (2001) provide an alternative estimator of ξ that does not require choosing a threshold u. They show that the bias in the Hill estimator is a linear and increasing function of k. Hence, a threshold-free estimate of ξ is the intercept β_0 in the regression

$$\hat{\xi}_k = \beta_0 + \beta_1 k + \nu(k), k = 1, .., K, \tag{32}$$

where $\hat{\xi}_k$ is the Hill estimator in Equation (31) with threshold $u = \hat{\epsilon}_{(k)}$. Since the variance of $\hat{\xi}_k$ depends on k, $\nu(k)$ is heteroskedastic. To correct for this, they estimate Equation (32) using weighted least squares with a $(K \times K)$ weighting matrix W, which has $\{\sqrt{1}, .., \sqrt{K}\}$ as diagonal elements and zeros elsewhere. We set $K = T/4$.

Given $\hat{\xi}$ we approximate the tail distribution F by setting $c = \frac{k}{T}u^{1/\hat{\xi}}$, which is derived from the condition $1 - F(u) = \frac{k}{T}$. The estimate of F is

$$\hat{F}(z) = 1 - \frac{k}{T}\left(\frac{z}{u}\right)^{-1/\hat{\xi}}. \tag{33}$$

Christoffersen and Gonçalves (2005) show that the $(1-p)$th quantile of $\hat{F}(z)$ is

$$\hat{c}_{1,p}^{Hill} = u\left(\frac{pT}{k}\right)^{-\hat{\xi}}, \tag{34}$$

and that

$$\hat{c}_{2,p}^{Hill} = \frac{\hat{c}_{1,p}^{Hill}}{1-\hat{\xi}}. \tag{35}$$

We substitute these estimates into Equations (11) and (12) to obtain the corresponding estimates of VaR_{T+1}^{p} and ES_{T+1}^{p}.

In Section B of the Online Appendix, we conduct a simulation study to estimate the optimal threshold, $\hat{\eta}$, for GPD estimation and, $\hat{u}$, for Hill estimation at the 99%, 97.5%, and 95% confidence levels. We find that GPD estimates are optimized by setting $\hat{\eta}$ equal to the 0.85 quantile of innovations for all three confidence levels and that Hill estimates are optimized by setting $\hat{u}$ equal to the same quantile as the confidence level, and we use these thresholds in this article.

2.2.7 Dynamic VaR and ES models

When estimating the parameters of the location-scale models above this is typically done by fitting to the whole conditional distribution of asset returns using Maximum Likelihood type methods. Such methods are efficient if the dynamics and the distributional specification are correct and in this case, the model will also yield optimal forecasts of the risk measures. However, if we think of the dynamic models only as approximations it is not obvious that the estimated model is optimal when the application is to forecast VaR and ES. Indeed, in this situation, it may be possible to improve the forecasted risk measures by estimating model parameters using an alternative metric, one that is consistent for VaR and ES, rather than using a (Quasi) Maximum Likelihood approach which focuses on the conditional mean and variance.

An appropriate metric for this problem was provided by Fissler and Ziegel (2016), who showed that the class of FZ loss functions given by

$$\begin{aligned} L_{FZ}(Y,v,e;p,G_1,G_2) = (1_{Y\le v}-p)(G_1(v)-G_1(Y)+\frac{1}{p}G_2(e)v) - G_2(e)(\frac{1}{p}\ 1_{Y\le v}Y - e) \\ - \mathcal{G}_2(e), \end{aligned} \tag{36}$$

where Y denotes the return, $-v$ is the VaR, $-e$ is the ES, G_1 is weakly increasing, G_2 is strictly increasing and strictly positive, and $\mathcal{G}_2' = G_2$, is consistent for VaR and ES. In other words, minimizing the expected FZ loss returns the true VaR and ES and

$$(-VaR_t, -ES_t) = \underset{(v,e)}{\operatorname{argmin}}\, E_{t-1}[L_{FZ}(Y_t,v,e;p,G_1,G_2)]. \tag{37}$$

To implement this approach for estimation one needs to choose the functions G_1 and G_2. Patton, Ziegel, and Chen (2019) suggest to set $G_1 = 0$ and $G_2 = -1/x$ to obtain

$$L_{FZ0}(Y,v,e;p) = -\frac{1}{pe}1_{Y\le v}(v-Y)+\frac{v}{e}+\log(-e)-1, \tag{38}$$

which they refer to as the FZ0 loss function, and they provide asymptotic theory for estimating VaR and ES models by minimizing this loss. Thus, one can now use this criterion to estimate parameters of any dynamic specifications like, for example, an ARMA-GARCH type model.

While GARCH type dynamics could be considered, Patton, Ziegel, and Chen (2019) specify instead a dynamic model for VaR and ES using the GAS model of Creal, Koopman, and Lucas (2013) and Harvey (2013) where the forcing variable is a function of the derivative and the Hessian of the FZ0 loss function instead of a log-likelihood. We first consider their one-factor GAS model for VaR and ES, where the risk measures, $v_t = a\exp(\kappa_t)$ and $e_t = b\exp(\kappa_t)$ with $b < a < 0$, are driven by a factor $\kappa_t = \log(\sigma_t)$, interpreted as the log volatility, with dynamics given by

$$\kappa_t = \omega + \beta\kappa_{t-1} + \gamma\frac{1}{b\exp(\kappa_{t-1})}\left(\frac{1}{p}\ 1_{Y_{t-1}\le a\exp(\kappa_{t-1})}Y_{t-1} - b\exp(\kappa_{t-1})\right). \tag{39}$$

The parameters (a, b, β, γ) can be estimated using the FZ0 loss function while setting $\omega = 0$ ensures identification, see Patton, Ziegel, and Chen (2019) for details. We also consider their Hybrid GAS/GARCH model with

$$\kappa_t = \omega + \beta\kappa_{t-1} + \gamma\frac{1}{b\exp(\kappa_{t-1})}\left(\frac{1}{p}\ 1_{Y_{t-1}\le a\exp(\kappa_{t-1})}Y_{t-1} - b\exp(\kappa_{t-1})\right) + \delta\log|Y_{t-1}|, \tag{40}$$

where parameters $(a, b, \beta, \gamma, \delta)$ are estimated using the FZ0 loss function and setting $\omega = 0$ again ensures identification.[12] We refer to these two models as the FZ1 and FZH models, respectively. The one-day VaR and ES forecasts are given by

$$\text{FZ} - VaR^p_{T+1} = -v_{T+1} = -a\exp(\kappa_{T+1}), \tag{41}$$

and

$$\text{FZ} - ES^p_{T+1} = -e_{T+1} = -b\exp(\kappa_{T+1}), \tag{42}$$

where κ_{T+1} follows Equation (39) in the FZ1 model and Equation (40) in the FZH.

2.3 Multiperiod VaR and ES Forecasts

Basel 2 and 2.5 capital charges are based on VaR with a 10-day time horizon whereas Basel 3 capital charges are based on ES with a 10 day time horizon. During Basel 2, banks would often approximate 10-day VaR by multiplying 1-day VaR by $\sqrt{10}$, which is correct only when returns are normally distributed. Due to the non-normality of returns, this scaling method is explicitly prohibited in Basel 3 (BCBS 2019). Basel 3 regulation though, does allow 10-day forecasts to be calculated using overlapping observations, which is how we calculate HS forecasts. Define the sum of 10-day losses conditional on time t as

$$L_t[10] = \sum_{k=t+1}^{t+10} L_k. \tag{43}$$

12 Patton, Ziegel, and Chen (2019) also consider a two-factor model for VaR and ES. However, they found the one-factor model performed better than the two-factor model, so we exclude it from our analysis.

The goal of multiperiod forecasts is to calculate the VaR and ES of $L_T[10]$ given the information at time T. The HS estimate for VaR^p_{T+10} is

$$\text{HS} - VaR^p_{T+10} = Q_{1-p}(\{L_t[10]\}), \tag{44}$$

where $Q_{1-p}(\{L_t[10]\})$ denotes the $(1-p)$th empirical quantile of the 10-day losses $\{L_t[10]\}_{t=1}^{T-10}$. The HS estimate for ES^p_{T+10} is

$$\text{HS} - ES^p_{T+10} = \frac{1}{\left(L_t[10] > HS - VaR^p_{T+10}\right)}\left(\sum_{L_t[10]>HS-VaR^p_{T+10}} L_t[10]\right), \tag{45}$$

where $(L_t[10] > HS - VaR^p_{T+10})$ denotes the number of 10-day losses exceeding $HS - VaR^p_{T+10}$.

For GARCH models, the volatility forecast and conditional distribution of 10-day losses are only available in closed form for normally distributed innovations. The 10-day GARCH variance forecast for normal innovations is

$$\sigma^2_{T+10} = \frac{\omega}{1-\gamma}\left(10 - \frac{1-\gamma^{10}}{1-\gamma}\right) + \frac{1-\gamma^{10}}{1-\gamma}\sigma^2_{T+1}, \tag{46}$$

where $\gamma = \alpha + \beta < 1$ (Tsay 2010). The normal estimate for VaR^p_{T+10} is

$$\text{Norm} - VaR^p_{T+10} = 10\mu + \sigma_{T+10}c^{Norm}_{1,p}, \tag{47}$$

and the estimate for ES^p_{T+10} is

$$\text{Norm} - ES^p_{T+10} = 10\mu + \sigma_{T+10}c^{Norm}_{2,p}. \tag{48}$$

Taking the limit as $\gamma \to 1$ the 10-day RM variance forecast is seen to be $\sigma^2_{T+10} = 10\sigma^2_{T+1}$. Hence, the RM estimate for VaR^p_{T+10} and ES^p_{T+10} is $\sqrt{10}\text{RM} - VaR^p_{T+1}$ and $\sqrt{10}\text{RM} - ES^p_{T+1}$, respectively. This relationship is referred to as the square root of time rule under RM.

In all other cases, multiperiod risk measures are obtained through simulation. We approximate the conditional distribution of $L_T[10] = L_{T+1} + .. + L_{T+10}$ by simulating future paths of losses

$$L_{T+k} = \mu + \sigma_{T+k}\epsilon_{T+k}, \tag{49}$$

where the conditional variance in GARCH models follows

$$\sigma^2_{T+k} = \omega + \alpha(L_{T+k-1} - \mu)^2 + \beta\sigma^2_{T+k-1}, \tag{50}$$

for $\sigma^2_{T+1}, .., \sigma^2_{T+10}$. We perform B = 2000 simulations, resulting in a sample $\{L_{b,T}[10]\}_{b=1}^{B}$ of 10-day losses indexed by b. The GARCH estimate for VaR^p_{T+10} is then

$$VaR^p_{T+10} = Q_{1-p}(\{L_{b,T}[10]\}), \tag{51}$$

$Q_{1-p}(\{L_{b,T}[10]\})$ where denotes the $(1-p)$th empirical quantile of the simulated 10-day losses $\{L_{b,T}[10]\}_{b=1}^{B}$. The GARCH estimate for ES^p_{T+10} is

$$ES^p_{T+10} = \frac{1}{\left(L_{b,T}[10] > VaR^p_{T+10}\right)}\left(\sum_{L_{b,T}[10]>VaR^p_{T+10}} L_{b,T}[10]\right), \tag{52}$$

where $(L_{b,T}[10] > HS - VaR^{p}_{T+10})$ denotes the number of simulated 10-day losses $\{L_{b,T}[10]\}_{b=1}^{B}$ exceeding VaR^{p}_{T+10}. The simulated innovations $\{\epsilon_{T+1}, .., \epsilon_{T+10}\}$ in Equation (49) are drawn directly from the estimated distribution, or in the case of the FHS model drawn with replacement from the empirical distribution of centered innovations $\hat{\epsilon}_t - \bar{\hat{\epsilon}}$. A similar approach is used for the FZ models with the only change that here the innovations are given by $\eta_t = L_t / \exp(\kappa_t)$.

For the EVT models, we follow the procedure in Mcneil and Frey (2000). For the GPD model, we estimate the positive and negative threshold $\eta^{\pm}$ as the 0.15 and 0.85 quantiles of $\hat{\epsilon}_t$, respectively. Using the thresholds, we estimate the positive and negative shape $\xi^{\pm}$ and scale $\sigma^{\pm}$ parameters. We draw innovations from the empirical distribution of innovations $\hat{\epsilon}_t$. If the drawn innovation is greater than η^{+}, we replace the innovation with $\eta^{+} + y^{+}$, where y^{+} is drawn from a GPD distribution with shape ξ^{+} and scale σ^{+}. Similarly, if the drawn innovation is less than η^{-}, we replace the innovation with $\eta^{-} - y^{-}$, where y^{-} is GPD distributed with shape ξ^{-} and scale σ^{-}. If the drawn innovation is between η^{-} and η^{+}, we use the innovation itself. We take a similar approach for the Hill model, estimating the positive and negative threshold $u^{\pm}$ as the p and $(1-p)$th quantiles of the innovations $\hat{\epsilon}_t$, respectively. Using the thresholds, we estimate the positive and negative shape $\xi^{\pm}$ parameter. If the drawn innovation is greater than u^{+}, we replace the innovation with $u^{+} + z^{+}$, where z^{+} is drawn from a Pareto distribution with a shape parameter ξ^{+}. If the drawn innovation is less than u^{-}, we replace the innovation with $u^{-} + z^{-}$, where z^{-} is drawn from a Pareto distribution with shape parameter ξ^{-}. If the drawn innovation is between u^{-} and u^{+}, we use the innovation itself.

3. Model Backtesting

Basel 3 changes the regulatory market risk measure from VaR at the 99% confidence level to ES at the 97.5% confidence level. With this change, banks are strongly motivated to identify which models are correctly specified under the new risk measure and regulatory supervisors are particularly interested in identifying the set of models that underestimate risk, since banks using these models may be undercapitalized prior to a financial shock. When Basel 3 was introduced, though, a major criticism against using ES for regulation was the lack of available backtesting due to the risk measure not being elicitable (Gneiting 2011).[13] However, Fissler, Ziegel, and Gneiting (2015) showed that ES is in fact jointly elicitable with VaR, and a recent literature on joint VaR and ES backtesting has been developed by academics. This section uses these recently developed and state-of-the-art methods, which we simply refer to as ES backtests, to examine which of the models are correctly specified and provide the best risk forecasts.

Since VaR is still used for regulatory backtesting and for setting the capital multiplier, we first consider individual backtests for VaR. The methods we use to backtest predicted VaR are standard and can be found in, for example, Christoffersen (2009). Specifically, we report the actual exceedances (Actual) and p-values from Kupiec (1995)'s Unconditional Coverage (UC) test, Christoffersen (1998)'s Conditional Coverage (CC) test, Christoffersen

13 A variable is elicitable if it can be defined as the minimizer of a mean scoring function. Gneiting (2011) shows that ES lacks elicitability while VaR is elicitable allowing backtesting VaR but not ES individually.

and Pelletier (2004)'s Duration (Dur) test, and Engle and Manganelli (2004)'s Dynamic Quantile (DQ) test. In terms of ES backtests, Nolde and Ziegel (2017) suggest that these can be separated into two categories: (i) traditional backtests which can be used to determine the correctly specified models, and (ii) comparative backtests which can be used to select the models that provide superior forecasts. For the traditional backtests, we next report p-values for Mcneil and Frey (2000)'s residual test (ER), Bayer and Dimitriadis (2018)'s strict, auxiliary, and one-sided regression tests (ESR), Nolde and Ziegel (2017)'s conditional calibration test (CCa), and Gordy and Mcneil (2020)'s spectral backtests. Finally, for the comparative backtests, we report average FZ0 losses with Hansen, Lunde, and Nason (2011)'s MCS, Diebold and Mariano (2002) t-statistics from Patton (2019)'s FZ0 backtest, and Ziegel et al. (2017)'s Murphy Diagrams.[14]

We empirically evaluate the various models' performance in terms of estimating one-day VaR and ES for our representative portfolio from January 1997 to February 2020.[15] For HS and RM, we use a 250-day rolling estimation window. While most banks use short windows for their HS estimation, dynamic models generally require a larger estimation window to reduce estimation error. We, therefore, choose to report results using a rolling estimation window of $T = 2000$ days and include data from 1989 for estimation purposes. In addition to results at the 99% and 97.5% confidence levels, we also consider a 95% confidence level. The 95% confidence level is not currently used for regulatory purposes. However, due to the larger tail sample, estimation error could be reduced across models at this level and hence this confidence level could be used for future regulation. For visual clarity, hypothesis tests that are rejected with 95% confidence are bold. Previewing our results, we reject that the HS, RM, Normal, STD, and GED models are correctly specified for VaR and ES backtests at traditional levels and find that these models provide poor forecasts. We also reject the FZ models in the conditional spectral backtests and find that these models provide inferior forecasts. We cannot reject the hypothesis of correct specification for most of the skewed GARCH models and find that these models also provide superior VaR and ES forecasts.

3.1 VaR Backtests

Panel A of Table 4 reports one-day VaR backtesting results at the 99% confidence level for which 57 exceedances are expected. The UC and CC p-values indicate we can reject the hypothesis that the HS, RM, Normal, STD, and GED models are correctly specified. These models have too many exceedances from underestimating VaR and are likely to have serially correlated exceedances. The Dur p-values indicate we can reject the hypothesis that HS has the correct duration between exceedances, likely because the model is misspecified against volatility clustering. Additionally, the DQ p-values indicate that the HS, RM, Normal, STD, and FZ1 models are misspecified. Panel B of Table 4 reports results at the 97.5% confidence level for which 144 exceedances are expected. The UC and CC p-values indicate we can reject the same set of models as with the 99% confidence level. The Dur p-

14 Further details on all these ES backtests can be found in Section C of the Online Appendix.

15 Although Basel sets capital charges based on 10-day risk estimates, multi-horizon backtests are challenging to conduct due to overlapping observations and changing portfolio compositions. We follow the existing literature and conduct backtests at a 1-day horizon which is also the horizon used for Basel backtesting.

Table 4 VaR backtests

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: Results at a 99% confidence level (expected exceedances = 57)														
Actual	80	116	59	105	80	64	78	65	65	52	59	59	53	54
UC	**0.01**	**0.00**	0.87	**0.00**	**0.01**	0.42	**0.01**	0.35	0.35	0.44	0.87	0.87	0.52	0.61
CC	**0.00**	**0.00**	0.88	**0.00**	**0.00**	0.68	**0.01**	0.62	0.62	0.59	0.88	0.88	0.66	0.73
Dur	**0.00**	0.17	0.51	0.09	0.08	0.49	0.23	0.39	0.33	0.61	0.51	0.51	0.08	0.45
DQ	**0.00**	**0.00**	0.33	**0.00**	**0.00**	0.45	0.07	0.20	0.20	0.32	0.33	0.33	**0.00**	0.49
Panel B: Results at a 97.5% confidence level (Expected exceedances = 144)														
Actual	178	200	141	204	197	149	180	141	151	131	141	141	165	147
UC	**0.01**	**0.00**	0.77	**0.00**	**0.00**	0.70	**0.00**	0.77	0.59	0.25	0.77	0.77	0.09	0.83
CC	**0.00**	**0.00**	0.21	**0.00**	**0.00**	0.31	**0.00**	0.21	0.31	0.06	0.21	0.21	0.05	0.06
Dur	**0.00**	**0.03**	0.05	0.07	**0.04**	0.05	**0.03**	0.07	0.08	**0.02**	0.05	0.05	**0.03**	0.07
DQ	**0.00**	**0.00**	0.24	**0.00**	**0.00**	0.12	**0.00**	0.12	0.16	0.11	0.24	0.24	**0.01**	**0.02**
Panel C: Results at a 95% confidence level (expected exceedances = 288)														
Actual	314	312	297	333	357	310	331	292	314	299	297	297	329	309
UC	0.14	0.17	0.63	**0.01**	**0.00**	0.21	**0.01**	0.85	0.14	0.55	0.63	0.63	**0.02**	0.23
CC	**0.00**	0.12	0.30	**0.02**	**0.00**	0.38	**0.04**	0.53	0.25	0.31	0.30	0.30	0.06	0.33
Dur	**0.00**	**0.02**	0.28	0.42	0.27	0.58	0.54	0.32	0.31	0.22	0.28	0.28	0.89	0.78
DQ	**0.00**	**0.00**	**0.02**	**0.00**	**0.00**	**0.04**	**0.00**	**0.02**	**0.01**	**0.01**	**0.02**	**0.02**	0.13	0.58

This table shows the VaR backtesting results of the representative portfolio from January 1997 to February 2020. Each panel reports results for a particular confidence level. In each panel, Row 1 shows the actual number of exceedances. Rows 2–5 display two-sided p-values for the UC, CC, Duration, and DQ backtests. Models with p-values <0.05 are in bold.

values indicate we can reject the hypothesis that the HS, RM, STD, GED, GPD, and FZ1 models have independent durations between exceedances. Additionally, the DQ p-values indicate that the HS, RM, Normal, STD, FZ1, and now also the FZH models are misspecified. We cannot reject the hypothesis of correct specification for the FHS, SSTD, SGED, SGT, Hill, and HillH models for either confidence level.

Next, Panel C of Table 4 reports the results at the 95% confidence level for which 288 exceedances are expected. The UC, CC, and Dur test results are similar to the previous confidence levels. Interestingly, the DQ p-values indicate we can reject the hypothesis of correct specification for every model except the two semiparametric FZ models. This finding is consistent with Manganelli and Engle (2001), who find that their Conditional Autoregressive Value at Risk (CAViaR) model outperforms GARCH models at the 95% confidence level. The two semiparametric FZ models are an extension of the CAViaR model and as such are expected to outperform GARCH models at this lower confidence level.

In summary, Table 4 shows that irrespective of the confidence level we can reject that the HS, RM, Normal, STD, and GED models are correctly specified. These models underestimate VaR and often have clustered exceedances. Also, since the symmetric GARCH and

FZ parametric models are often rejected, we conclude that modeling skewness is crucial for estimating VaR in financial returns. At the extreme 95% confidence level, the FZH model is the only model to not be rejected by any test. However, the FHS, SSTD, SGED, SGT, Hill, and HillH models have the most accurate VaR estimates for the 99% and 97.5% confidence levels, the relevant levels for Basel regulation.

3.2 Traditional ES Backtests

Panel A of Table 5 reports traditional ES backtest results at the 99% confidence level. The ER p-values indicate that we can reject the hypothesis that the HS, RM, Normal, GED, and SGED models have the correct ES estimates on average. Since we evaluate the one-sided hypothesis of the unconditional ER test, we can conclude that these models systematically underestimate ES. The ESR Strict and Aux p-values indicate we can reject the hypothesis that the HS, RM, Normal, and GED models have accurate ES estimates. Specifically, we can reject the hypothesis of zero intercept and unit intercept when regressing exceedances from these models on ES. The ESR Int p-values for the one-sided intercept backtest show that these models along with the STD and SGED models have intercepts that are too low, confirming that these models underestimate ES on average. The CCa p-values indicate we can reject the hypothesis that the HS, RM, Normal, and GED models have accurate VaR and ES estimates. Since we conduct the one-sided CCa test where the null hypothesis is that the VaR and ES estimates are weakly greater than their true values on average, we conclude that the rejected models underestimate risk. We cannot reject the hypothesis that VaR and ES are correctly specified for the FHS, SSTD, SGT, GPD, Hill, HillH, FZ1, and FZH models in any traditional ES backtest at this level. The traditional ES backtest results are nearly identical for the 97.5% and 95% confidence levels in Panels B and C, except that the HillH model is rejected at these confidence levels.

Panel A of Table 6 reports unconditional spectral backtest results for the wide interval from (0.95, 0.995) using the Uniform (Uni), Arcsin (Arc), and Epanechnikov (Epa) continuous kernel density functions. The continuous kernel p-values indicate that we can reject the hypothesis that the RM, Normal, STD, GED, and HillH models have uniformly distributed probability integral transform (PIT) values. Given the results from the traditional backtests in Table 5, we conclude that these models have thinner tails than actual losses have, which likely results in underrepresented tail PIT values. Panel B of Table 6 reports unconditional spectral backtest results for the narrow interval from (0.97, 0.98), which is the neighborhood around the 97.5% confidence level used for Basel 3. The p-values for all three kernels indicate we reject the same set of models as for the wide interval. Panel C of Table 6 reports unconditional spectral backtest results for the uniform 3-level points (0.95, 0.975, 0.99), the main confidence levels of interest in our article. The p-values indicate that we reject the hypothesis that the RM, Normal, STD, and GED models have uniformly distributed PIT values. The set of models rejected is consistent across all the unconditional backtests, indicating these models have nonuniform PIT-values and are unlikely to be correctly specified. We cannot reject that the HS, FHS, SSTD, SGED, SGT, GPD, Hill, FZ1, and FZH models have uniformly distributed PIT-values in any unconditional spectral backtest.

Panel D of Table 6 reports conditional spectral backtest results for the wide interval. The continuous kernel p-values indicate that we can reject the hypothesis that the HS, RM, Normal, STD, GED, HillH, FZ1, and FZH models have uniformly distributed PIT-values

Table 5 Traditional ES backtests

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: Results at a 99% confidence level														
ER	**0.02**	**0.00**	0.53	**0.00**	0.10	0.41	**0.00**	**0.01**	0.41	0.20	0.57	0.14	0.31	0.07
ESR strict	**0.00**	**0.00**	0.98	**0.00**	0.09	0.79	**0.01**	0.26	0.73	0.96	0.99	0.62	0.28	0.87
ESR Aux	**0.00**	**0.00**	0.93	**0.00**	0.09	0.84	**0.01**	0.23	0.80	0.94	0.90	0.69	0.19	0.92
ESR Int	**0.00**	**0.00**	0.27	**0.00**	**0.01**	0.13	**0.00**	**0.02**	0.11	0.32	0.29	0.09	0.32	0.15
CCa	**0.01**	**0.00**	0.97	**0.00**	0.07	0.87	**0.01**	0.17	0.85	0.63	1.00	0.84	0.76	0.62
Panel B: Results at a 97.5% confidence level														
ER	**0.04**	**0.00**	0.74	**0.00**	0.25	0.39	**0.00**	**0.02**	0.48	0.62	0.84	**0.00**	0.93	0.86
ESR strict	**0.01**	**0.00**	0.90	**0.00**	**0.02**	0.79	**0.00**	0.46	0.77	0.76	0.71	0.08	0.99	0.90
ESR Aux	**0.00**	**0.00**	0.85	**0.00**	**0.02**	0.84	**0.00**	0.49	0.81	0.70	0.65	0.11	0.95	0.85
ESR Int	**0.00**	**0.00**	0.43	**0.00**	**0.00**	0.14	**0.00**	0.06	0.13	0.53	0.56	**0.01**	0.44	0.55
CCa	**0.01**	**0.00**	1.00	**0.00**	**0.01**	0.83	**0.00**	0.36	0.79	0.35	1.00	0.10	1.00	0.99
Panel C: Results at a 95% confidence level														
ER	**0.04**	**0.00**	0.84	**0.00**	**0.02**	0.42	**0.00**	0.12	0.54	0.91	0.94	**0.00**	0.96	0.93
ESR Strict	**0.00**	**0.00**	0.81	**0.00**	**0.00**	0.54	**0.00**	0.54	0.53	0.76	0.45	**0.00**	1.00	0.81
ESR Aux	**0.00**	**0.00**	0.90	**0.00**	**0.00**	0.58	**0.00**	0.57	0.55	0.80	0.54	**0.00**	0.98	0.85
ESR Int	**0.00**	**0.00**	0.39	**0.00**	**0.00**	0.07	**0.00**	0.06	0.06	0.46	0.63	**0.00**	0.26	0.37
CCa	0.05	**0.00**	1.00	**0.00**	**0.00**	0.42	**0.00**	0.42	0.38	1.00	1.00	**0.00**	1.00	1.00

This table shows the traditional ES backtesting results of the representative portfolio from January 1997 to February 2020. Each panel reports results for a particular confidence level. In each panel, Row 1 shows the one-sided p-values for the exceedance residual test. Rows 2 and 3 show the two-sided p-values for the Strict and Auxiliary ES regression backtests. Rows 4 and 5 show the one-sided p-values for the Intercept ES regression and Conditional Calibration backtests. Models with p-values below 0.05 are in bold.

and are serially independent. Interestingly, the HS, FZ1, and FZH models pass the unconditional backtests, but are rejected in the conditional backtest. This indicates that these models exhibit correlated spectrally transformed PIT-values. These models fail to use all available information when forecasting, resulting in temporal dependence between PIT-values.[16] Panel E of Table 6 reports conditional spectral backtest results for the narrow interval. The set of rejected models are the same as for the wide interval, reaffirming that these models have nonuniform or dependent PIT-values and are unlikely to be correctly specified. Panel F of Table 6 reports conditional spectral backtest results for the uniform three-level points. The p-values indicate that we reject the hypothesis that the HS, RM, Normal, STD, GED, FZ1, and FZH models have uniformly distributed and independent PIT-values. We cannot reject that the FHS, SSTD, SGED, SGT, GPD, and Hill models have uniformly distributed and independent PIT-values in any conditional spectral backtest.

In summary, Tables 5 and 6 show that the HS, RM, Normal, STD, GED, SGED, HillH, FZ1, and FZH models are likely not correctly specified for VaR and ES at the 99%, 97.5%,

16 This is confirmed by these models inferior performance in comparative backtests below.

Table 6 Spectral backtests

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: UC on a wide interval (0.95, 0.995)														
Uni	0.19	**0.00**	0.87	**0.00**	**0.00**	0.17	**0.00**	0.85	0.17	0.80	0.72	**0.00**	0.90	0.93
Arc	0.23	**0.00**	0.94	**0.00**	**0.00**	0.16	**0.00**	0.74	0.15	0.93	0.68	**0.01**	0.91	1.00
Epa	0.17	**0.00**	0.82	**0.00**	**0.00**	0.20	**0.00**	0.96	0.21	0.66	0.75	**0.00**	0.90	0.88
Panel B: UC on a narrow interval (0.97, 0.98)														
Uni	0.11	**0.00**	0.62	**0.00**	**0.00**	0.43	**0.00**	0.69	0.46	0.28	0.78	**0.00**	0.99	0.93
Arc	0.13	**0.00**	0.61	**0.00**	**0.00**	0.40	**0.00**	0.68	0.43	0.29	0.78	**0.00**	0.99	1.00
Epa	0.09	**0.00**	0.63	**0.00**	**0.00**	0.48	**0.00**	0.69	0.49	0.27	0.78	**0.00**	0.97	0.86
Panel C: UC on a discrete three-level points (0.95, 0.975, 0.99)														
Uni	0.21	**0.00**	0.92	**0.00**	**0.00**	0.31	**0.00**	0.83	0.21	0.77	0.73	0.11	0.88	0.92
Panel D: CC on a wide interval (0.95, 0.995)														
Uni	**0.00**	**0.00**	0.32	**0.00**	**0.00**	0.09	**0.00**	0.15	0.08	0.32	0.82	0.06	**0.00**	**0.01**
Arc	**0.00**	**0.00**	0.37	**0.00**	**0.00**	0.11	**0.00**	0.19	0.10	0.38	0.86	0.18	**0.00**	**0.02**
Epa	**0.00**	**0.00**	0.29	**0.00**	**0.00**	0.08	**0.00**	0.14	0.08	0.27	0.77	**0.02**	**0.00**	**0.00**
Panel E: CC on a narrow interval (0.97, 0.98)														
Uni	**0.00**	**0.00**	0.23	**0.00**	**0.00**	0.12	**0.00**	0.12	0.13	0.18	0.85	**0.01**	**0.00**	**0.00**
Arc	**0.00**	**0.00**	0.24	**0.00**	**0.00**	0.13	**0.00**	0.12	0.13	0.17	0.82	**0.01**	**0.00**	**0.00**
Epa	**0.00**	**0.00**	0.22	**0.00**	**0.00**	0.11	**0.00**	0.12	0.13	0.19	0.87	**0.01**	**0.00**	**0.00**
Panel F: CC on a discrete three-level points (0.95, 0.975, 0.99)														
Uni	**0.00**	**0.00**	0.36	**0.00**	**0.00**	0.12	**0.00**	0.15	0.08	0.52	0.93	0.52	**0.00**	**0.02**

This table shows the spectral backtest results of the representative portfolio from January 1997 to February 2020. Each panel reports results for a particular test type and interval. Panels A and B show p-values for the unconditional Z-test using a uniform, arcsin, and Epanechnikov kernel. Panel C shows p-values for the unconditional Z-test using a discrete uniform kernel. Panels D and E show p-values for the conditional Martingale Difference test using a uniform, arcsin, and Epanechnikov kernel and the conditioning variable transformation $h(p) = |2p - 1|^4$ with four lags. Panel F shows p-values for the conditional Z-test using a discrete uniform kernel. HS and GARCH models have the same PIT values across confidence levels. FZ PIT values are estimated at the 95% confidence level. Models with p-values below 0.05 are in bold.

and 95% confidence levels. These models either underestimate tail risk or have dependent forecasts. The FHS, SSTD, SGT, GPD, and Hill models are, on the contrary, never rejected by the traditional ES backtests. In particular, we cannot reject that these models have independent forecasts or PIT-values, suggesting that they are using the full information set when making forecasts. These models can accommodate the heavy tails and skewness of financial returns and appear to be correctly specified for VaR and ES. While the traditional

backtests in this section determines which models are incorrectly specified for VaR and ES, the next section compares across models to evaluate the set of best forecasting models.

3.3 Comparative ES Backtests

Table 7 shows the results for comparative ES backtests using the FZ0 loss function. Patton, Ziegel, and Chen (2019) show that FZ0 is the only FZ loss function that generates loss differences that are homogeneous of degree zero, a property that has been shown in volatility forecasting applications to lead to higher power in Diebold and Mariano (2002) tests. We consider results at the 97.5% confidence level here and, since the results are qualitatively similar, report 99% and 95% confidence levels in Section D of the Online Appendix. Panel A in Table 7 reports the average FZ0 loss for each model with a star besides models in the 75% MCS. The FHS, SSTD, SGED, SGT, GPD, and Hill models have the lowest average loss of −0.47 and are the only models that belong to the MCS. This is consistent with the traditional backtesting results, where the same set of models fails to be rejected as correctly specified in every backtest. The Normal, STD, GED, HillH, FZ1, and FZH models, on the contrary, have average losses between −0.46 and −0.42 and do not belong to the MCS. This is consistent with the traditional backtesting results, where these models are sometimes rejected as correctly specified. HS and RM have relatively higher average losses of −0.27 and −0.37 and are not in the MCS which is consistent with their rejection for nearly every traditional backtest.

Next, Panel B in Table 7 reports Diebold–Mariano test statistics on average FZ0 losses. Along each row, a negative number indicates the row model outperforms the column model. A t-stat greater than 1.96 in absolute value indicates the loss difference is statistically different from zero at the 95% confidence level. The HS row is positive and statistically significant for each column model, indicating that HS performs worse than every other model. The RM row shows the model underperforms every column model except for HS. The SSTD row, on the contrary, is negative for every column model indicating that SSTD has the lowest average loss. However, the loss difference is not statistically significant for models in the MCS. Interestingly, the FZ1 and FZH columns are positive and statistically significant for the models in the MCS, indicating they perform worse than this group of skewed GARCH models. This finding contradicts the results in Patton, Ziegel, and Chen (2019) who find that semiparametric FZ models outperform GARCH models. In their analysis, however, models are estimated once whereas we use a rolling window estimation procedure. Our results show that after accounting for time-varying parameters and changes to the distribution of innovations, GARCH models have superior forecasts compared to semiparametric FZ models.[17]

Figure 1 plots selected Murphy diagrams at the 99% and 97.5% confidence levels, comparing pairwise FHS to HS, FZH to HS, and FHS to FZH across the entire class of consistent loss functions. As a complement to the Murphy diagrams, Table 8 reports minimal Westfall–Young (WY) p-values, formally testing for forecasting dominance across a grid of thresholds. Figure 1a and d compare FHS to HS at the 99% and 97.5% confidence levels

17 Section D of the Online Appendix shows that skewed GARCH models also outperform FZ models at the 99% and 95% confidence levels, though the loss differences are not statistically significant at the 95% confidence level. Section E of the Online Appendix confirms that skewed GARCH models have lower losses than FZ models for the risk factor portfolios and liquidity horizon portfolios.

Table 7 Comparative ES backtests at the 97.5% confidence level

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: Average FZ0 loss														
Loss	−0.27	−0.37	−0.47*	−0.43	−0.45	−0.47*	−0.46	−0.47*	−0.47*	−0.47*	−0.47*	−0.46	−0.42	−0.44
Panel B: Diebold–Mariano test statistics for FZ0 loss														
HS	NA	2.46	4.09	3.51	4.03	4.20	4.10	4.17	4.19	4.04	4.08	4.13	3.50	3.54
RM	**−2.46**	NA	3.55	2.97	3.89	3.83	4.00	3.73	3.81	3.40	3.51	3.79	1.88	2.23
FHS	**−4.09**	**−3.55**	NA	**−2.45**	−1.39	0.50	−1.11	0.32	0.41	−0.43	0.03	−0.54	**−2.74**	**−2.26**
Norm	**−3.51**	**−2.97**	2.45	NA	3.99	3.00	3.92	2.78	2.98	2.25	2.39	2.74	−0.10	0.42
STD	**−4.03**	**−3.89**	1.39	**−3.99**	NA	2.05	2.24	1.74	2.01	1.18	1.34	1.57	−1.73	−1.06
SSTD	**−4.20**	**−3.83**	−0.50	**−3.00**	**−2.05**	NA	−1.80	−0.40	−0.42	−0.58	−0.42	−1.21	**−2.97**	**−2.39**
GED	**−4.10**	**−4.00**	1.11	**−3.92**	**−2.24**	1.80	NA	1.50	1.75	0.92	1.06	1.25	**−2.03**	−1.34
SGED	**−4.17**	**−3.73**	−0.32	**−2.78**	−1.74	0.40	−1.50	NA	0.18	−0.51	−0.27	−0.87	**−2.93**	**−2.35**
SGT	**−4.19**	**−3.81**	−0.41	**−2.98**	**−2.01**	0.42	−1.75	−0.18	NA	−0.50	−0.35	−1.20	**−2.93**	**−2.35**
GPD	**−4.04**	**−3.40**	0.43	**−2.25**	−1.18	0.58	−0.92	0.51	0.50	NA	0.47	−0.18	**−2.71**	**−2.21**
Hill	**−4.08**	**−3.51**	−0.03	**−2.39**	−1.34	0.42	−1.06	0.27	0.35	−0.47	NA	−0.47	**−2.72**	**−2.25**
HillH	**−4.13**	**−3.79**	0.54	**−2.74**	−1.57	1.21	−1.25	0.87	1.20	0.18	0.47	NA	**−2.63**	**−2.09**
FZ1	**−3.50**	**−1.88**	2.74	0.10	1.73	2.97	2.03	2.93	2.93	2.71	2.72	2.63	NA	0.88
FZH	**−3.54**	**−2.23**	2.26	−0.42	1.06	2.39	1.34	2.35	2.35	2.21	2.25	2.09	−0.88	NA

This table shows the comparative ES backtesting results of the representative portfolio from January 1997 to February 2020. Results are based on the FZ0 loss function at the 97.5% confidence level. Panel A shows the average FZ0 loss with a star beside models in the 75% MCS. Panel B shows *t*-statistics from Diebold–Mariano tests comparing FZ0 average losses. A negative value indicates that the row model has lower average loss than the column model. The *t*-statistics greater than 1.96 in absolute value indicate the loss difference is significantly different from zero at the 95% confidence level. Models with *t*-stat below −1.96 are in bold.

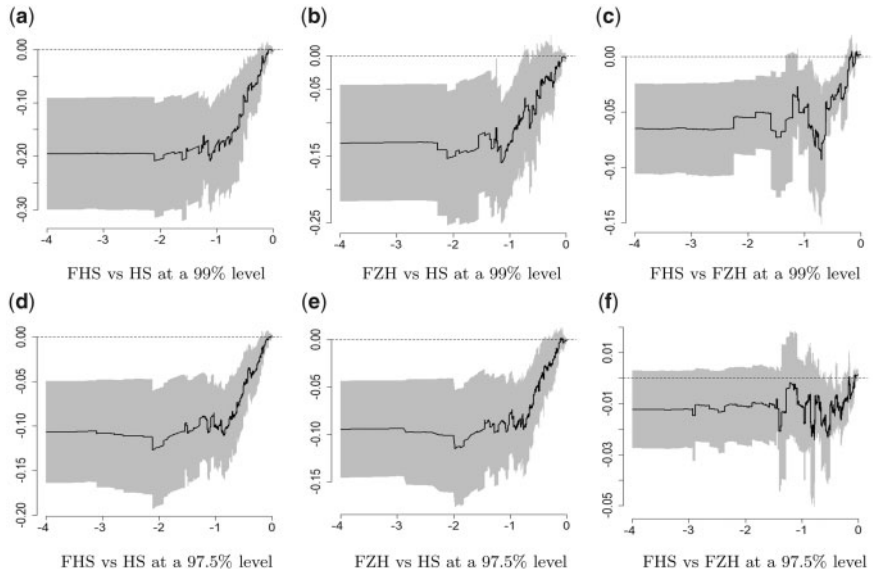

Figure 1 Murphy diagrams at the 97.5% confidence level. (**a**) FHS versus HS at a 99% level. (**b**) FZH versus HS at a 99% level. (**c**) FHS versus FZH at a 99% level. (**d**) FHS versus HS at a 97.5%. level. (**e**) FZH versus HS at a 97.5% level. (**f**) FHS versus FZH at a 97.5% level. This figure shows Murphy diagrams of the representative portfolio from January 1997 to February 2020 for the 99% and 97.5% confidence levels. In each panel, the vertical axis is the difference in average loss between models, where a negative value indicates the first model outperforms the second model. The horizontal axis is the threshold value, where each threshold corresponds to a different consistent loss function. Confidence intervals at the 95% level are plotted in gray.

Table 8 Tests of forecast dominance

Hypothesis	99% confidence level	97.5% confidence level	95% confidence level
HS weakly dominates FHS	**0.002**	**0.000**	**0.000**
FHS weakly dominates HS	0.940	0.702	0.808
HS weakly dominates FZH	**0.020**	**0.004**	**0.004**
FZH weakly dominates HS	0.982	0.912	0.406
FZH weakly dominates FHS	0.050	**0.048**	**0.002**
FHS weakly dominates FZH	0.688	0.904	0.800

This table shows the tests of forecast dominance of the representative portfolio from January 1997 to February 2020 for the 99%, 97.5%, and 95% confidence levels. Each row shows the minimal WY p-values for the hypothesis of weak dominance. Models with p-values below 0.05 are in bold.

and show that the loss difference is negative at nearly every threshold, indicating FHS outperforms HS. The 95% confidence interval shows that the difference is nearly always statistically different from zero. The WY p-values in Row 1 of Table 8 show that we can reject the hypothesis of HS dominating FHS at 99%, 97.5%, and 95% confidence levels. The WY p-values in Row 2 show that we cannot reject FHS dominating HS at any confidence level.

Next, Figure 1b and e show that FZH outperforms HS. The WY p-values in Row 3 of Table 8 show that we can reject the hypothesis that HS weakly dominates FZH, while Row 4 shows that we cannot reject that FZH weakly dominates HS.

Finally, Figure 1c and f plot the Murphy diagram comparing FHS to FZH at 99% and 97.5% confidence levels. The negative loss difference shows that FHS outperforms FZH across different consistent loss functions. Rows 5 and 6 of Table 8 report WY p-values for the hypothesis of weak dominance between FHS and FZH. Row 6 shows that we cannot reject the hypothesis that FHS dominates FZH at 99%, 97.5%, and 95% confidence levels. However, Row 5 shows we can reject the hypothesis that FZH dominates FHS at 97.5% and 95% confidence levels only. The comparative backtest found FHS outperforms FZH for the FZ0 loss function, while the Murphy diagram and WY p-values confirm that FHS outperforms FZH across a large family of loss functions.

In summary, the comparative ES backtests show that the FHS, SSTD, SGED, SGT, GPD, and Hill models consistently have the best performance across the 99%, 97.5%, and 95% confidence levels. The HS, RM, Normal, STD, GED, HillH, FZ1, and FZH models have statistically higher losses compared to the best performing models. We also find that the best performing GARCH models had lower average losses than the FZ models, and we fail to reject the hypothesis that FHS weakly dominates FZH. The set of best performing models is broadly consistent with the set of correctly specified models in the traditional backtests. Considering all the results reported in this section, we conclude that skewed GARCH models are likely correctly specified and perform the best in terms of forecasting risk measures. The semiparametric FZ1 and FZH models are generally outperformed by the skewed GARCH models. The symmetric Normal, STD, and GED models perform worse than the previous models and are likely not correctly specified. The HS and RM models, which are frequently used by banks, are nearly always rejected as being correctly specified and are the worst-performing models in terms of forecasting the relevant risk measures.

4. Regulatory Capital Requirements

Assuming that capital is costly and hence banks are interested in minimizing their capital requirements, Basel 3 incentivizes banks to select models that produce low Stressed ES.[18] However, at the same time, Basel 3 also penalizes models with too many VaR exceedances through the backtesting multiplier which increases capital requirements and incentivizes banks to use more conservative models. Given this trade-off, it's not clear which model minimizes capital for banks. Moreover, from a regulatory supervisor's perspective, banks using incorrectly specified models that underestimate risk is problematic during a financial crisis, since the banks' risk management models will fail at the same time which increases systemic risk. Considering the backtesting results from Section 3, this section determines whether Basel regulation is in fact incentivizing banks to choose models that are correctly specified.

18 While lower capital requirements may be privately optimal for banks, higher capital requirements may be socially beneficial in reducing the likelihood of systemic crisis. See Birn et al. (2020) for a review on the costs and benefits of bank capital.

We address the question about correctly incentivizing banks by carefully calculating the capital requirements for a representative bank with the diverse portfolio of assets described in Section 2.1. We do this for each model under the Basel 3 regulatory framework to determine which models minimize capital requirements. We then use the same data to calculate requirements under Basel 2 and Basel 2.5 to determine whether the set of models that minimize capital change across the three regimes. Finally, we measure the volatility and the peak-to-trough variation as the maximum difference of capital requirements in Basel 2, 2.5, and 3 to determine whether this regulation is successful at increasing the stability of the capital requirements.

We empirically determine the capital requirements of each model for the representative bank portfolio from January 1997 to February 2020. We calculate capital requirements for Basel 2, 2.5, and 3 by fixing the regime and closely following the regime's capital calculations over the entire period. Consistent with backtesting, we use a rolling estimation window of 250 days for HS and RM and 2000 days for the dynamic GARCH and FZ models. In each table, columns are bold if the models are rejected in the traditional ES or spectral backtesting of Section 3. Previewing our results, we find that HS and FZH have the lowest capital requirement for Basel 3. In fact, the results show that incorrectly specified models generally have lower capital requirements than correctly specified ones and Basel 3, therefore, does not incentivize banks to use correctly specified models. Moreover, we find that while average capital requirements nearly quadruple from Basel 2 to Basel 2.5, they decreased significantly from Basel 2.5 to Basel 3 only for incorrectly specified models. Finally, we show that capital stability generally increases with the regulatory changes. However, the models that have the lowest volatility, the FZ1 and FZH models, are not among the correctly specified models.

4.1 Basel 3 Capital Requirements

Table 9 shows the average capital requirements of each model calculated under the Basel 3 framework. We also provide a detailed breakdown of the intermediary calculations by risk factor and liquidity horizon to verify that our findings are robust to various portfolio specifications. Panel A of Table 9 reports the average Basel 3 capital requirement, backtest multiplier, and IMCC. Row 1 demonstrates that correctly specified models have significantly higher average Basel 3 capital than rejected models. HS has the lowest average Basel 3 capital of 17.21%, despite having the worst backtesting performance in Section 3. The light-tailed RM and GARCH Normal models also have low average capital of 19.47% and 21.69%, despite having poor backtesting performance. The FZ1 and FZH models outperform these models during backtesting and also have low average capital of 17.42% and 19.96%. The FHS, SSTD, SGT, GPD, and Hill models generally have the highest average capital requirements of 26.61–32.38%, despite failing to be rejected in any ES backtest and having the lowest average FZ0 losses in Table 7. While FHS weakly dominates HS and FZH across the set of consistent loss functions, using this model requires 1.78 times the average daily capital. SSTD has the lowest average capital among correctly specified models, but using this model still requires 1.55 times the capital relative to HS and FZH.

The multiplier and IMCC in Panel A of Table 9 explain why correctly specified models have higher average capital than rejected models. Despite poor backtesting performance, HS has an average multiplier of 1.57. The average multiplier of correctly specified models

Table 9 Basel 3 capital requirements

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: Basel 3 average capital requirement														
Basel 3 Capital	**17.21**	**19.47**	30.64	**21.69**	**25.77**	26.61	**26.33**	**27.48**	27.38	32.38	29.85	**32.46**	**19.96**	**17.42**
Basel 3 Multiplier	**1.57**	**1.65**	1.53	**1.63**	**1.58**	1.55	**1.58**	**1.55**	1.56	1.52	1.53	**1.53**	**1.53**	**1.53**
IMCC	**10.96**	**11.82**	20.04	**13.27**	**16.29**	17.20	**16.67**	**17.76**	17.66	21.37	19.48	**21.22**	**13.14**	**11.43**
Panel B: Liquidity-adjusted stressed ES by risk factor														
All	**10.18**	**9.92**	16.94	**11.11**	**13.70**	15.38	**14.45**	**15.78**	14.96	18.98	17.23	**19.27**	**11.45**	**9.59**
Interest	**0.25**	**0.39**	0.47	**0.38**	**0.45**	0.43	**0.42**	**0.42**	0.45	0.47	0.44	**0.43**	**0.45**	**0.50**
Equity	**5.82**	**6.71**	12.27	**7.72**	**9.81**	9.67	**9.46**	**10.22**	10.91	13.57	12.43	**13.06**	**8.15**	**6.68**
Commodities	**1.46**	**1.55**	1.60	**1.28**	**1.41**	1.47	**1.44**	**1.46**	1.44	1.46	1.55	**1.47**	**1.48**	**1.40**
Foreign exch.	**0.37**	**0.79**	0.67	**0.65**	**0.67**	0.74	**0.70**	**0.72**	0.72	0.64	0.66	**0.67**	**0.47**	**0.51**
Credit	**3.85**	**4.29**	8.13	**5.38**	**6.54**	6.71	**6.86**	**6.92**	6.84	7.62	6.64	**7.54**	**4.28**	**4.17**
Panel C: Stressed ES of representative portfolio by liquidity horizon														
LH 10	**6.43**	**6.21**	7.61	**5.73**	**6.70**	7.74	**6.83**	**7.20**	7.32	7.78	8.15	**7.32**	**6.71**	**6.35**
LH 20	**4.98**	**4.57**	5.86	**4.31**	**5.10**	5.69	**5.14**	**5.69**	5.59	6.14	5.45	**5.85**	**4.56**	**3.64**
LH 40	**3.03**	**2.82**	6.56	**4.12**	**5.26**	5.94	**5.45**	**5.85**	5.84	7.48	5.87	**7.02**	**3.52**	**2.99**
LH 60	**2.42**	**2.36**	6.43	**3.76**	**4.73**	5.16	**5.43**	**6.17**	5.00	7.60	6.83	**7.15**	**3.29**	**2.46**
LH 120	**0.33**	**0.44**	0.50	**0.40**	**0.42**	0.45	**0.43**	**0.45**	0.47	0.42	0.47	**0.40**	**0.35**	**0.37**

This table shows the Basel 3 capital requirement results of the representative bank portfolio from January 1997 to February 2020. Panel A shows the average daily Basel 3 capital requirement in percentage terms, backtesting multiplier, and IMCC in percentage terms. Panel B shows in percentage terms the average liquidity-adjusted Stressed ES of the representative portfolio and separated by risk factor. Panel C shows in percentage terms the Stressed ES of the representative portfolio for each individual liquidity horizon. Models rejected by the traditional ES or spectral backtests are bold.

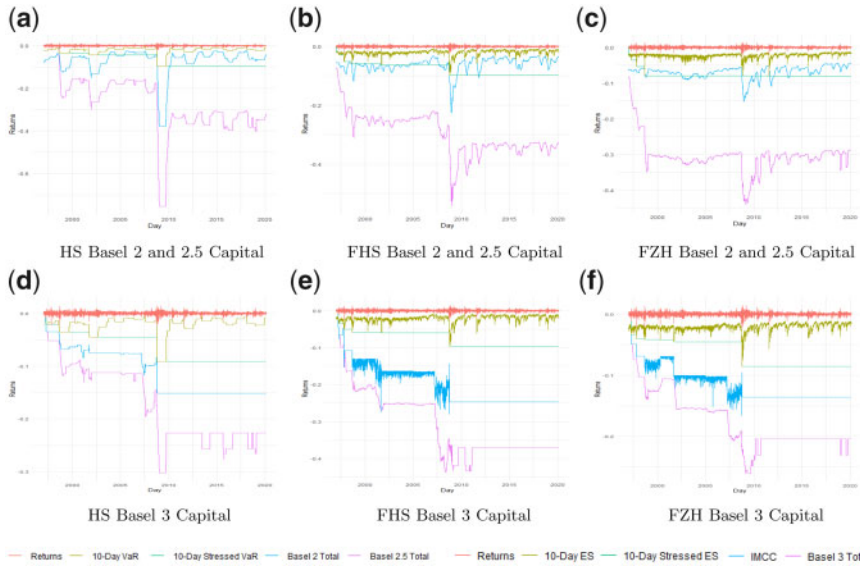

Figure 2 Basel capital requirements. (**a**) HS Basel 2 and 2.5 capital. (**b**) FHS Basel 2 and 2.5 capital. (**c**) FZH Basel 2 and 2.5 capital. (**d**) HS Basel 3 capital. (**e**) FHS Basel 3 capital. (**f**) FZH Basel 3 capital. This figure shows the returns, negative risk measures, and capital requirements of the representative portfolio from January 1997 to February 2020 for HS, FHS, and FZH. The top panels plot the daily returns, negative 10-day VaR, Basel 2 capital requirement, 10-day Stressed VaR, and Basel 2.5 capital requirement. The bottom panels plot the daily returns, negative 10-day ES, 10-day Stressed ES, IMCC, and Basel 3 capital requirement.

ranges from 1.52 to 1.56. Since Basel 3 capital is equal to IMCC times the multiplier, we see that differences in average Basel 3 capital across models are almost entirely attributed to differences in IMCC. HS has the lowest IMCC of 10.96%, while FHS has nearly double the IMCC at 20.04% due to its dynamic volatility and heavy tails. The difference between HS and FHS is illustrated in Panels (d) and (e) of Figure 2, which shows that FHS has a higher average capital and IMCC on every day after the 1998 Asian crisis. SSTD has the lowest IMCC among correctly specified models at 17.20%, which is still substantially higher than HS. The RM and Normal models have the highest multipliers at 1.65 and 1.63, but much lower average capital requirements than SSTD due to their low IMCC. The FZ models have both low average multipliers of 1.53 and low IMCC, resulting in low Basel 3 capital. Generally, the least costly model for banks is the one that minimizes IMCC, or equivalently minimizes Stressed ES at the 97.5% confidence level. The HS, RM, Normal, FZ1, and FZH models minimize Stressed ES, resulting in significantly lower Basel 3 average capital requirements. However, these models are either rejected using traditional backtests or provide inferior forecasts of risk measures.

Panel B of Table 9 shows the liquidity-adjusted Stressed ES of the representative bank portfolio as well as by individual risk factors. From Equation (8), IMCC is calculated by taking a weighted sum of the representative portfolio and risk factor portfolios' liquidity-adjusted Stressed ES. Row 1 of Panel B shows that for the representative portfolio, correctly specified models have high liquidity-adjusted Stressed ES ranging from 14.96% to 18.98%.

The rejected HS, RM, Normal, FZ1, and FZH models have the lowest liquidity-adjusted Stressed ES, the same set of models that minimize IMCC and Basel 3 capital. Rows 2–6 of Panel B show that HS has the lowest ES for the interest, equity, foreign exchange, and credit risk factors, while the Normal model has the lowest ES for commodities. The FZH model has the lowest ES for the representative portfolio of 9.59%, but it has moderate levels of ES for the individual risk factors explaining why its IMCC is higher than HS. The correctly specified models have high ES for every risk factor, verifying that our results are not sensitive to risk factor weights. FHS has more than twice the ES of HS for equity and credit, the two largest risk factors. Across risk factors, the models that minimize liquidity-adjusted Stressed ES are rejected by backtests, while correctly specified models have the highest ES.

Panel C of Table 9 shows the Stressed ES of the representative bank portfolio for each liquidity horizon. The liquidity-adjusted Stressed ES of the representative portfolio is calculated as a function of these liquidity portfolios. The panel shows that the HS, RM, Normal, FZ1, and FZH models minimize Stressed ES across liquidity horizons. Since differences in Basel 3 capital are driven by Stressed ES, this explains why these models have the lowest average Basel 3 capital requirements. The ES backtests in Section 3 are only for the representative portfolio (LH 10). For this portfolio, correctly specified models have the highest Stressed ES of 7.32% to 8.15%. We show in Section E of the Online Appendix that for the LH 20–60 portfolios, skewed GARCH and FZ models are likely to be correctly specified. The HS, RM, Normal, STD, GED, SGED models are rejected for the LH 20–60 portfolios, explaining their lower Stressed ES. For the LH 20–60 portfolios, FZ1 and FZH are never rejected and also have a low average Stressed ES. These results also show that every model except STD is rejected for the LH 120 portfolio. However, the HS, FZ1, and FZH models have the lowest Stressed ES for LH 120.

In Basel 3, capital requirements are minimized for models that consistently minimize Stressed ES at the 97.5% confidence level across risk factors and liquidity horizons. HS, RM, Normal, FZ1, and FZH consistently minimize Stressed ES across portfolios, resulting in low Basel 3 capital requirements. The HS, RM, Normal, FZ1, and FZH models are misspecified since they are rejected in nearly every ES backtest, have the largest FZ0 losses, or provide inferior forecasts of risk measures. The FHS, SSTD, SGT, GPD, and Hill models are never rejected in the traditional backtests and minimize FZ0 losses. However, they have high Stressed ES estimates, resulting in much higher average Basel 3 capital requirements. Thus, the first result is that under Basel 3 there is little incentive for a capital requirement minimizing bank to choose correctly specified models. In particular, given the level of the Basel 3 multiplier, this tool that regulators can use to punish misspecified models has only a marginal effect on average capital requirements and banks have little incentive to choose conservative and correctly specified models.

4.2 Evolution of Basel Capital Requirements

Table 10 summarizes the average Basel 2, 2.5, and 3 capital requirements, risk measures, and Basel backtesting results. Panel A of Table 10 shows the average 10-day VaR at the 99% confidence level and average capital requirements of the entire sample calculated under the Basel 2 framework. Row 1 shows that the HS average 10-day VaR is 2.17%, which is roughly 2.5 times the 10-day volatility ($0.276\sqrt{10} = 0.87$) of the representative portfolio in Table 3. All dynamic models have a lower 10-day VaR compared to HS, since

Table 10 Basel 2, 2.5, and 3 average capital requirements

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: Basel 2 average capital requirements														
VaR	**2.17**	**1.80**	2.06	**1.23**	**1.73**	1.86	**1.71**	**1.83**	1.86	2.04	2.14	**1.89**	**1.92**	**2.08**
Basel 2 capital	**7.12**	**5.98**	6.09	**4.98**	**5.38**	5.61	**5.35**	**5.56**	5.60	5.98	6.12	**5.91**	**5.93**	**6.88**
Panel B: Basel 2.5 average capital requirements														
Stressed VaR	**6.45**	**6.18**	7.76	**6.70**	**6.60**	7.03	**6.49**	**6.85**	7.04	7.35	7.89	**7.48**	**5.78**	7.75
Basel 2.5 capital	**27.34**	**26.27**	29.73	**26.83**	**26.17**	27.26	**25.78**	**26.67**	27.31	28.19	30.15	**28.71**	**23.44**	**30.48**
Panel C: Basel 3 average capital requirements														
Stressed ES	**6.43**	**6.21**	7.61	**5.73**	**6.70**	7.74	**6.83**	**7.20**	7.32	7.78	8.15	**7.32**	**6.71**	**6.35**
Basel 3 capital	**17.21**	**19.47**	30.64	**21.69**	**25.77**	26.61	**26.33**	**27.48**	27.38	32.38	29.85	**32.46**	**19.96**	**17.42**
Panel D: Basel backtests on one-day VaR														
Exceedances	**3.49**	**4.93**	2.48	**4.48**	**3.42**	2.82	**3.42**	**2.82**	2.91	2.22	2.48	**2.48**	**2.26**	**2.04**
Basel 2 multiplier	**3.15**	**3.30**	3.07	**3.26**	**3.16**	3.10	**3.16**	**3.10**	3.11	3.04	3.07	**3.07**	**3.05**	**3.06**

This table shows the Basel 2, 2.5, and 3 risk measures and capital requirements of the representative bank portfolio from January 1997 to February 2020. Panel A shows the average daily VaR estimate and Basel 2 capital requirement in percentage terms. Panel B shows the average daily Stressed VaR estimate and Basel 2.5 capital requirement in percentage terms. Panel C shows the average Stressed ES (LH 10) and Basel 3 capital requirements in percentage terms. Panel D shows the average number of exceedances and the Basel 2.5 multiplier. Models rejected by the traditional ES or spectral backtests are bold.

dynamic volatility decreases risk levels faster after a shock. This phenomenon is illustrated in Figure 2a and b, which plots HS and FHS 10-day VaR. After a shock like that following the 2008 Subprime Mortgage crisis, the FHS model quickly decreases VaR levels, while HS maintains high levels of VaR for the entire estimation window. The model with the lowest VaR though is the Normal model. Row 2 shows that the Normal model also has the lowest average Basel 2 capital requirement of 4.98%, despite the Basel backtest results in Panel D showing that this model has the second highest average exceedances and multiplier of 4.48 and 3.26, respectively. In contrast to Basel 3 results, HS has the highest average Basel 2 requirement of 7.12% due to its high 10-day VaR and capital multiplier of 3.15. The correctly specified models have high average Basel 2 capital of 5.60–6.12%, despite having low multipliers of 3.04–3.11. This shows that Basel 2 failed to incentivize banks to choose correctly specified models.

Next, Panel B of Table 10 reports the average 10-day Stressed VaR and Basel 2.5 capital requirement. Stressed VaR is defined as the maximum historical 10-day VaR at the 99% confidence level and the panel shows that the average Stressed VaR is significantly higher than VaR. In contrast to nonstressed VaR, all of the GARCH models have higher Stressed VaR than HS does, since the stressed measure never decreases. Put differently, the VaR-minimizing benefit of dynamic volatility disappears for Stressed VaR. The top panels of Figure 2 show that the stressed period for HS, FHS, and FZH is set during the Asian crisis, the Dot-Com crisis, and the 2008 recession resulting in high values of Stressed VaR. FZ1 has the lowest Stressed VaR across all models at 5.78%. Row 2 shows that the FZ1 model also has the lowest average Basel 2.5 capital, partially due to its low multiplier. HS has a low Stressed VaR but a high multiplier resulting in moderate Basel 2.5 capital. The correctly specified FHS, SSTD, SGT, GPD, and Hill models have a high Stressed VaR but a low multiplier and as a result reasonable capital requirements. Consequently, the correctly specified SSTD and SGT models require less average capital than HS. These results show that Basel 2.5 has better incentives to use correctly specified models relative to Basel 2.

Finally, Panels A–C of Table 10 together allow us to compare capital calculated under Basel 2, 2.5, or 3. We first note that models generally require more than four times the average capital under Basel 2.5 than under Basel 2. As illustrated in the top panels of Figure 2, this increase in capital is driven by shocks during the 1998, 2001, and 2008 crises leading to a large Stressed VaR. Since Stressed VaR never decreases, capital levels stay elevated after the 2008 crisis across all models. The comparison of Basel 2.5 to Basel 3 is more nuanced. For HS, there is a significant decrease in required capital from 27.34% under Basel 2.5 to 17.21% under Basel 3, and Figure 2a and d illustrates that HS requires more capital under Basel 2.5 on each day. The capital requirements under Basel 3 are also much lower than under Basel 2.5 for the FZ and light-tailed RM and Normal models but remain mostly unchanged from Basel 2.5 to Basel 3 for the skewed GARCH models.

So what drives these differences from Basel 2.5 to Basel 3? First of all, comparing HS Stressed 10-day VaR at the 99% level in Panel B with Stressed 10-day ES at the 97.5% level in Panel C shows that the two stressed risk measures are nearly identical. However, Table 9 shows that HS has substantially lower Stressed ES for the 40-, 60-, and 120-day liquidity-horizon portfolios, also causing low liquidity-adjusted Stressed ES across several risk factors. The capital requirements under Basel 3 are also much lower than under Basel 2.5 for the FZ and light-tailed RM and Normal models due to their relatively low Stressed ES estimates for longer liquidity-horizon portfolios and low liquidity-adjusted Stressed ES.

Interestingly, it appears that the requirement under Basel 3 to penalize low liquidity assets (i.e., assets with long liquidity horizons), additionally in fact disincentivizes banks from using the correctly specified skewed GARCH models due to their consistently high Stressed ES across liquidity horizons. Instead, under Basel 3 banks are incentivized to use misspecified models that consistently minimize Stressed ES across liquidity horizons.

4.3 Basel Capital Variability

Table 11 reports several measures of capital variability across the Basel 2, 2.5, and 3 regimes. We measure the daily capital volatility to capture day-to-day fluctuations in capital and report volatility standardized by capital to compare across regimes as well. We also measure the maximum annual differences in capital to capture peak-to-trough variation in capital requirements. Panel A of Table 11 shows the capital volatility, which is suitable for comparing across models holding the regime fixed. Row 1 shows that FZH has the lowest Basel 2 volatility of 1.64% followed by FZ1 with 1.85%. The non-normal GARCH models have Basel 2 volatility ranging from 2.16% for GPD to 2.38% for FHS. The RM and Normal models have significantly higher volatility at 2.73% and 2.91%, but HS has the absolute highest Basel 2 volatility with 6.57%. Row 2 shows that FZ1 and FZH have the lowest Basel 2.5 volatility at 4.06% and 4.46%. Compared to this, HS has more than three times the volatility of the FZ models and double the volatility of most GARCH models. Figure 2a shows that the high HS volatility is caused by its short estimation window and the Basel multiplier. Figure 2c, on the contrary, shows that FZH has stable estimates of 10-day VaR and is generally unaffected by the Basel multiplier. Row 3 shows that HS has a moderate Basel 3 volatility of 7.17% whereas FHS has the highest volatility at 8.85% under this regulation. FZ1 and FZH continue to have the lowest Basel 3 volatility at 5.02% and 4.66%. These numbers are close to the Basel 2.5 volatility levels for the FZ models whereas Basel 3 volatility is significantly higher than Basel 2.5 volatility for most of the GARCH models. Seen across the regulatory regimes, the table shows that the FZ models have the lowest capital volatility for Basel 2, 2.5, and 3, and banks seeking to minimize capital volatility in Basel 3, in particular, since it may be costly to raise capital during times of high volatility, should consider using the FZH model.

Figure 3 visualizes the trade-off between capital minimization and stability of capital requirements by plotting the mean against the volatility of model capital. We color code the rejected models in red and the correctly specified models in blue. For each regime, the bottom models minimize mean capital and the left models minimize capital variability. Panel (a) shows that HS in the top-right maximizes both mean and volatility of Basel 2 capital. The GARCH models are clustered with the Normal model minimizing mean and the correctly specified models having high means. The FZ models minimize volatility with a moderate mean capital. Panel (b) shows that HS maximizes Basel 2.5 volatility with a moderate mean capital. The GARCH models are clustered with moderate mean and volatility of capital. FZ models minimize Basel 2.5 volatility and FZ1 has the lowest mean capital as evidenced by its position in the bottom left, and banks can minimize both the mean and volatility of Basel 2.5 capital requirements by using the FZ1 model. Panel (c) shows that the FZ models minimize Basel 3 volatility with low means as evidenced by their positions in the bottom left. HS has the lowest mean capital, but has higher volatility than the FZ models. GARCH models have both high mean and volatility as evidenced by their positions in the

Table 11 Basel 2, 2.5, and 3 capital variability

	HS	RM	FHS	Norm	STD	SSTD	GED	SGED	SGT	GPD	Hill	HillH	FZ1	FZH
Panel A: Volatility in percentage terms														
Basel 2 Vol	**6.57**	**2.91**	2.38	**2.73**	**2.33**	2.27	**2.32**	**2.25**	2.27	2.16	2.36	**2.31**	**1.85**	**1.64**
Basel 2.5 Vol	**13.43**	**6.15**	6.85	**8.39**	**6.71**	6.14	**5.87**	**5.89**	5.89	5.71	6.65	**6.13**	**4.06**	**4.46**
Basel 3 Vol	**7.17**	**5.72**	8.85	**7.51**	**7.54**	7.63	**7.92**	**7.66**	7.02	7.36	6.12	**7.26**	**5.02**	**4.66**
Panel B: Volatility standardized by average capital														
Basel 2 Vol/CA	**0.92**	**0.49**	0.39	**0.55**	**0.43**	0.40	**0.43**	**0.40**	0.40	0.36	0.39	**0.39**	**0.31**	**0.24**
Basel 2.5 Vol/CA	**0.49**	**0.23**	0.23	**0.31**	**0.26**	0.23	**0.23**	**0.22**	0.22	0.20	0.22	**0.21**	**0.17**	**0.15**
Basel 3 Vol/CA	**0.42**	**0.29**	0.29	**0.35**	**0.29**	0.29	**0.30**	**0.28**	0.26	0.23	0.21	**0.22**	**0.25**	**0.27**
Panel C: Maximum annual difference in percentage terms														
Basel 2 max. diff.	**33.58**	**17.38**	15.63	**18.60**	**15.22**	14.19	**15.16**	**14.43**	14.45	13.57	15.75	**15.13**	**10.20**	**9.18**
Basel 2.5 max. diff.	**58.05**	**24.96**	24.07	**33.11**	**25.45**	20.76	**21.37**	**21.12**	21.14	18.76	23.74	**21.42**	**17.31**	**16.23**
Basel 3 max. diff.	**15.71**	**13.72**	16.94	**18.85**	**20.87**	17.52	**21.07**	**19.74**	22.51	23.91	29.06	**21.61**	**52.79**	**10.26**

This table shows the Basel 2, 2.5, and 3 variability measures of the representative bank portfolio from January 1997 to February 2020. Panel A shows the daily volatility of capital in percentage terms. Panel B shows the daily volatility divided by the average capital requirement in percentage terms. Panel C shows the maximum annual difference (250-day difference) of capital requirements in percentage terms. Models rejected by the traditional ES or spectral backtests are bold.

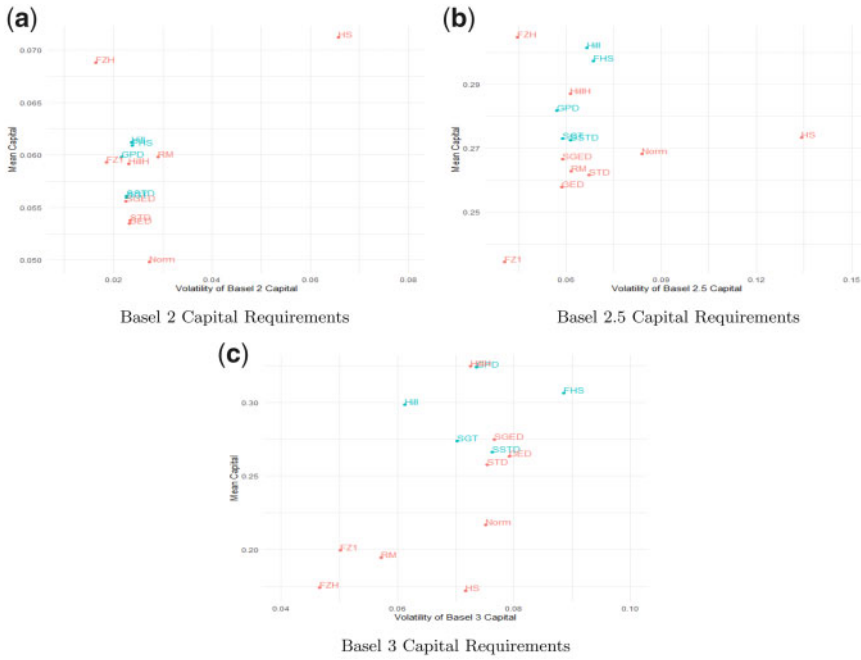

Figure 3 Basel models mean-volatility plots. (**a**) Basel 2 capital requirements. (**b**) Basel 2.5 capital requirements. (**c**) Basel 3 capital requirements. This figure shows mean and volatility plots of the representative portfolio from January 1997 to February 2020 for Basel 2, 2.5, and 3. In each panel, the vertical axis is the mean capital requirement and the horizontal axis is the volatility of capital requirements. Models in the bottom left minimize mean and volatility of capital requirements, while models at the top right maximize mean and volatility.

top right. Basel 3 disincentivizes banks to use these correctly specified models, since they have the highest means and volatilities, despite never being rejected in any ES backtest and having the lowest FZ losses. Banks are instead incentivized to use FZ models, which minimize mean and variability of capital, but perform worse than GARCH models in the comparative backtests. Many banks may remain with HS, since the model minimizes mean capital and is the simplest to implement. This may be problematic to regulators, since HS has the worst ES backtest performance across all models and systematically underestimates ES. However, since there is negligible punishment for misspecification and the Basel backtests only VaR, banks have little incentive to move away from HS. Regulators may be able to incentivize banks to move toward correctly specified models if they penalize misspecification in VaR and ES.

Next, we compare stability across Basel regimes using volatility standardized by capital. We use standardized volatility to capture models with either high volatility or low capital. Panel B of Table 11 shows that standardized volatility significantly declines between Basel 2 and Basel 2.5 as a result of including Stressed VaR. This decline is most apparent for HS, where standardized volatility decreases from 0.92 to 0.49. Rows 2 and 3 show that HS capital volatility further declines to 0.42 from Basel 2.5 to 3, demonstrating that Basel 3 indeed

induces more stability when using HS. However, nearly every other model experiences an increase in capital volatility from Basel 2.5 to Basel 3. Figure 2d illustrates the stability of HS under Basel 3 relative to Basel 2.5. Figure 2e and f show that for FHS and FZH, the increase in Basel 3 volatility is largely driven by variation in $\frac{ES_{F,C}}{ES_{R,C}}$, the ratio that Basel 3 uses to adjust the reduced portfolio to the current. This point-in-time adjustment ratio causes day-to-day capital changes in dynamic models as evidenced by the erratic IMCC changes between 1999 and 2008, but is set to 1 after the reduced set of portfolios becomes the full set in 2008. The increase in standardized volatility for FZ models also occurs due to their significant decrease in capital requirements from Basel 2.5 to Basel 3. In general, most of the models are more stable in Basel 3 than the previous regimes.

Finally, Panel C of Table 11 shows the maximum annual difference in percentage terms. The max difference measures the largest annual peak-to-trough variation in capital requirements, which occurs during the 2008 financial crisis for most models. Peak-to-trough variation in minimum capital requirements is a common measure of regulatory procyclicality (Gordy and Howells 2006). Rows 1 and 2 show that the maximum difference increases from Basel 2 to Basel 2.5 due to the introduction of Stressed VaR. The top panels in Figure 2 illustrate that the new stressed period in 2008 combined with the multiplier causes a large jump in capital requirements. This is particularly apparent for HS in Panel (a), where the peak-to-trough variation in Basel 2.5 capital requirements is 58%. While Basel 2.5 was implemented after 2008, the results show that staying with Basel 2.5 can cause a large peak-to-trough variation in capital requirements if a future shock moves the stressed period, although this risk is somewhat mitigated by the higher Basel 2.5 capital requirements. Rows 2 and 3 show that the maximum difference substantially decreases from Basel 2.5 to Basel 3 for most models. This decrease is largest for HS, where the peak-to-trough variation in Basel 3 is only 15.71% as shown in Figure 2d. FZH has the lowest maximum difference across all models for every Basel regime, making the model suitable for banks seeking to minimize large jumps in capital requirements.[19]

In summary, our results show that capital requirements became significantly more stable from Basel 2 to Basel 2.5 due to the introduction of Stressed VaR. Basel 3 will further increase the stability of capital requirements by decreasing nonstandardized and standardized volatility across most models and could also reduce the procyclicality of capital requirements as evidenced by the lower peak-to-trough variation across most models under this regime compared to previous regimes. Procyclicality is particularly concerning to regulators, since it may amplify severe shocks and increase systemic risk. However, the results also show that the models with the most stable capital requirements across regulatory regimes and across variability metrics are generally not the correctly specified models. In particular, FZ models most often result in the least variability although they provide inferior forecasts of the risk measures. The exception to this is when considering the standardized volatility of the Basel 3 capital which is the lowest for the GPD and Hill models. Thus, while banks can minimize both the mean and volatility of Basel 3 capital by using the FZH model, regulators could attempt to incentivize the use of correctly specified models under this regime by penalizing models with high standardized volatility.

19 The FZ1 model had a large shock during the dot-com crisis resulting in high one-day Basel 3 capital requirements.

5. Conclusion

In 2019, the BCBS finalized the Basel 3 regulatory regime, which changes the regulatory measure of market risk from VaR at the 99% confidence level to ES at the 97.5% confidence level and adds new complex calculations based on liquidity horizons and risk factors resulting in the most complex market risk framework to date. As a result, banks are eager to know which models minimize these new capital requirements. Additionally, regulatory supervisors are eager to know if the models incentivized by the new regulations are correctly specified. Finally, regulators are also interested in whether Basel 3 increases the stability and reduces the procyclicality of capital requirements. This article answers these questions using cutting-edge ES backtests by carefully calculating regulatory capital under the new Basel 3 framework for realistic portfolios banks may hold.

Our backtesting results show that HS, RM, and symmetric GARCH models are misspecified and systematically underestimate VaR and ES. In particular, HS is rejected by every traditional backtest and has the worst comparative backtest results. A new class of FZ models due to Patton, Ziegel, and Chen (2019) perform much better, but are rejected by conditional spectral backtests and underperform the class of skewed GARCH models traditionally used in empirical finance in comparative backtests. GARCH FHS, Skewed Student's t, Skewed Generalized t, GPD, and EVT based Hill models are not rejected in any traditional ES backtest. These skewed GARCH models also have the best performance for comparative backtests, often belonging to the MCS and weakly dominating the class of FZ models.

Our regulatory capital results demonstrate that Basel 3 incentivizes banks to choose models that minimize Stressed ES at the 97.5% level, since there is nearly no penalty for using misspecified models. As a result, HS is the model that minimizes capital requirements for Basel 3 despite failing every backtest. The Hybrid FZ model requires the second lowest regulatory capital and leads to the most stable capital requirements whereas correctly specified GARCH models require the highest regulatory capital and have the highest capital requirement variability using most metrics. Our results thus show that, although the proposed regulation under Basel 3 seems to incentivize banks to use models that have more stable capital requirements through time, the models that minimize average capital requirements appear misspecified and produce inferior forecasts of the regulatory risk measures.

Our findings have important implications for the current regulation. In particular, we show that Basel 3 regulation strongly disincentivizes banks from using correctly specified models. Although the same holds for previous regulatory regimes, the changes suggested in Basel 3 makes the relative differences even larger. We identify two possible reasons for this. First of all, it appears that the requirement under Basel 3 to penalize low liquidity assets, that is, assets with long liquidity horizons, additionally in fact disincentivizes banks from using the correctly specified skewed GARCH models due to their consistently high Stressed ES across liquidity horizons. Second, given the low level of the Basel 3 multiplier banks have little incentive to choose conservative and correctly specified models. Thus, if regulator's objective is to also incentivize the use of correctly specified models they would have to reconsider the effect of these changes.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

References

Adrian, T., and H. S. Shin. 2014. Procyclical Leverage and Value-at-Risk. *Review of Financial Studies* 27: 373–403.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9: 203–228.

Balkema, A., and I. de Haan. 1974. Residual Life Time at Great Age. *The Annals of Probability* 2: 792–804.

Barone-Adesi, G., F. Bourgoin, and K. Giannopoulos. 1998. Don't Look Back **11**: 100–104.

Bayer, S., and T. Dimitriadis. 2018. Regression Based Expected Shortfall Backtesting. Forthcoming in Journal of Financial Econometrics.

BCBS. 1996a. "Amendment to the Capital Accord to Incorporate Market Risks," Bank for International Settlements. Available at https://www.bis.org/publ/bcbs24.pdf.

BCBS. 1996b. "Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements." Bank for International Settlements. Available at http://www.bis.org/publ/bcbs22.pdf.

BCBS. 2009. "Revisions to the Basel II Market Risk Framework." Bank for International Settlements. Available at https://www.bis.org/publ/bcbs193.pdf.

BCBS. 2011. "Interpretive Issues with respect to the Revisions to the Market Risk Framework." Bank for International Settlements. Available at http://www.bis.org/publ/bcbs193a.pdf.

BCBS. 2013. "Fundamental Review of the Trading Book: A Revised Market Risk Framework." Bank for International Settlements. Available at https://www.bis.org/publ/bcbs265.pdf.

BCBS. 2014. "Analysis of the Trading Book Hypothetical Portfolio Exercise." Bank for International Settlements. Available at https://www.bis.org/publ/bcbs288.pdf.

BCBS. 2019. "Minimum Capital Requirements for Market Risk." Bank for International Settlements. Available at https://www.bis.org/bcbs/publ/d457.pdf.

Berkowitz, J., P. Christoffersen, and D. Pelletier. 2011. Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science* 57: 2213–2227.

Berkowitz, J., and J. O'brien. 2002. How Accurate Are Value-at-Risk Models at Commercial Banks? *The Journal of Finance* 57: 1093–1111.

Birn, M., O. de Bandt, S. Firestone, M. Gutiérrez Girault, D. Hancock, T. Krogh, H. Mio, et al. 2020. The Costs and Benefits of Bank Capital - a Review of the Literature. *Journal of Risk and Financial Management* 13: 74.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Bollerslev, T., and V. Todorov. 2014. Time-Varying Jump Tails. *Journal of Econometrics* 183: 168–180.

Bollerslev, T., and J. M. Wooldridge. 1992. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric Reviews* 11: 143–172.

Brunnermeier, M. K., and Y. Sannikov. 2016. "The I Theory of Money." NBER Working paper No. 22533. https://www.nber.org/papers/w22533.pdf (accessed August 21, 2020).

Christoffersen, P. 1998. Evaluating Interval Forecasts. *International Economic Review* 39: 841–862.

Christoffersen, P. 2009. "Value-at-Risk Models." In T. Andersen, R. Davis, J. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*. Berlin: Springer, pp. 753-766.

Christoffersen, P., and S. Gonçalves. 2005. Estimation Risk in Financial Risk Management. *The Journal of Risk* 7: 1–28.

Christoffersen, P., J. Hahn, and A. Inoue. 2001. Testing and Comparing Value-at-Risk Measures. *Journal of Empirical Finance* 8: 325–342.

Christoffersen, P., and D. Pelletier. 2004. Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics* 2: 84–108.

Christoffersen, P. F. 2011. *Elements of Financial Risk Managemnet*. Amsterdam: Elsevier.

Creal, D., S. J. Koopman, and A. Lucas. 2013. Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics* 28: 777–795.

Diebold, F. X., and R. S. Mariano. 2002. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 20: 134–144.

Diebold, F. X., T. Schuermann, and J. D. Stroughair. 2000. Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. *The Journal of Risk Finance* 1: 30–35.

EBA. 2012. *EBA Guidlines on Stressed Value at Risk*. Paris: European Banking Authority.

Engle, R. F. 1983. Estimates of the Variance of U.S. Inflation Based on the ARCH Model. *Journal of Money, Credit and Banking* 15: 286–301.

Engle, R. F., and S. Manganelli. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics* 22: 367–381.

Falato, A., D. Iercosan, and F. Zikes. 2019. "Banks as Regulated Traders." Federal Reserve Board Working paper. · https://www.federalreserve.gov/econres/feds/files/2019005pap.pdf (accessed August 21, 2020).

Fissler, T., and J. F. Ziegel. 2016. Higher Order Elicitability and Osband's Principle. *The Annals of Statistics* 44: 1680–1707.

Fissler, T., J. F. Ziegel, and T. Gneiting. 2015. Expected Shortfall is Jointly Elicitable with Value at Risk - Implications for Backtesting. Risk Magazine January: 58-61.

Gencay, R., and F. Selcuk.2004. Extreme Value Theory and Value-at-Risk: Relative Performance in Emerging Markets. *International Journal of Forecasting* 20: 287–303.

Gneiting, T. 2011. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* 106: 746–762.

Gordy, M. B., and B. Howells. 2006. Procyclicality in Basel II: Can We Treat the Disease without Killing the Patient? *Journal of Financial Intermediation* 15: 395–417.

Gordy, M. B., and A. J. Mcneil. 2020. Spectral Backtests of Forecast Distributions with Application to Risk Management. *Journal of Banking & Finance* 116: 105817.

Gourieroux, C. 1997. *ARCH Models and Financial Applications*. Berlin: Springer Series in Statistics.

Hansen, B. 1994. Autoregressive Conditional Density Estimation. *International Economic Review* 35: 705–730.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. The Model Confidence Set. *Econometrica* 79: 453–497.

Harvey, A. C. 2013. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*, Vol. 52. Cambridge: Cambridge University Press.

Heid, F. 2007. The Cyclical Effects of the Basel II Capital Requirements. *Journal of Banking & Finance* 31: 3885–3900.

Hill, B. 1975. A Simple General Approach to Inference about the Tail of a Distribution. *The Annals of Statistics* 3: 1163–1174.

Huisman, R., K. G. Koedijk, C. J. M. Kool, and F. Palm. 2001. Tail-Index Estimates in Small Samples. *Journal of Business & Economic Statistics* 19: 208–216.

Hull, J., and A. White. 1998. Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk. *The Journal of Risk* 1: 5–19.

Morgan, J. P. 1996. *Riskmetrics - Technical Document*. Technical Document, 4th ed. New York, NY: JP Morgan.

Kelly, B., and H. Jiang.2014. Tail Risk and Asset Prices. *Review of Financial Studies* 27: 2841–2871.

Kupiec, P. 1995. Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives* 3: 73–84.

Manganelli, S., and R. F. Engle. 2001. "Value at Risk Models in Finance." Working Paper Series No. 75. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp075.pdf (accessed August 21, 2020).

Mcneil, A., and R. Frey. 2000. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance* 7: 271–300.

Mehta, A., M. Neukirchen, S. Pfetsch, and T. Poppensieker. 2012. "Managing Market Risk: Today and Tomorrow." McKinsey & Company Working Papers on Risk 32. https://www.mckinsey.com/business-functions/risk/our-insights/managing-market-risk-today-and-tomorrow (accessed August 21, 2020).

Nolde, N., and J. F. Ziegel. 2017. Elicitability and Backtesting: Perspectives for Banking Regulation. *The Annals of Applied Statistics* 11: 1833–1874.

O'brien, J., and P. J. Szerszeń. 2017. An Evaluation of Bank Measures for Market Risk before, during and after the Financial Crisis. *Journal of Banking & Finance* 80: 215–234.

Patton, A. J. 2019. Comparing Possibly Misspecified Forecasts. *Journal of Business & Economic Statistics* 1–23.

Patton, A. J., J. F. Ziegel, and R. Chen. 2019. Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk). *Journal of Econometrics* 211: 388–413.

Pérignon, C., Z. Y. Deng, and Z. J. Wang. 2008. Do Banks Overstate Their Value-at-Risk? *Journal of Banking & Finance* 32: 783–794.

Pérignon, C., and D. R. Smith. 2008. A New Approach to Comparing VaR Estimation Methods. *The Journal of Derivatives* 16: 54–66.

Pérignon, C., and D. R. Smith. 2010. The Level and Quality of Value-at-Risk Disclosure by Commercial Banks. *Journal of Banking & Finance* 34: 362–377.

Shim, J. 2013. Bank Capital Buffer and Portfolio Risk: The Influence of Business Cycle and Revenue Eiversification. *Journal of Banking & Finance* 37: 761–772.

Theodossiou, P. 1998. Financial Data and the Skewed Generalized t-Distribution. *Management Science* 44: 1650–1661.

Theodossiou, P. 2018. "Risk Measures for Investment Values and Returns Based on Skewed-Heavy Tailed Distributions: Analytical Derivations and Comparison." SSRN Working paper.

Tsay, R. S. 2010. *Analysis of Financial Time Series*. 3rd ed. Hoboken, NJ: Wiley.

Ziegel, J. F., F. Krüger, A. Jordan, and F. Fasciati. 2017. Murphy Diagrams: Forecast Evaluation of Expected Shortfall. ·Journal of Financial Econometrics 8:95–120.

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 97–127
doi: 10.1093/jjfinec/nbaa030
Article

Dynamic Global Currency Hedging*

Bent Jesper Christensen[1,3,5] and Rasmus Tangsgaard Varneskov[2,3,4]*

[1]Aarhus University, [2]Copenhagen Business School, [3]CREATES, [4]Multi Assets at Nordea Asset Management and [5]Dale T. Mortensen Center

Address correspondence to Rasmus Tangsgaard Varneskov, Department of Finance, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg, Denmark, or e-mail: rtv.fi@cbs.dk

Received June 28, 2019; revised May 31, 2020; editorial decision July 30, 2020; accepted August 6, 2020

Abstract

This article proposes a model for discrete-time currency hedging based on continuous-time movements in portfolio and foreign exchange rate returns. The vector of optimal currency exposures is given by the negative realized regression coefficients from a one-period conditional expectation of the intraperiod quadratic covariation matrix for portfolio and exchange rate returns. Empirical results from an extensive hedging exercise for equity investments illustrate that currency exposures exhibit important time variation, leading to substantial volatility reductions when hedging, without sacrificing returns. A risk-averse investor is willing to pay several hundred annual basis points to switch from existing hedging methods to the proposed dynamic strategies.

Key words: Currency hedging, foreign exchange rates, high-frequency data, quadratic covariation, realized currency beta

JEL classification: C32, C58, G11, G15

Currency hedging is the hottest thing in investing right now.
Article headline, Business Insider UK, March, 2015.

* We are grateful to Kris Jacobs, two anonymous referees, Torben G. Andersen, Ian Dew-Becker, Asbjørn Trolle Hansen, Ravi Jagannathan, Kurt Kongsted, Viktor Todorov, Claus Vorm, and seminar participants at Kellogg School of Management for many useful comments and suggestions. Financial support from the Center for Research in Econometric Analysis of TimE Series (CREATES), funded by the Danish National Research Foundation (DNRF78), is gratefully acknowledged. Both authors are also Danish Finance Institute (DFI) research fellows. R.T.V. is with both Copenhagen Business School and Nordea Asset Management. The views expressed here are those of the authors and not necessarily those of any of the affiliated institutions.

The potential benefits from international diversification have been recognized in the academic finance literature since the seminal work by Grubel (1968) and Levy and Sarnat (1970). Many empirical studies, however, find little, if any, statistically significant diversification benefits from investments across developed countries in more recent times, unless carried out using specific investment styles, for example, by leveraging size, value, and momentum anomalies.[1] A possible explanation for these somewhat discouraging results (seen from the perspective of an investor) is the continuing integration of international financial markets, resulting in higher correlation among international assets and, thereby, reducing the potential for harvesting diversification benefits, see, for example, Longin and Solnik (1995), Bekaert et al. (2009), and Christoffersen et al. (2012).

Most of the aforementioned studies, however, neglect an important component of international investments: The currency exposure implicit in the international equity portfolio holdings. In other words, international investments in a given foreign country are exposed to exchange rate movements, and investors need to decide whether, and how, to hedge this additional source of risk. In practice, investment professionals often choose to hedge a certain fraction of their currency exposure, popular choices being the half hedge and the full hedge. Some studies have analyzed hedging strategies that go beyond simple rule-of-thumb guides. In particular, Glen and Jorion (1993), de Roon et al. (2003), Campbell et al. (2010), Schmittmann (2010), Kroencke et al. (2014), and Opie and Dark (2015) analyze diversification benefits from optimal hedging strategies based on the theory originally proposed by Anderson and Danthine (1981), albeit with mixed empirical results.[2] Whereas they all reject leaving international investments unhedged, the first two studies find no significant evidence that a static, optimal, volatility-minimizing hedging strategy provides diversification benefits beyond what can be achieved by fully hedging international equity investments. However, when implementing a pseudodynamic hedging strategy according to which the optimal currency exposure to a given foreign country depends on the level of its interest rates relative to those in the domestic country, thus mimicking some form of carry trade, they find significant gains over full hedging. Campbell et al. (2010) find that a static volatility-minimizing hedging strategy significantly reduces the risk of international equity investments, compared to the gains from full hedging, and a similar pseudodynamic hedging strategy provides additional diversification benefits. However, the latter are economically modest, judging by their Sharpe ratios (SRs), and often statistically insignificant. Furthermore, their subsample analysis suggests that optimal currency exposures are quite sensitive to the specific sample under consideration. Similar results are obtained by Schmittmann (2010) and Opie and Dark (2015) from different countries' perspectives and across various horizons, corroborating the conclusions. Finally, Kroencke et al. (2014) take a deeper look into the diversification benefits from using traditional currency investment styles such as carry trade, momentum, and value strategies in said framework, thus promoting the pseudodynamic aspect of the optimal hedging strategies. They find significant diversification benefits, particularly when including foreign exchange rates for countries outside of the G10. However, the diversification benefits from their two-step

1 This includes the mean–variance analyses in, among others, Britten-Jones (1999), Errunza et al. (1999), Eun et al. (2008), Eun et al. (2010), and Fama and French (2012).

2 Optimal in this setting is to be understood in a mean–variance sense, that is, as the solution to a quadratic optimization problem for an investor seeking to maximize her risk-return tradeoff.

procedure stem from the speculative asset allocation in the second step, not from the hedging itself, for which their results resemble those of Campbell et al. (2010). Thus, they reflect the profitability of the three currency investment styles over the last 30–40 years.

In this article, we reconsider the first step, that is, we focus on enhancing the diversification benefits from volatility-minimizing hedging strategies, conditionally on a given portfolio, not on speculative currency investments.[3] To set the stage, we highlight two important aspects of previous approaches to currency hedging that demand further attention. First, all aforementioned studies of optimal currency exposure rely on the theoretical results from Anderson and Danthine (1981), who assume that asset prices are observed at the same frequency as that at which the investor rebalances her portfolio, that is, the frequency at which hedging decisions are made. This implies, for example, that if an investor rebalances her portfolio at a monthly frequency, then movements in asset prices occur at the monthly frequency, as well. Hence, this approach neglects all information from asset price movements occurring at higher frequencies, for example, daily or intradaily. Second, the hedging strategies are often studied in their "static," or unconditional, form, suggesting that optimal, volatility-minimizing currency exposures should be constant, often over a time span of 30–40 years, and estimated using full sample information. When such hedging strategies are given a time-varying flavor, it is by conditioning on variables related to currency investment styles, such as past interest rate differentials. The latter approach is labeled "pseudodynamic" for two reasons: (i) All intertemporal movements in the optimal currency exposures are determined by slowly varying conditioning variables. Hence, no traditional time series modeling (ARMA, GARCH, or stochastic volatility) is actually performed. (ii) The implementation of the hedging strategies is often in-sample, that is, the functional link to the interest differentials is estimated using full sample information, then used for conditional hedging decisions.[4] Hence, neither the static nor the pseudodynamic implementation of the optimal hedging strategies is designed for real-time investment decisions, and they provide inadequate descriptions of the dynamic properties of optimal currency exposures. The main exception to these caveats is Opie and Dark (2015), who perform an out-of-sample analysis in which they compare rule-of-thumb hedges to a static optimal hedging strategy and a dynamic strategy based on a multivariate VAR-GARCH model, both implemented with rolling windows to estimate the currency exposures. The models are based on daily data, and rebalancing occurs if the changes in exposure generate increases in utility. Similarly to Campbell et al. (2010), they show that the two optimal strategies produce the lowest portfolio volatility, but also display statistically indistinguishable performance. That is, they find no additional improvements from actual dynamic

3 A related body of work considers optimal hedging of spot exchange rate risk using equivalent currency futures contracts in *conditional* frameworks resemblant of that developed by Anderson and Danthine (1981), see, for example, Baillie and Bollerslev (1989). However, this problem is distinct from the present setting of strategic utilization of currency exposures to improve the performance of an existing portfolio.

4 Note that the implementation of the two-step optimal hedging strategies using currency investment styles in Kroencke et al. (2014) does not suffer from the caveat in (ii), as the conditioning variables for the investment styles are contemporaneously available when the investor rebalances her portfolio. However, the caveat in (i) still describes their analysis. Similar comments apply to the robustness check in Campbell et al. (2010, Section 6).

modeling. However, despite addressing the second caveat, their framework, as well as analysis, does not explicitly treat the issue of sampling versus rebalancing frequency, but rather implicitly via a utility function.

This article directly addresses both caveats by introducing a new economic model for *discrete-time* currency hedging that not only allows the assets of interest—the portfolio and foreign currencies—to exhibit *within-period* movements, but actively utilizes additional "infill" information to construct accurate measures of optimal, volatility-minimizing, currency exposures. In particular, the latter are shown to be the negative realized regression coefficients from a one-period conditional expectation of the intraperiod quadratic covariation (QC) matrix for portfolio and currency returns, labeled the *realized currency betas*. The model, hence, facilitates dynamic hedging strategies, depending exclusively on the dynamic evolution of the ex post QC matrix. This has the strong theoretical implication that interest rate differentials have *no* asymptotic impact on the optimal currency demands for a given international portfolio, in stark contrast with existing hedging theory, for example, Anderson and Danthine (1981), Glen and Jorion (1993), and Campbell et al. (2010).[5] Moreover, as the proposed strategies do *not* rely on information about local trends in currencies in their construction, they are notably different from traditional currency investment styles, such as carry, momentum, and value trading. From a theoretical perspective, the development of the realized currency beta hedging framework involves establishing new results for optimal currency exposures based on the notion of QC measures and infill asymptotic limits. From a practical perspective, the theory suggests that an investor should sample as frequently as possible within fixed time intervals between portfolio rebalances to construct accurate estimates the QC matrix and model its dynamics. Hence, this article proposes to implement the new hedging strategies using modern, yet simple, nonparametric techniques to accurately measure and dynamically model historical QC matrices, imposing only few parametric restrictions on the underlying processes.

The new dynamic hedging strategies are analyzed in an extensive empirical exercise, covering different international equity portfolios, as well as a balanced fixed income–equity portfolio, and different rebalancing horizons, sampling frequencies, and currency investment universes (sets of foreign currencies available for hedging purposes). This produces several new and striking results that may be summarized as follows: (i) The optimal currency exposures display substantial time-variation, which is tied to important economic events such as the 2008–2009 global financial crisis, the European sovereign debt crisis, and the "bloody Christmas" of 2018 global stock sell-off. The Swiss Franc and Japanese Yen are, on average, the most important hedging currencies, and the Canadian Dollar and Euro the main funding currencies. (ii) The proposed dynamic hedging strategies produce statistically significant, as well as economically substantial, volatility reductions for *all* baseline portfolios, compared to fully hedging currency exposure, as well as to existing, static and pseudodynamic, approaches to optimal hedging. (iii) These volatility reductions come without sacrificing returns, especially when implemented using intradaily data to estimate the QC between assets, thereby delivering SRs that are 61% larger than key benchmarks. (iv) The estimated economic gains to the new hedging strategies are equally

5 Even the dynamic VAR-GARCH implementation of currency hedging strategies in Opie and Dark (2015) depends implicitly on interests rate differentials via the conditional mean specification for currency excess returns.

substantial, being 120–465 annual basis points (BP) after transaction costs over full hedging—depending on baseline portfolio and investor risk-aversion—and 120–520 annual BP over existing hedging procedures. (v) The quality of the input QC measure seems to be more important for designing profitable realized currency beta hedging strategies than the dynamic model specification or an expansion of the set of hedging currencies, with the former being the second most important feature. (vi) The currency overlay behind the dynamic realized currency beta investment strategy is negatively correlated with the FX carry trade, and only modestly correlated with momentum and value investments. Interestingly, the empirical analysis strongly suggests that the unwinding of carry trades, at least partially, fund the strong performance of the proposed dynamic hedging strategy during the global financial crisis of 2008–2009.

All of the findings (i)–(vi) are new to the literature on global currency hedging. In particular, the empirical hedging results go well beyond those in existing studies, such as Glen and Jorion (1993), Campbell et al. (2010), and Opie and Dark (2015), by not only showing how dynamic hedging strategies can be designed to obtain better volatility *and* risk-return trade-offs than full hedging and static optimal procedures, but also by estimating the economic gains from such strategies to a risk-averse investor, documenting important time-variation in optimal currency exposures, showing how this links to key economic events, and by providing results that speak to the relative importance of currency universe, dynamic model, and sampling frequency. Moreover, this article is the first to leverage intradaily data to construct currency hedging strategies, and this feature is paramount for designing procedures that deliver significant volatility reductions and superior economic performance.

The finding that dynamic hedging strategies based on intradaily rather than daily data improves portfolio performance is consistent with Fleming, Kirby, and Ostdiek (2001, 2003), who study dynamic asset allocation between S&P 500, Treasury bond, and gold futures.[6] However, in addition to the present analysis being one of hedging rather than asset allocation, the elicitation of gains from intradaily data in the international investments and currency trading case is more challenging than in their single-country analysis, due to assets being traded on different exchanges with only partially overlapping trading hours. Our results demonstrating that dynamic rather than static modeling of exchange rate covariances leads to economic gains for a risk-averse investor are also consistent with Della Corte et al. (2009), who analyze asset allocation between fixed income and currencies by applying different univariate dynamic models to monthly data. In contrast with these studies, the problem of currency hedging can be viewed as a constrained, or conditional, asset allocation exercise in which currency exposures are selected for a given baseline portfolio, unlike their unconstrained approaches. Hence, it mirrors the problem faced by many institutional

6 It is also consistent with Andersen et al. (2003), who consider VaR estimation using 30-minute returns on two currencies, and with Chiriac and Voev (2011), and Varneskov and Voev (2013), who carry out mean–variance analyses using intradaily data on DJIA stocks. However, none of these studies considers the interaction between equity investments and currency exposures, nor do they use intradaily data to design currency hedging strategies. Finally, it is consistent with Christoffersen and Diebold (2000), who show that volatility forecastability is important for risk management, see also Andersen et al. (2013) for a comprehensive review.

investors where one investment team takes a strategy, or portfolio, of another as given, and carries out hedging to improve its performance.

Even though individual currencies have traditionally been viewed as poor investment vehicles with low return and high volatility, there has been a recent surge of academic papers in a separate strand of the literature on exchange rate modeling, showing that systematic application of traditional currency trading, in particular carry trade, momentum, and value investments, may be highly profitable, even on a risk-adjusted basis, see, for example, the recent contributions by Lustig and Verdelhan (2007), Brunnermeier et al. (2009), Burnside et al. (2011), Lustig et al. (2011), Menkhoff et al. (2012a, 2012b), Moskowitz et al. (2012), Asness et al. (2013), and many references therein. The dynamic hedging strategies proposed in this article similarly constitute systematic trading opportunities in currencies. However, they are designed with the specific purpose of improving the performance of an already existing baseline portfolio. Moreover, as the realized currency betas are asymptotically invariant to changes in interest rate differentials and only use information about the covariance between foreign exchange rate and portfolio returns in their construction, that is, no information about local trends in the former, they are notably different from traditional investment styles. In fact, their favorable correlation properties, noted in (vi), suggest not only that the proposed dynamic strategies may provide a hedge for the carry trade, but also that there may be intriguing opportunities to combine the four different methods in designing tactical foreign currency exchange rate trading.

The outline of the article is as follows. Section 1 introduces the new economic model and the assumptions, then derives the theoretical foundation for the proposed dynamic currency hedging strategies. Section 2 discusses the nonparametric implementation procedure. Section 3 introduces the data and provides empirical evidence of time variation in the optimal currency exposures. The risk-return performance and economic benefits from implementing different hedging strategies are examined in Section 4. Section 5 relates the returns to the dynamic hedging strategies to those from traditional currency investment styles, and Section 6 concludes. The Online Appendix provides additional theory, proofs of the theoretical results, various robustness checks, and implementation details.

1 The Dynamic Modeling Framework

This section introduces a model for *discrete-time hedging* based on *continuous-time* within-period movements in the underlying portfolio and foreign exchange rate returns. The model is intended to capture the decision problem of an investor who rebalances, or rehedges, her portfolio in fixed time intervals, but observes both portfolio and exchange rate movements within each interval. Optimal currency exposures are established using *infill* asymptotic theory for a general class of continuous-time price processes. The discrete-time framework follows along the lines of Anderson and Danthine (1981) and Campbell et al. (2010). However, as shown below, allowing for the continuous-time within-period movements in the processes of interest not only generalizes the framework considerably—it also simplifies the optimal, volatility-minimizing, hedging decision.

1.1 Discrete-Time Decision-Making

Suppose that at each discrete point in time $t = 1, 2, \ldots, T$, an investor holds a position $w_{c,t}$ in country c's equities, $c = 0, \ldots, C$, from time t until $t+1$, when the holding pays a gross continuously compounded return of $R_{c,t+1}$.[7] For simplicity, let $c = 0$ index the home country, which is assumed to be the United States, and let $S_{c,t+1}$ be the corresponding time $t+1$ spot exchange rate quoted in USD per foreign currency unit. In this setting, the U.S. investor earns an unhedged return $R^u_{c,t+1} = R_{c,t+1}S_{c,t+1}/S_{c,t}$ on her country c investment. To hedge the latter against currency risk, the investor buys a holding of the one-period forward exchange rate $F_{c,t}$, equivalently measured in USD per foreign currency unit, at time t in country c. Let $\theta_{c,t}$ be the dollar value of this holding per USD invested in the equity portfolio. Thus, the investor gets to exchange $\theta_{c,t}/S_{c,t}$ units of $R_{c,t+1}w_{c,t}/S_{c,t}$ back into USD at the exchange rate $F_{c,t}$, and the remaining $(R_{c,t+1}w_{c,t}/S_{c,t} - \theta_{c,t}/S_{c,t})$ units of foreign currency at the spot exchange rate $S_{c,t+1}$. This suggests writing the hedged portfolio return as

$$R^h_{t+1} = \sum_{c=0}^{C} w_{c,t}R^u_{c,t+1} + \sum_{c=0}^{C} \theta_{c,t}\frac{F_{c,t}}{S_{c,t}} - \sum_{c=0}^{C} \theta_{c,t}\frac{S_{c,t+1}}{S_{c,t}}. \tag{1}$$

Notice that the choice of domestic hedge ratio, $\theta_{0,t}$, is arbitrary, since $S_{0,t} = F_{0,t} = 1$, for all t. Hence, for ease of exposition, the hedge ratios are normalized to sum up to one, implying that

$$\sum_{c=0}^{C} w_{c,t} = 1, \quad \theta_{0,t} = 1 - \sum_{c=1}^{C} \theta_{c,t}, \tag{2}$$

for all t. Maintaining an assumption of absence of arbitrage throughout, it follows by covered interest rate parity (CIP) that $F_{c,t}/S_{c,t} = (1 + I_{0,t})/(1 + I_{c,t})$, where $I_{c,t}$ denotes the nominal short-term risk-free interest rate. This identity may be inserted in Equation (1).

The form of the portfolio return in Equation (1) also allows for speculative positions in exchange rates if, for example, the currency demand $\theta_{c,t}$ is driven by, for example, a model for local trends in $S_{c,t}$, regardless of its correlation with the portfolio return. Hence, to avoid confusion going forward, the label "hedging" in this article signifies that currency demands are determined with the explicit objective of reducing the risk of the portfolio return, thus seeking currencies with favorable correlation properties. In other words, currency *hedgers* and *speculators* are distinguished according to whether they emphasize correlation properties or local trends, respectively, when selecting foreign exchange rate exposure.

1.2 Intraperiod Dynamics

Suppose that the processes of interest—equities, currencies, and bonds—are defined on a filtered probability space, $(\Omega, \mathcal{F}, (\mathcal{F}_{t,\tau}), \mathbb{P})$, where $\tau \in [t, t+1]$ is the within-period time indicator. In the absence of arbitrage, prices are assumed to follow semimartingales, for

7 The exposition is laid out for equities. This may without loss of generality, however, be changed to other assets held in foreign countries, such as corporate bonds, commodities, derivatives, etc., as long as the assumptions on the assets, as outlined below, are satisfied. A balanced bond–equity portfolio is examined in the empirical analysis.

example, Back (1991). Hence, denote by $P_{c,\tau}$ the price of the equity holdings in country c, measured in local currency, and $B_{c,\tau}$ the price of a country c-denominated riskless bond. Then, for $c = 0, \ldots, C$, the system of equity, currency, and bond prices obeys

$$dP_{c,\tau}/P_{c,\tau} = \mu_{c,\tau}d\tau + \sigma_{c,\tau}dW_{c,\tau}, \tag{3}$$

$$dS_{c,\tau}/S_{c,\tau} = \alpha_{c,\tau}d\tau + \varphi_{c,\tau}dY_{c,\tau}, \tag{4}$$

$$dB_{c,\tau}/B_{c,\tau} = \lambda_{c,\tau}d\tau, \tag{5}$$

in which $(\mu_{c,\tau}, \alpha_{c,\tau})$ and $(\sigma_{c,\tau}, \varphi_{c,\tau})$ capture the within-period drift and stochastic volatility of equity and currency returns, $W_{c,\tau}$ and $Y_{c,\tau}$ are standard Brownian motions adapted to $(\mathcal{F}_{t,\tau})$, and $\lambda_{c,\tau}$ models the instantaneous return from holding a short-term riskless bond.[8] Moreover, for $c \neq k$, we define the QCs $d[W_c, W_k]_\tau = \sigma_{c,k,\tau}d\tau$ and $d[Y_c, Y_k]_\tau = \varphi_{c,k,\tau}d\tau$. Finally, let $d[W_c, Y_k]_\tau = \psi_{c,k,\tau}d\tau$ for all c, k. The theoretical analysis, then, necessitates additional structure on the system:

Assumption 1. *For all $c, k \in \{0, \ldots, C\}$, the components of (3)–(5) satisfy:*

a. *$\mu_{c,\tau}$, $\alpha_{c,\tau}$, and $\lambda_{c,\tau}$ are $\mathcal{F}_{t,\tau}$-predictable and locally bounded;*

b. *$\sigma_{c,\tau}$ and $\varphi_{c,\tau}$ are $\mathcal{F}_{t,\tau}$-adapted, locally bounded, càdlàg, and strictly greater than zero;*

c. *$\sigma_{c,k,\tau}$, $\varphi_{c,k,\tau}$, and $\psi_{c,k,\tau}$ are $\mathcal{F}_{t,\tau}$-adapted, locally bounded, and càdlàg.*

The present setting generalizes the previously developed framework for deriving optimal currency exposure in Anderson and Danthine (1981), Glen and Jorion (1993), de Roon et al. (2003), and Campbell et al. (2010), by allowing for stochastic drift and volatility, as well as intraperiod movements in equites and currencies. The latter are assumed to belong to a general class of continuous Brownian semimartingales, which, again, is commonly used in the literature on high-frequency volatility and covariance estimation, since it nests many continuous-time models in financial economics.[9] For example, the class accommodates the widely documented presence of leverage effects, that is, nonzero correlation between innovations in the price process and the stochastic volatility process. The modeling system implies that in a given time interval, $[t, t+1]$, between the rebalancing times of the portfolio of equities, currencies, and bonds, asset prices are allowed to evolve according to intraperiod trajectories, which will be important for determining the investor's optimal, volatility-minimizing, currency position.

Remark 1. *While it is convenient to work with locally riskless bonds, it is important to note that the analytical results below are not contingent on a diffusive component being absent in Equation (5). All our results are asymptotically invariant to replacing the latter with $dB_{c,\tau}/B_{c,\tau} = \lambda_{c,\tau}d\tau + \varsigma_{c,\tau,\epsilon}dZ_{c,\tau}$, where $dZ_{c,\tau}$ is a standard Brownian motion, which may have nontrivial QC with $dW_{k,\tau}$ and $dY_{k,\tau}$, and $\varsigma_{c,\tau,\epsilon} = \varsigma_{c,\tau} \times (d\tau)^\epsilon$ captures stochastic volatility, with $\epsilon > 0$ and $\varsigma_{c,\tau}$ satisfying conditions similar to Assumption 1(b). This models*

8 The time t subscript is dropped for notational simplicity when describing the intraperiod price system (3)–(5), since the representation is valid for all intervals, with $\tau \in [t, t+1]$, $t = 1, \ldots, T$.

9 See, for example, Andersen and Benzoni (2009) and Andersen et al. (2006, 2013) for reviews.

the volatilities of short-term bonds as an order of magnitude smaller than the corresponding volatilities for equity and currency prices. For $\epsilon \to \infty$, *Equation (5) is recovered.*

Remark 2. *The vector price system (3)–(5) may be extended to include jumps. This only leads to minor changes in the interpretation of the results. The role of jumps is discussed in Section A of the Online Appendix, where the theoretical results, provided below, are also extended.*

Before deriving the optimal currency exposures, it is important to characterize the path of the hedged portfolio return at each time $\tau \in [t, t+1]$. Hence, with V_τ denoting the value of the hedged portfolio at time τ, the evolution of its instantaneous return may be described using Equations (3)–(5),

$$\frac{dV_\tau}{V_\tau} = \sum_{c=0}^{C} w_{c,t} \frac{d(P_{c,\tau}S_{c,\tau})}{P_{c,\tau}S_{c,\tau}} + \sum_{c=0}^{C} \theta_{c,t} \frac{d(B_{0,\tau}/B_{c,\tau})}{B_{0,\tau}/B_{c,\tau}} - \sum_{c=0}^{C} \theta_{c,t} \frac{dS_{c,\tau}}{S_{c,\tau}}. \tag{6}$$

As in Campbell et al. (2010), it simplifies the problem to work in logarithms and use matrix notation. Hence, let $r^h_{t+1} = \ln(R^h_{t+1})$ and $x_{c,\tau} = \ln(X_{c,\tau})$ for $X \in \{P, S, V\}$. Similarly, let $\boldsymbol{w}_t = (w_{0,t}, \ldots, w_{C,t})'$ be the $(C+1) \times 1$ vector of portfolio weights, $\boldsymbol{\Theta}_t = (\theta_{0,t}, \ldots, \theta_{C,t})'$ the corresponding $(C+1)$-dimensional vector of currency hedging positions, $\boldsymbol{x}_\tau = (x_{0,\tau}, \ldots, x_{C,\tau})'$ for $x \in \{p, s, \lambda\}$, and $\boldsymbol{\lambda}_{0,\tau} = \boldsymbol{\iota}\lambda_{0,\tau}$, with $\boldsymbol{\iota}$ a $(C+1) \times 1$ vector of ones. Furthermore, to explicitly capture the fact that an investor can alter her currency exposure by lending and borrowing (going long or short in bonds or forward contracts), define $\boldsymbol{\beta}_t \equiv (\beta_{0,t}, \ldots, \beta_{C,t})' = \boldsymbol{w}_t - \boldsymbol{\Theta}_t$ as the net exposures to the foreign currencies. For example, $\beta_{c,t} = 0$ corresponds to having a fully hedged position in country c's equities, and $\beta_{c,t} = w_{c,t}$ to leaving the exposure completely unhedged. In general, $\beta_{c,t} > 0$ implies that the investor demands exposure to currency c and, equivalently, she wants to be underexposed if $\beta_{c,t} < 0$. Note that Equation (2) implies $\boldsymbol{\beta}_t'\boldsymbol{\iota} = 0$, that is, the dynamic currency hedging portfolio is a zero investment, meaning that all long positions in currencies are financed by shorting bonds in funding currencies, similarly as in traditional currency investment styles. Finally, a regularity condition is imposed on the elements of the vectors $\boldsymbol{w}_t$ and $\boldsymbol{\Theta}_t$ to simplify the further theoretical analysis.

Assumption 2. *For all* $t = 1, \ldots, T$, $\sup_{c=0,\ldots,C} |w_{c,t}| + \sup_{c=0,\ldots,C} |\theta_{c,t}| < \infty$.

Assumption 2 innocuously states that both the equity portfolio weight in and currency exposure to country c, $c = 0, \ldots, C$, must be finite.[10] The following proposition provides a representation result for the *within-period* currency hedged log-returns, dv_τ.[11]

10 Strictly speaking, the condition $\sup_{c=0,\ldots,C} |\theta_{c,t}| < \infty$ should be shown endogenously in the dynamic model, since $\theta_{c,t}$ will depend on the components of the intraperiod price system (3)–(5). However, by assuming it from the outset, rather than showing it endogenously, the proofs of Propositions 1–4 may be shortened considerably, without loss of intuition.

11 We will be using the nomenclature $o_p(d\tau)$ to describe higher-order terms of the form $(d\tau)^2$, $d\tau \times dW_{c,\tau}$, $d\tau \times dY_{c,\tau}$, etc., which have no asymptotic impact in the further analysis below.

Proposition 1. *Suppose the representation (6) and Assumptions 1 and 2 hold. Then*

$$dv_\tau = \boldsymbol{w}_t'(d\boldsymbol{p}_\tau + \boldsymbol{\lambda}_{0,\tau}d\tau - \boldsymbol{\lambda}_\tau d\tau) + \boldsymbol{\beta}_t'(d\boldsymbol{s}_\tau - \boldsymbol{\lambda}_{0,\tau}d\tau + \boldsymbol{\lambda}_\tau d\tau) + \Sigma_\tau^h d\tau + o_p(d\tau),$$

where Σ_τ^h is $\mathcal{F}_\tau$-adapted, locally bounded, and càdlàg.

Proposition 1, similarly to the representation in Campbell et al. (2010, Equation (1)), provides a decomposition of the hedged log-return into three components; the first is the instantaneous excess return on a *fully* hedged portfolio; the second term represents the instantaneous excess return on currencies, which depends on the selected exposure, $\boldsymbol{\beta}_t$; and the last term is a Jensen's inequality correction. However, unlike in the corresponding framework in Anderson and Danthine (1981) and Campbell et al. (2010), the instantaneous log-return on the hedged portfolio is allowed to evolve stochastically in the interval $\tau \in [t, t+1]$, implying that the one-period log-return may be written

$$r_{t+1}^h = \int_t^{t+1} dv_\tau, \quad t = 1, \ldots, T, \tag{7}$$

thus formally providing a link between their framework and ours. Equation (7) suggests that the one-period log-return on a hedged portfolio may be interpreted as the sum of returns at a higher frequency. For an investor with a monthly investment horizon, this could, for example, be a sum of daily log-returns.

1.3 Optimal Dynamic Currency Exposure

The optimal dynamic selection of currency exposure requires the choice of an appropriate objective function. Usually, in portfolio selection problems, this involves choosing the portfolio weights to minimize portfolio variance subject to certain constraints. Similarly to the one-period log-return Equation (7), which is measured by cumulating returns at higher frequency, a measure of its variance must also reflect the stochastic intraperiod movements in dv_τ. In this setting, quadratic variation (QV) offers such a variability measure, see, for example, Andersen et al. (2010). Formally, suppose the intraperiod hedged log-return dv_τ is observed on a discrete partitioning τ_i of the time interval, $t = \tau_0 < \tau_1 < \ldots < \tau_n = t+1$. The QV of r_{t+1}^h is, then, defined as

$$[dv_\tau]_{t+1} \equiv \text{plim}_{n\to\infty} \sum_{i=1}^{n} (v_{\tau_i} - v_{\tau_{i-1}})^2 = \lim_{h\to 0} \int_t^{t+1} \mathbb{E}[\mathcal{M}(dv_\tau)^2 | \mathcal{F}_{t,\tau-h}] d\tau, \tag{8}$$

with $\mathcal{M}(\cdot)$ isolating the martingale component of dv_τ, for $\sup_i\{\tau_{i+1} - \tau_i\} \to 0$ as $n \to \infty$, see, for example, Jacod and Shiryaev (2003).[12] QV captures the entire realized ex post variation of the hedged log-returns, and its use will simplify the computations of the optimal currency exposures via the next result.

12 The QC between two appropriately dimensioned vector processes $\boldsymbol{x}_{\tau_i}$ and $\boldsymbol{y}_{\tau_i}$, for a similar partition of the sample $\tau_i \in [t, t+1]$, $i = 0, \ldots, n$, is analogously defined as $[\boldsymbol{x}, \boldsymbol{y}]_{t+1}$, that is, as the probability limit of a sum of outer products of their increments as the distance between observations tends to zero.

Proposition 2. *Suppose the conditions of Proposition 1 hold. Then, as* $n \to \infty$,

$$[dv_\tau]_{t+1} = [\boldsymbol{w}_t' d\boldsymbol{p}_\tau + \boldsymbol{\beta}_t' d\boldsymbol{s}_\tau]_{t+1}.$$

Proposition 2 shows that the QV of the hedged log-return depends *only* on the QVs of the fully hedged log-return and the total currency exposure return, as well as on their QC. Hence, there is no impact from movements in nominal short-term risk-free interest rate differentials nor from the Jensen's inequality induced term, $\Sigma_\tau^h d\tau$. This distinct advantage of the proposed *within-period* model for equities, currencies, and bonds is due to the fact that drift components have no asymptotic impact on QV in the *infill* asymptotic limit. As a result, Proposition 2 provides a variance measure that contrasts starkly with the corresponding long-span variance measure used for the development of the existing currency hedging theory by Anderson and Danthine (1981), and applied in Glen and Jorion (1993), de Roon et al. (2003), and Campbell et al. (2010), and which depends on period-by-period movements in short-term interest rate differentials.

Since the vector of dynamic net currency exposures, $\boldsymbol{\beta}_t$, represents a zero-investment portfolio, it suffices to determine the $C \times 1$ vector of foreign currency exposures $\tilde{\boldsymbol{\beta}}_t = (\beta_{1,t}, \ldots, \beta_{C,t})'$, which spans the unique elements of $\boldsymbol{\beta}_t$. Formally, and consistently with our distinction between foreign exchange hedgers and speculators, exposures are selected to minimize the one-period conditional QV of the hedged log-return, that is, as

$$\tilde{\boldsymbol{\beta}}_t^* = \underset{\boldsymbol{\beta}_t | \boldsymbol{w}_t}{\operatorname{argmin}}\, \ell_t(\boldsymbol{\beta}_t, \boldsymbol{w}_t), \quad \ell_t(\boldsymbol{\beta}_t, \boldsymbol{w}_t) = \frac{1}{2}\mathbb{E}_t[[dv_\tau]_{t+1}]. \tag{9}$$

Before stating the optimality result, let $\tilde{\boldsymbol{s}}_\tau = (s_{1,\tau}, \ldots, s_{C,\tau})'$ denote the vector of currencies corresponding to the unique exposures $\tilde{\boldsymbol{\beta}}_t$. Then, the following proposition solves Equation (9).

Proposition 3. *Suppose the conditions of Proposition 2 hold, and that* $\mathbb{E}_t[[d\tilde{\boldsymbol{s}}_\tau]_{t+1}]$ *is positive definite for all* $t = 1, \ldots, T$. *Then the limiting, unique, optimal currency exposures are determined by*

$$\tilde{\boldsymbol{\beta}}_t^* = -\mathbb{E}_t[[d\tilde{\boldsymbol{s}}_\tau]_{t+1}]^{-1}\mathbb{E}_t[[d\tilde{\boldsymbol{s}}_\tau, \boldsymbol{w}_t' d\boldsymbol{p}_\tau]_{t+1}].$$

Proposition 3 demonstrates that the vector of optimal currency exposures is the negative vector of realized regression coefficients from an implicit projection of the fully hedged log-return on the vector of foreign exchange rate innovations, which is embedded in the one-period conditional expectation of the QC matrix. This former is labeled the *realized currency beta*, in analogy with the market exposure measured by the CAPM beta. However, it is stressed that while the market beta reflects the uncertainty of a given asset in terms of its sensitivity to market movements, the realized currency beta reflects the hedging

potential from having active, and systematic, currency exposure in a *given* equity portfolio and is, as a result, *not* a deep characteristic of a currency.

In addition, Proposition 3 suggests that realized currency betas may be computed dynamically using only *within-period* equity and foreign exchange rate data by first obtaining a time series of their QC estimates, and then specifying an appropriate dynamic model for these, to obtain one-step-ahead conditional expectations. This is a highly desirable property, since it implies that the optimal currency exposures are not only asymptotically invariant to short-term interest rate differentials, but also to the validity of CIP, which is otherwise used to substitute out forward rates with interest rate differentials in Equations (1) and (6).[13] Importantly, this invariance separates the optimal currency exposures from the popular carry trade investments, which are designed with long positions in baskets of currencies with high short-term interest rates and short in baskets of currencies with low interest rates, resting on the failure of the uncovered interest rate parity. The resulting realized currency beta hedging strategies may, thus, be viewed as an alternative to traditional currency investments, such as carry, momentum, and value trading strategies, which rely solely on the modeling of local trends, rather than cross-asset covariances. Empirical comparisons are made in Section 5.

Finally, focusing solely on volatility reduction in the objective function (9) has two additional advantages. First, it mirrors the problem faced by many institutional investors, where one investment team takes a trading strategy as given (here, an equity portfolio), then executes a hedging procedure to improve its risk profile. Second, from an econometric perspective, Engle and Colacito (2006) show that the economic value of time-varying covariances can only be consistently measured in a mean–variance setting by the minimum variance portfolio. Hence, the objective function (9) facilitates consistent evaluation of the proposed intraperiod model for currency hedging.

Remark 3. *Although the exposition is given from the perspective of a U.S. investor, it is important to note that the realized currency betas are dynamically invariant to base currency. This implies that, for example, a U.K. investor with the same equity portfolio will be choosing identical optimal currency exposures. This invariance result is formally shown in the Online Appendix.*

2 Estimating Optimal Currency Exposures

Dynamic implementation of the proposed realized currency beta hedging strategy requires both estimation of the latent QC matrix over each discrete time interval between portfolio rebalances, and subsequent dynamic modeling of the covariance matrices. Hence, two different nonparametric approaches to QC estimation, which may be applied to data sampled at different frequencies, are discussed first. Second, a simple

13 While Akram et al. (2008) find that CIP holds approximately at daily or lower frequencies, Du et al. (2018) find persistent deviations from the no-arbitrage condition. Although realized currency betas are invariant to CIP, the latter will have an impact on whether to apply forward rates or interest rate differentials to evaluate the return performance of optimal hedging strategies ex post. In the empirical analysis below, the investor is assumed to trade FX forwards, and forward rates are applied, to avoid concerns about CIP violations.

filtering procedure for the construction of one-period-ahead conditional expectations of the QC matrix is then detailed.

2.1 Measuring QC

Suppose that the vector $\boldsymbol{x}_{\tau_i} = (\boldsymbol{w}_t' d\boldsymbol{p}_{\tau_i}, d\tilde{\boldsymbol{s}}_{\tau_i}')'$ is observed at the $n+1$ discrete time points from the portfolio rebalancing at t to the next, that is, at $\tau_i \in [t, t+1]$, $i = 0, \ldots, n$, then the realized covariance (RC) estimator, introduced by Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004), represents the empirical approximation to the computations (8). Formally, the estimator is defined as

$$RC(\boldsymbol{x}) = \sum_{i=1}^{n} \Delta \boldsymbol{x}_{\tau_i} \Delta \boldsymbol{x}_{\tau_i}', \tag{10}$$

where $\Delta = 1 - L$ is the usual first difference operator. Under mild conditions on the vector price system in Equations (3)–(5), $RC(\boldsymbol{x}) \mathbb{P}[\Delta \boldsymbol{x}]_{t+1}$ for $\sup_i\{\tau_{i+1} - \tau_i\} \to 0$ as $n \to \infty$. Its associated central limit theory demonstrates that convergence occurs at the optimal rate, $n^{1/2}$, to a mixed Gaussian distribution. Implicit in these statements, however, is that the individual entries in $\boldsymbol{x}_{\tau_i}$ are observed synchronously and without measurement errors. This approximation may not be too damaging if the rebalancing horizon is, for example, weekly or monthly, and the intraperiod observations are recorded daily or even intradaily at sufficiently sparse intervals.[14] If the data are sampled intradaily at higher frequencies, on the contrary, market microstructure (MMS) effects and nonsynchronicity-related errors drive a wedge between the observed equity prices and exchange rates and their theoretical counterparts, leading the individual entries of standard covariance matrix estimators such as RC to diverge. Hence, if the data are available intradaily at frequencies higher than the conventional five-minute rule-of-thumb, it is pertinent to use an estimator that actively mitigates the impact from these measurement errors while maintaining good efficiency properties. A class of estimators fitting these requirements is the flat-top realized kernels, proposed by Varneskov (2016, 2017).

The notion of measurement errors may be quantified as follows: Let the observable, synchronized, intradaily observations follow an additive noise model of the form $\boldsymbol{y}_{\tau_i} = \boldsymbol{x}_{\tau_i} + \boldsymbol{u}_{\tau_i}$, where $\boldsymbol{u}_{\tau_i}$ summarizes the effects from an array of market imperfections, including synchronization errors, and is referred to as MMS noise.[15] Next, let $\Gamma_h(\boldsymbol{y}) = \sum_{i=|h|+1}^{n} \Delta \boldsymbol{y}_{\tau_i} \Delta \boldsymbol{y}_{\tau_{i-|h|}}'$ for $h \geq 0$ and $\Gamma_h(\boldsymbol{y}) = \Gamma_{-h}(\boldsymbol{y})'$ for $h < 0$ be the realized autocovariance of $\boldsymbol{y}$ for given lag h. The class of flat-top realized kernels is designed to eliminate the noise-induced bias and variance of the RC estimator by weighting higher-order realized autocovariances appropriately as

14 It is generally not recommended to sample much more sparsely than daily since the asymptotic approximation of negligible drift, or local trends, may be poor at such frequencies. If the series display nonnegligible drift, this obviously needs to be taken into account when computing the QC estimates.

15 Besides synchronization errors, the MMS noise captures both exogenous effects, such as bid–ask bounce movements, and endogenous effects, such as asymmetric information and strategic learning among market participants.

$$RK^*(\boldsymbol{y}) = RC(\boldsymbol{y}) + \sum_{h=1}^{n-1} k(h/H)\{\Gamma_h(\boldsymbol{y}) + \Gamma_{-h}(\boldsymbol{y})\}, \tag{11}$$

for a bandwidth parameter $H = an^{1/2}$, $a > 0$, and, in particular, a nonstochastic kernel function, $k(\cdot)$, designed as

$$k(z) = 1_{\{|z| \leq k\}} + \lambda(|z| - k)1_{\{|z| > k\}}, \tag{12}$$

with $k = H^{-\xi}$, $\xi \in (0, 1)$, a shrinking function of the bandwidth H, and $\lambda(\cdot)$ a second-order smooth kernel function, satisfying some mild regularity conditions, an example being the Parzen kernel. The properties of these heteroskedasticity and autocorrelation consistent (HAC) style estimators depend crucially on the kernel function, and by selecting $k(z)$ as in Equation (12), the resulting class of flat-top realized kernels achieves optimal asymptotic properties in this setting, such as consistency, asymptotic unbiasedness, and mixed Gaussianity at the optimal rate of convergence, $n^{1/4}$, under mild assumptions on the MMS noise and (possibly random) sampling times.[16] If optimally designed, the estimator is, in addition, efficient in a Cramér–Rao sense. As a result, it performs well in finite samples, even for sparse observations available at one- to five-minute frequencies. Implementation details are provided in the Online Appendix.

When intradaily observations are only available for a certain part of a day, the trading window, and there is no recorded trading during weekends, holidays, etc., the estimates from the flat-top realized kernel may be supplemented with the squared close-to-open return since the preceding (trading) day. This approach essentially combines the estimates from $RC(\boldsymbol{x})$ and $RK^*(\boldsymbol{y})$.

2.2 A Simple Filtering Approach to Covariance Modeling

A number of different procedures to construct one-step conditional expectations of the QC matrix have been proposed in the literature. However, rather than searching for the best covariance model, the aim of this article is to provide a baseline approach for dynamic implementation of the realized currency betas, which is simple, of nonparametric flavor, and easy to implement for QC estimates with different degrees of measurement errors, such that it can accommodate within-period sampling at both daily and intradaily frequencies. In particular, the procedure that is introduced here adapts the rolling window estimator proposed by Foster and Nelson (1996) and Andreou and Ghysels (2002) in the univariate case, and extended to the multivariate case in Fleming et al. (2001), to the present setting. To this end, let $\boldsymbol{\Sigma}_t$ and $\hat{\boldsymbol{\Sigma}}_t$ be short-hand notation for the latent conditional QC matrix and a generic estimator of this, respectively. Then, the use of rolling window estimators implies the relation $\boldsymbol{\Sigma}_t = \sum_{i=1}^{\infty} \varpi_{t-i} \odot \hat{\boldsymbol{\Sigma}}_{t-i}$, where ϖ_{t-i} is a symmetric $(C+1) \times (C+1)$ matrix of weights, and $\odot$ is the Hadamard product. As proved by Foster and Nelson (1996, Theorem 5) under weak assumptions, the mean-squared error optimal covariance weights are given by $\varpi_{t-i} = \gamma \exp(-\gamma i)\iota\iota'$, such that $\boldsymbol{\Sigma}_t = \exp(-\gamma)\boldsymbol{\Sigma}_{t-1} + \gamma \exp(-\gamma)\hat{\boldsymbol{\Sigma}}_{t-1}$. The resulting covariance estimate is, however, generally downward biased since $(1+\gamma)\exp(-\gamma) < 1$ for $\gamma > 0$. Hence, a new bias-corrected version is introduced,

16 The presence of additive MMS noise slows down the best attainable rate of convergence from $n^{1/2}$ to $n^{1/4}$.

$$\boldsymbol{\Sigma}_t = (1 - (1+\gamma)\exp(-\gamma))\overline{\boldsymbol{\Sigma}}_{t-1} + \exp(-\gamma)\boldsymbol{\Sigma}_{t-1} + \gamma\exp(-\gamma)\hat{\boldsymbol{\Sigma}}_{t-1}, \tag{13}$$

where $\overline{\boldsymbol{\Sigma}}_{t-1}$ is the prevailing mean of $\hat{\boldsymbol{\Sigma}}_{t-1}$.[17] In other words, this version of the rolling window covariance estimator may be thought of as an exponentially weighted multivariate GARCH model for the time series of QC estimates, $\hat{\boldsymbol{\Sigma}}_t$, whose rate of decay is determined by a single parameter γ. Despite being parsimoniously parametrized, this approach allows for persistent time-variation in $\boldsymbol{\Sigma}_t$, while implicitly reducing the impact of measurement errors in $\hat{\boldsymbol{\Sigma}}_t$. Despite its simplistic structure, Varneskov and Voev (2013) show that the forecasting performance of Equation (13), without the bias-correction, is insignificantly different from that of more sophisticated multivariate Cholesky decomposed HAR and ARFIMA models when evaluated using a global minimum variance criterion for a portfolio of ten stocks and various RC measures.[18] As a robustness check, however, the empirical analysis of currency hedging strategies, presented in the next sections, has been carried out using a multivariate HAR model in place of the MGARCH model. The results are very similar, albeit with slightly worse risk-return properties, and are provided in Section F of the Online Appendix.

3 Data, Implementation, and Summary Statistics

This section introduces the data, which consist of daily and intradaily observations covering the time span from January 2000 through December 2019. Furthermore, it provides details on the construction of the QC estimates and forecasts, as well as the computation of the realized currency betas. Moreover, bid–ask spread data are used to estimate transaction costs for the evaluation of the dynamic hedging strategies. Finally, novel evidence of time variation in optimal currency exposures is presented.

3.1 Data Collection and Construction

The empirical analysis is performed for a U.S. investor who holds either an equity portfolio or a balanced portfolio with fixed income and equity, and may use (a subset of) the G10 currencies to hedge her foreign currency exchange rate exposure. In particular, two sets of currencies are considered. The first includes the very liquid Canadian Dollar (CAD), Swiss Franc (CHF), Euro (EUR), Great Britain Pound (GBP), and Japanese Yen (JPY). The second set further allows active investments in the Australian Dollar (AUD), the Norwegian Krona (NOK), the New Zealand Dollar (NZD), and the Swedish Krona (SEK). These currency investment universes are labeled G06 and G10, respectively. For each exchange rate, the last *daily* spot price (bid, ask, trade) quoted on Bloomberg during New York Stock Exchange (NYSE) trading hours is obtained, along with the corresponding one-month forward points, to construct forward prices, as well as estimates of the transaction costs associated with FX trading. Summary statistics for forward returns, spot returns, and implied returns on

17 Specifically, the present approach differs from the procedure in Fleming et al. (2003) by replacing the outer product of returns, or RC, with a generic QC estimator, similarly to the study in Varneskov and Voev (2013). Moreover, it differs from the latter by introducing a bias correction.

18 Using a statistical loss function, on the contrary, Varneskov and Voev (2013) find statistically significant gains from using multivariate Cholesky decomposed HAR and ARFIMA models over the multivariate GARCH model.

Table 1. Excess FX returns

Summary statistics for excess FX returns

	AUD	CAD	CHF	EUR	GBP	JPY	NOK	NZD	SEK
Mean (IR)	2.16	0.15	−1.61	−0.55	0.44	−2.05	0.91	2.51	−0.34
Mean (SR)	0.31	0.52	2.32	0.44	−1.01	−0.33	−0.52	1.20	−0.54
Mean (ER)	2.47	0.67	0.71	−0.12	−0.58	−2.37	0.39	3.72	−0.88
Std. Dev.	12.43	8.75	11.12	9.60	9.10	9.75	11.75	12.65	11.75
Skewness	−0.34	−0.10	3.73	0.05	−0.76	0.07	−0.15	−0.30	−0.04
Kurtosis	12.66	6.01	120.73	4.64	13.71	7.11	5.62	5.82	5.56
SR	0.20	0.08	0.06	−0.01	−0.06	−0.24	0.03	0.29	−0.07
Mean (BA)	6.92	4.45	6.47	3.08	4.22	3.78	15.13	12.76	11.58
Std. Dev.	6.60	3.90	7.84	3.20	5.67	5.00	17.33	11.61	10.83
Q05	1.52	1.05	1.21	0.82	0.68	1.04	3.67	3.99	4.08
Q95	17.95	9.81	19.31	9.45	12.24	9.66	41.47	30.62	29.79

This table presents annualized summary statistics for daily excess FX returns, in logarithms, from January 2000 through December 2019. Moreover, the mean excess return (ER) is decomposed into interest rate (IR) differentials and spot FX returns (SR). The bottom part of the table provides the mean and standard deviation as well as the 5% and 95% quantiles of estimated bid–ask (BA) spreads in BP. Specifically, the latter are computed as $10,000(\texttt{ask} - \texttt{bid})/((\texttt{ask} + \texttt{bid})/2)$ for the forward prices.

interest rate differentials are provided in Table 1, along with estimates of transaction costs in BP per annum. Note that *all* return series are log-transformed.[19]

Table 1 illustrates that currencies such as the CHF and JPY have lower returns on interest rates than the USD, and vice versa for the AUD and NZD. Moreover, the spot return fails to offset the interest rate differentials, consistent with the former two being funding currencies for carry trades, and the latter being high-yielding currencies in the G10 universe, see, for example, Lustig et al. (2011). For all individual currencies, the excess returns are modest compared to their volatility, generating annualized SRs of no higher than 0.3. However, an extensive literature (see the introduction for references) has documented that systematic FX trading can be very profitable, on a risk-adjusted basis. The CHF exhibits extreme kurtosis relative to the remaining currencies, caused by the Swiss National Bank's abandonment of the EUR peg on January 15, 2015. The estimated transaction costs, computed as $10,000(\texttt{ask} - \texttt{bid})/((\texttt{ask} + \texttt{bid})/2)$ for the forward prices, show that, on average, it is cheaper to trade G06 currencies than G10. Following convention in the literature, for example, Lyons (2001), the cost of trading is fixed at half the average spread, that is, at 3 BP for the G06 currencies and 6 BP for the G10 set. The results, however, are robust to increasing these numbers.

Three different baseline portfolios are considered. The first is the S&P 500, whose currency exposure is determined implicitly through the international investments and cash

19 Since all summary statistics are computed directly from log-returns, there is no need to consider the Jensen's inequality correction in Proposition 1. The latter is only important when transferring log-returns back into gross return form. Moreover, as seen by the general identity from Itô's lemma, $R_t = r_t + [r]_t/2$, the use of log-returns leads to a conservative assessment of the benefits from applying the proposed dynamic hedging strategies.

flows of its constituents. The second is an equally weighted portfolio in the DAX, FTSE 100, and S&P 500. All equity index investments are carried out via futures contracts. The third portfolio has a 60% weight on the S&P 500 and 40% on 10-year Treasury bond futures. This balanced, risk-parity-style, benchmark has been popularized by mutual funds, such as Vanguard, and is a prominent benchmark for passive investors. As such, the setup resembles the one in Campbell et al. (2010), who consider currency hedging for the S&P 500 and equal-weighted (EW) international equity investments, that is, the first two portfolios. The main differences are that they use either monthly or quarterly observations, while we collect *daily* data on the futures contracts. Moreover, they include the AUD in their set of (seven) hedging currencies, whereas we consider both the G06 and G10 sets of currencies. Finally, the investment horizons considered here are weekly and monthly, rather than monthly and quarterly.

In addition to the daily futures data for the baseline portfolios and FX forwards, we collect intradaily, one-minute data for the baseline portfolios and spot prices of the G06 currencies. The intradaily data are generally nonsynchronous, and *refresh time sampling* is used to align observations, see, for example, Varneskov (2016, Definition 1). Then, from either the daily or intradaily data, estimates of QC are constructed using RC or flat-top realized kernels, respectively.[20] Hence, our final dataset consists of either weekly or monthly observations of returns on the baseline portfolios and FX forwards, QC estimates for G06 and G10 constructed from daily data, as well as QC estimates for G06 constructed from intradaily data. This facilitates assessments of whether the new dynamic realized currency beta hedging theory generates improved portfolio performance, whether there are differences across baseline portfolios, or across different sets of hedging currencies, G06 versus G10, and whether there are additional gains from leveraging intradaily data to increase precision of the QC estimates.

3.2 Implementing Realized Currency Betas

The empirical analysis features several benchmark implementations of the dynamic realized currency beta hedging strategies. First, to extrapolate expectations of the QC matrix, the dynamic MGARCH specification in Equation (13) is estimated using standard multivariate Gaussian maximum likelihood techniques in conjunction with either the RC estimates from Equation (10) for the G06 currency set in place of $\hat{\mathbf{\Sigma}}_t$, the corresponding for the G10 set, or the flat-top realized kernel estimates from Equations (11)–(12) for the G06 set. Specifically, the smoothing parameter γ is estimated recursively using an expanding window of observations and a two-year initialization period. Given this, a QC forecast is generated from Equation (13) and used to compute the realized currency betas from Proposition 3. These three adaptive dynamic hedging (ADH) methods are labeled ADH-06, ADH-10, and ADH-HF to indicate the FX set for daily data-based QC estimates and high-frequency data-based estimates, respectively. In addition, a standard MGARCH model is included as a benchmark in the same G06 framework. This model is nested in the setting (13) by replacing $\hat{\mathbf{\Sigma}}_t$ with the outer product of returns (weekly or monthly, according to rebalancing frequency) and is labeled ADH-SM for Standard MGARCH.

20 Details on the cleaning of the intradaily data as well as its characteristics, for example, number of synchronized observations, noise-to-signal ratios, etc., are provided in Section E of the Online Appendix.

Table 2. Average exposures

Average exposures for currency hedging strategies

	CAD	CHF	EUR	GBP	JPY	AUD	NZD	SEK	NOK	USD
S&P 500										
CMV-06	−0.96	0.83	−0.69	0.09	−0.08	–	–	–	–	0.81
CMV-10	−0.45	0.70	0.34	0.17	−0.04	−0.40	−0.03	−1.03	0.22	0.53
ADH-SM	−0.99	0.53	−0.46	0.01	0.06	–	–	–	–	0.86
ADH-06	−0.69	0.61	−0.49	−0.06	0.34	–	–	–	–	0.29
ADH-10	−0.42	0.54	0.05	0.04	0.32	−0.27	−0.12	−0.32	−0.05	0.23
ADH-HF	−0.43	0.30	−0.26	−0.03	0.39	–	–	–	–	0.04
ROL-HF	−0.42	0.43	−0.35	−0.01	0.36	–	–	–	–	−0.01
EW										
CMV-06	−0.99	1.04	−0.91	0.30	0.07	–	–	–	–	0.48
CMV-10	−0.48	0.93	0.14	0.39	0.11	−0.39	−0.05	−1.00	0.17	0.20
ADH-SM	−1.39	0.76	−0.58	0.13	0.12	–	–	–	–	0.96
ADH-06	−0.63	0.59	−0.44	0.01	0.30	–	–	–	–	0.17
ADH-10	−0.36	0.52	0.13	0.10	0.29	−0.30	−0.04	−0.33	−0.11	0.10
ADH-HF	−0.48	0.37	−0.32	0.01	0.44	–	–	–	–	−0.02
ROL-HF	−0.47	0.46	−0.38	0.03	0.44	–	–	–	–	−0.08
Balanced										
CMV-06	−0.54	0.45	−0.44	0.09	−0.12	–	–	–	–	0.56
CMV-10	−0.26	0.38	0.13	0.13	−0.10	−0.21	−0.01	−0.60	0.14	0.40
ADH-SM	−0.57	0.27	−0.28	0.04	−0.03	–	–	–	–	0.56
ADH-06	−0.39	0.32	−0.27	−0.04	0.13	–	–	–	–	0.25
ADH-10	−0.23	0.28	0.03	0.02	0.12	−0.14	−0.09	−0.18	−0.03	0.21
ADH-HF	−0.24	0.15	−0.15	−0.02	0.17	–	–	–	–	0.10
ROL-HF	−0.23	0.22	−0.20	−0.01	0.16	–	–	–	–	0.07

This table presents the average exposures for seven different approaches to currency hedging and the three different baseline portfolios using a monthly rebalancing frequency. The baseline portfolios are the S&P 500, an EW basket of DAX, FTSE 100, and S&P 500 futures contracts, and a balanced portfolio with 60% S&P 500 and 40% 10-year U.S. Treasury bond futures. The different hedging strategies are detailed in Section 3.2. The sample spans January 2000 through December 2019. The dynamic covariance models are estimated using an expanding window with a two-year initialization period.

These dynamic strategies are compared to a fully hedged baseline portfolio to examine if active currency hedging adds economic value to the existing investments. Moreover, a *real-time* version of the strategy in Campbell et al. (2010), labeled CMV, is implemented, using the negative slope from a regression of the portfolio returns on excess FX returns. This is included for both the G06 and G10 sets of currencies, labeled CMV-06 and CMV-10, respectively, and utilizes an expanding window of either weekly or monthly observations, depending on rebalancing frequency. Finally, to examine the relative importance of the dynamic model for intraday-based QC estimates, a further benchmark is added, in which the QC expectations are based on the average flat-top realized kernel estimates using an expanding window up to and including ten years, and subsequently a rolling ten-year window. This simple procedure circumvents the forecasting step (13) and is dubbed the ROL-HF estimator.

Table 3. Risk-return performance

Risk-return performance for currency hedging strategies

	Full	CMV-06	CMV-10	ADH-SM	ADH-06	ADH-10	ADH-HF	ROL-HF	CON-HF
S&P 500									
Mean	5.93	5.31	4.39	2.71	6.34	4.14	7.87	7.24	8.00
Std. Dev.	14.29	12.79	12.47	15.49	12.22	11.58	11.76	12.85	11.79
Skewness	−0.92	−0.78	−0.37	−1.54	−0.74	−0.43	−0.81	−0.81	−0.82
Kurtosis	5.23	4.89	3.77	9.42	6.73	5.97	6.33	4.66	6.29
SR	0.41	0.41	0.35	0.17	0.52	0.36	0.67	0.56	0.68
EW									
Mean	4.40	3.41	2.10	1.29	4.02	1.81	6.73	6.01	6.90
Std. Dev.	14.96	14.02	13.80	16.59	12.74	12.23	13.27	13.99	13.33
Skewness	−0.93	−0.54	−0.43	−0.16	−0.45	−0.65	−0.58	−0.57	−0.55
Kurtosis	5.24	6.09	5.33	4.18	7.35	6.31	7.77	5.87	7.58
SR	0.29	0.24	0.15	0.08	0.32	0.15	0.51	0.43	0.52
Balanced									
Mean	5.03	4.89	4.38	2.92	5.05	3.71	5.96	5.73	5.96
Std. Dev.	8.15	7.37	7.20	8.59	7.13	6.65	6.88	7.49	6.88
Skewness	−1.00	−0.39	−0.19	−1.52	−0.40	−0.13	−0.73	−0.80	−0.73
Kurtosis	6.55	4.13	3.79	8.60	6.61	6.52	6.38	5.35	6.38
SR	0.62	0.66	0.61	0.34	0.71	0.56	0.87	0.76	0.87

This table presents annualized risk-return performance for nine different approaches to currency hedging and three different baseline portfolios using a monthly rebalancing frequency. The baseline portfolios are the S&P 500, an EW basket of DAX, FTSE 100, and S&P 500 futures contracts, and a balanced portfolio with 60% S&P 500 and 40% 10-year U.S. Treasury bond futures. The different hedging strategies are detailed in Section 3.2. The sample spans January 2000 through December 2019. The dynamic covariance models are estimated using an expanding window with a two-year initialization period.

As a preliminary gauge of the similarities and differences between strategies, the average currency exposures are provided in Table 2 for each benchmark portfolio. Some interesting discrepancies between average exposures obtained from the various hedging procedures appear. First, for the S&P 500 portfolio, the CMV-06 strategy takes large long positions in the CHF and USD, funded by short positions in CAD and EUR. The ADH-HF procedure holds JPY rather than USD, and the ADH-06 strategy has approximately equal (long) exposures to the JPY and USD. This is consistent with the CHF, JPY, and USD being considered as safe haven currencies during episodes of financial turmoil. Second, the absolute magnitude of the average FX exposures vary, with exposures in CMV-06 larger than those in ADH-06, which in turn are larger than the corresponding in ADH-HF. Third, when the FX set is extended from G06 to G10 for the CMV and ADH strategies, this generates a reduction in the short positions in CAD and EUR (whose exposure actually changes sign), in favor of shorting the AUD and SEK. Moreover, the CMV procedure indicates positive exposure to NOK. Fourth, whereas the average exposures are similar in magnitude and sign for the S&P 500 and EW portfolios, the optimal exposures for the balanced portfolio are significantly reduced. This is consistent with the latter having lower volatility (cf. Table 3), implying that the hedging strategies must match a lower volatility target.

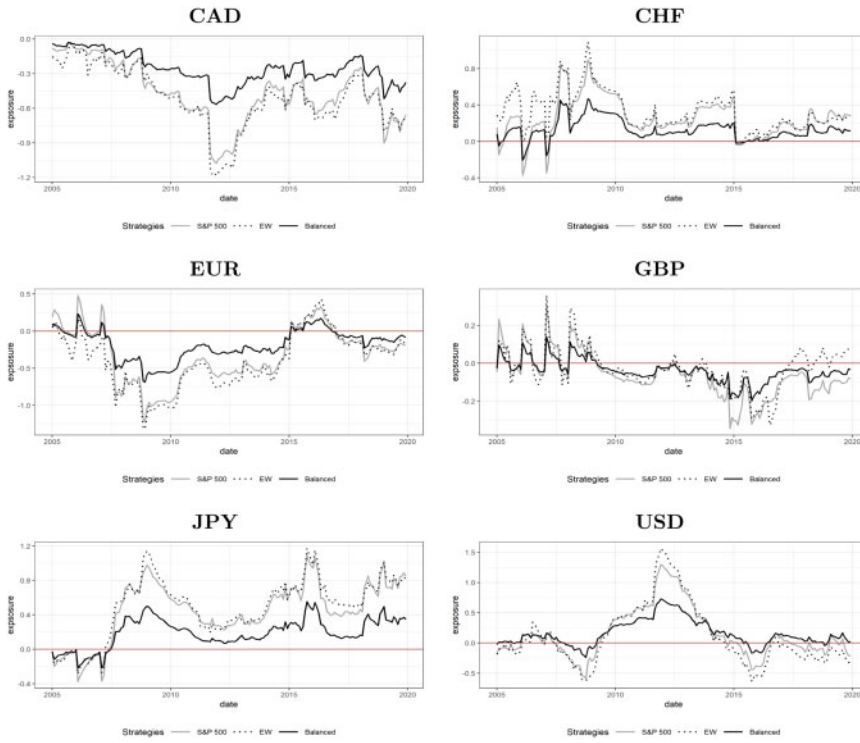

Figure 1. Exposure plots. This figure depicts the optimal currency exposures to the CAD, CHF, EUR, GBP, JPY, and USD, respectively, computed using the ADH-HF procedure for three different baseline portfolios and a monthly rebalancing frequency. The exposures are depicted for January 2005 through December 2019.

Figure 1 depicts the optimal exposures for the ADH-HF procedure applied to the three baseline portfolios. While this underscores the propensity of the hedging strategy to go long CHF and JPY, and funding this by shorting the EUR and CAD, there are interesting dynamic patterns. First, the exposures to the CHF and JPY are much larger during the global financial crisis of 2008–2009 than during the years leading up to and following it. Hence, the strategy takes on larger "safe haven bets" during financial turmoil. Second, the status of the CHF as a hedging currency is dramatically affected by the Swiss National Banks's abandonment of its EUR peg during January 2015, and while the optimal exposure remains positive afterward, it leaves the hedging demand much smaller. The optimal exposure to the JPY, on the contrary, remains large after the financial crisis, with further spikes occurring in early 2016, when fears of a slowdown of China's economy lead to global stock sell-offs, and during late 2018, when, again, stocks exhibit massive sell-offs over Christmas. Third, the optimal exposure to the USD is flat both before 2010 and after 2015. However, the model suggests to hold USD during the European sovereign debt crisis. Finally, Figure 1 also illustrates that it is optimal to hold more than a $\pm 100\%$ exposure to certain currencies over short portions of the sample period. This may not be feasible for all institutional investors. Hence, as a robustness check, we will also examine the performance of a

modified version of the ADH-HF strategy, in which, if the optimal exposure is larger than ± 1, it is fixed at ± 1. This constrained version is labeled the CON-HF strategy.

4 Benefits from Dynamic Global Currency Hedging

This section demonstrates that beside displaying interesting time-variation closely tied to important economic events (cf. Figure 1), the estimated dynamic realized currency betas give rise to hedging strategies that provide economic benefits to an investor beyond what is achieved by either fully hedged equity investments or existing optimal hedging strategies from Campbell et al. (2010), which ignore within-period variation in the economic system. Specifically, the gains from dynamic currency hedging are illustrated from three different perspectives. First, standard risk-return results are provided. Second, the statistical significance of the volatility reductions is formally tested. Third, the economic benefits to a risk-averse investor are assessed. The results are described for monthly returns. The corresponding results for weekly returns are very similar and available in Section F of the Online Appendix, together with robustness checks using the multivariate HAR model.

4.1 Risk-Return Benefits

As an initial gauge of the benefits to dynamic currency hedging, Table 3 reports the annualized mean return, standard deviation, and SR, along with skewness and kurtosis, for an equity investor who implements one of the eight hedging strategies, for each of the three baseline portfolios. There are several striking findings, which are described in detail for the S&P 500 portfolio and subsequently generalized to the remaining two portfolios. First, all strategies except ADH-SM provide nontrivial volatility reductions relative to full hedging. Second, the dynamic ADH-06 and ADH-HF strategies generate larger volatility reductions than the corresponding CMV-06 procedure, with the HF version performing best. Specifically, the ADH-HF strategy delivers a 100 BP improvement in volatility over CMV-06, in addition to average returns that are 150 BP higher. Consequently, the SR of ADH-HF is 61% higher than those for the fully and CMV-06 hedged portfolios, and 29% higher than for ADH-06. Third, while the CMV-06 procedure delivers substantial volatility reductions, these come at a cost of returns, thus failing to improve the SR relative to the fully hedged portfolio. This is consistent with the empirical findings in Campbell et al. (2010), despite the different setup. Fourth, when comparing equivalent strategies using either the G06 or G10 sets of currencies, the latter is seen to generate the largest volatility reductions, but these are very costly, and the strategies deliver worse overall risk-return performance. From Table 2, this follows since these strategies, on average, substitute a large part of the short position in the CAD to the high-yielding AUD, which is very expensive.

The results in Table 3 demonstrate that a better risk-return trade-off can be achieved by the proposed realized currency beta framework. Moreover, they show that obtaining precise high-frequency data-based estimates of QC delivers the best results, judging by the SR, and that daily data-based QC estimates also provide economic value over standard benchmarks. They further illustrate that the selection of currency universe is important for overall performance, and that active trading in high-yielding currencies can be expensive. Interestingly, the results also indicate that reasonable performance can be achieved by the simple ROL-HF procedure, which has higher SR than ADH-06,

Table 4. Volatility reductions

Volatility reductions for currency hedging strategies

	Full	CMV-06	ADH-SM	ADH-06	ADH-HF	ROL-HF	CON-HF
S&P 500							
FTRK	15.47	15.51	16.26	13.72	13.29*	14.34	13.30*
RC	15.37	13.85	14.61	12.35*	12.51*	13.69	12.52*
EW							
FTRK	16.02	15.71	17.65	13.69	13.49*	14.75	13.50*
RC	14.83	13.75	15.55	12.28*	12.43	13.52	12.45
Balanced							
FTRK	8.70	9.02	9.42	7.99	7.74*	8.23	7.74*
RC	8.75	8.28	8.60	7.41*	7.46*	8.03	7.46*

This table presents the average, annualized, ex post volatility for seven approaches to currency hedging and three baseline portfolios using a monthly rebalancing frequency. The ex post QC measures are constructed using either RC based on daily data or the flat-top realized kernel (FTRK) based on intradaily data. The baseline portfolios are the S&P 500, an EW basket of DAX, FTSE 100, and S&P 500 futures contracts, and a balanced portfolio with 60% S&P 500 and 40% 10-year U.S. Treasury bond futures. The hedging strategies are detailed in Section 3.2. The sample spans January 2000 through December 2019. The dynamic covariance models are estimated using an expanding window with a two-year initialization period.
*Signifies that the hedging strategy belongs to the 10% MCS of Hansen et al. (2011).

suggesting that the input covariance measure is relatively more important than the dynamic model. Finally, and not surprisingly, given Figure 1, the results for ADF-HF and CON-HF are almost identical. The few and small violations of ± 1 exposure have little impact on overall performance.

In sum, the risk-return benefits from applying the realized currency beta hedging strategies, especially those leveraging intradaily data, appear substantial. Importantly, they are also not confined to the S&P 500 portfolio. The remaining portions of Table 3 show that the same performance pattern appears for the EW and balanced portfolios. Specifically, the relative volatility and SR gains are similar in magnitude for ADH-HF, thus delivering superior risk-return performance. The next two subsections test the significance of the gains and assess them from the perspective of a risk-averse economic agent.

4.2 Significance Testing of Volatility Reductions

This subsection elaborates on the risk-return results by testing whether the volatility reductions achieved by the dynamic hedging strategies are statistically significant. Specifically, the best set of hedging models is determined by their ability to reduce volatility of the baseline portfolio. To this end, the model confidence set (MCS) approach of Hansen et al. (2011) is applied to identify the models in a manner that is robust against multiple testing biases. The testing procedure requires the selection of an ex post QC proxy and is implemented using both the RC and the flat-top realized kernel estimates. Since the hedging strategies using the G10 universe consistently exhibit lower SRs than those actively trading G06 currencies, and since flat-top realized kernel estimates are only available for the latter, this

exercise will focus on the G06 universe. Moreover, the MCS is configured with the T-max statistic and a 10% significance level (see Hansen et al., 2011 for details).

Results are collected in Table 4. The table shows the average ex post volatility for each of the QC measures, and indicates (by an asterisk) whether a given strategy belongs to the MCS. The message is clear. The ADH-06 and ADH-HF strategies consistently deliver the largest volatility reductions, significantly outperforming the fully hedged portfolio, as well as the CMV-06, ADH-SM, and ROL-HF strategies, for all three baseline portfolios. The ADH-HF strategy achieves the most significant volatility reductions when the flat-top realized kernel estimates are used as QC proxy, while the ADH-06 is also included in the set of best performing strategies when the RC estimate is used to calculate ex post volatility. In fact, the latter is the best strategy for the EW portfolio in this case, although its average volatility edge over ADH-HF is small. As for the risk-return results in Table 3, there is no significant difference between ADH-HF and CON-HF.

The volatility reductions from applying the dynamic ADH-HF strategy rather than fully hedging currency exposure or applying the existing CMV-06 approach to currency hedging are not only economically meaningful, for the S&P 500 portfolio of the order 250 BP and 100 BP for the two strategies (cf. Table 3), they are also statistically significant, based on the results in Table 4. Moreover, whereas ADH-HF achieves the highest SRs in Table 3 for all baseline portfolios, the results in Table 4 show that similar volatility reductions can be achieved by ADH-06. The advantage in terms of SR from using intradaily data in the implementation, thus, stems mainly from enhancing the returns to the hedging strategy.

4.3 Economic Benefits from Dynamic Currency Hedging

The economic benefits are assessed via three different measures. First, using the SRs in Table 3, the gain is quantified as the number of BP an investor is better off at a 10% volatility level, that is, by

$$\varpi_{b,s} = (\mathtt{SR}_{\mathtt{b}} - \mathtt{SR}_{\mathtt{s}}) \times 10 \times 100, \tag{14}$$

where $\mathtt{SR}_{\mathtt{b}}$ and $\mathtt{SR}_{\mathtt{s}}$ are the SRs of a benchmark strategy b and an alternative strategy s, respectively. However, as emphasized by Han (2006) and Della Corte et al. (2009), the SR may underestimate the performance of dynamic portfolio strategies. In particular, as the SR is computed using the full sample realized portfolio return and standard deviation, it may not adequately describe the *conditional risk* faced by an investor at each point in time. Hence, following Fleming et al. (2001), the economic value of a benchmark hedging strategy b relative to the alternative s is also assessed by determining the fee that may be subtracted from the hedged portfolio return corresponding to the benchmark each period, while still leaving average utility unchanged, compared to that achieved by investing according to the alternative hedging strategy. In other words, this fee equals the amount a risk-averse investor is willing to pay in order to switch from the alternative strategy s to the hedging benchmark b. Formally, as in Della Corte et al. (2009), let $Z_t^b = 1 + r_t^b$ and $Z_t^s = 1 + r_t^s$ be the payoffs to the benchmark and the alternative hedging strategy, respectively, then the switching fee $\Phi_{b,s}$ solves

$$\sum_{t=T_1+1}^{T}\left\{(Z_t^b-\Phi_{b,s})-\frac{\delta}{2(1+\delta)}(Z_t^b-\Phi_{b,s})^2\right\}=\sum_{t=T_1+1}^{T}\left\{Z_t^s-\frac{\delta}{2(1+\delta)}(Z_t^s)^2\right\}. \tag{15}$$

Specifically, Equation (15) equates the average realized period-by-period utility across the benchmark and alternative hedging strategies for an investor with quadratic preferences and relative risk-aversion indexed by the parameter δ. In the empirical application, $\delta \in \{3, 8\}$ is fixed (see Fleming et al., 2001; Della Corte et al., 2009, for detailed discussions of this preference specification).[21]

Finally, while both $\varpi_{b,s}$ and $\Phi_{b,s}$ speak to the (conditional) properties of means and volatilities for returns to the currency hedging strategies, it is worth examining the robustness to higher moments. For example, from Table 3, kurtosis and negative skewness are reduced by an expansion of the currency universe, for otherwise identical strategies. Thus, the hedging strategies are further evaluated using a measure capturing features of the whole return distribution, specifically, the Omega ratio statistic introduced by Bernardo and Ledoit (2000) and Keating and Shadwick (2002), and studied extensively by Caporin et al. (2018). The ratio is defined as

$$\Omega_s(\varrho)=\frac{\mathbb{E}[r_t^s-\varrho|r_t^s>\varrho]}{\mathbb{E}[\varrho-r_t^s|r_t^s\leq\varrho]}\times\frac{1-F_s(\varrho)}{F_s(\varrho)}, \tag{16}$$

with the conditional expectation taken with respect to a threshold ϱ, and $F_s(\varrho)$ the cumulative distribution function. Hence, the statistic quantifies the ratio of favorable and unfavorable outcomes with respect to a given threshold, where $\varrho \in \{0, 0.1\}$ is selected, following Caporin et al. (2018). As above, define $\Omega_{b,s}(\varrho)=\Omega_s(\varrho)/\Omega_b(\varrho)$ to assess the Omega ratio relative to a benchmark strategy.

Table 5 reports the estimates of $\varpi_{b,s}$, $\Phi_{b,s}$, and $\Omega_{b,s}(\varrho)$ using ADH-HF as the benchmark strategy for all three baseline portfolios. Interestingly, when considering the results for gains in SR, $\varpi_{b,s}$, the ADH-HF strategy delivers 210–250 BP improvements over full hedging, 200–350 BP over the existing CMV procedures, and 150–190 BP over ADH-06. These numbers are substantial, clearly illustrating the value of using intradaily data and the proposed economic model that actively utilizes within-period information to carry out dynamic currency hedging. The differences are smaller relative to the ROL-HF strategy, 70–100 BP, further underscoring the importance of QC measure quality relative to dynamic model specification. The ADH-HF and CON-HF strategies are economically identical.

When turning to the corresponding estimates of switching fees, $\Phi_{b,s}$, thus speaking to the relative conditional risk of the hedging strategies, the qualitative rankings are identical.[22] However, there are differences between the absolute magnitudes of the $\varpi_{b,s}$ and $\Phi_{b,s}$ estimates. The results for the S&P 500 and EW equity portfolios indicate even larger return differences than before. For the balanced portfolio, on the contrary, the estimated switching

21 Fleming, Kirby, and Ostdiek (2001, 2003) fix $\delta \in \{1, 10\}$ and Della Corte et al. (2009) let $\delta \in \{2, 6\}$. A higher value of δ, such as ten, implies that the investor is willing to pay a higher fee for strategies that reduce volatility.

22 Note that the switching fee estimates in Table 5 are new to the currency hedging literature and, thus, provide further perspectives on the economic value a risk-averse investor receives from implementing also the existing optimal hedging procedures in Glen and Jorion (1993), de Roon et al. (2003), and Campbell et al. (2010).

Table 5. Economic gains

Economic benefits of currency hedging strategies

	Full	CMV-06	CMV-10	ADH-SM	ADH-06	ADH-10	ADH-HF	ROL-HF	CON-HF
S&P 500									
$\varpi_{b,s}$	255.06	254.74	318.02	494.84	151.06	311.96	0.00	106.49	−8.72
$\Phi_{b,s}(3)$	294.68	294.97	374.87	669.15	170.37	366.85	0.00	104.57	−11.31
$\Phi_{b,s}(8)$	465.80	360.71	419.66	923.56	199.82	356.25	0.00	175.52	−9.22
$\Omega_{b,s}(0)$	0.82	0.82	0.78	0.69	0.90	0.79	1.00	0.92	1.01
$\Omega_{b,s}(0.1)$	0.84	0.82	0.78	0.70	0.90	0.79	1.00	0.93	1.01
EW									
$\varpi_{b,s}$	213.25	264.07	355.11	429.66	191.55	358.77	0.00	77.41	−10.20
$\Phi_{b,s}(3)$	305.27	363.10	484.79	692.49	250.14	451.80	0.00	101.56	−14.15
$\Phi_{b,s}(8)$	427.19	415.37	520.96	936.57	214.06	384.15	0.00	152.88	−10.30
$\Omega_{b,s}(0)$	0.83	0.80	0.74	0.70	0.85	0.74	1.00	0.93	1.01
$\Omega_{b,s}(0.1)$	0.84	0.80	0.74	0.71	0.85	0.74	1.00	0.93	1.01
Balanced									
$\varpi_{b,s}$	248.43	203.09	258.29	526.20	158.09	308.95	0.00	101.99	0.00
$\Phi_{b,s}(3)$	121.73	118.13	165.19	343.91	96.47	220.81	0.00	36.97	0.00
$\Phi_{b,s}(8)$	170.97	136.07	176.83	410.97	105.57	212.79	0.00	59.87	0.00
$\Omega_{b,s}(0)$	0.84	0.85	0.82	0.68	0.91	0.80	1.00	0.93	1.00
$\Omega_{b,s}(0.1)$	0.86	0.86	0.82	0.69	0.90	0.79	1.00	0.94	1.00

This table presents estimates of the economic gains arising from applying the ADH-HF strategy in place of eight alternative currency hedging strategies. The gains are computed for three baseline portfolios and a monthly rebalancing frequency. Specifically, as described in Section 4.3, $\varpi_{b,s}$ quantifies the SR difference, $\Phi_{b,s}(\delta)$ the switching fee for a risk-averse investor with quadratic utility function and risk-aversion parameter $\delta \in \{3, 8\}$, and $\Omega_{b,s}(\varrho)$ the relative Omega ratio for $\varrho \in \{0, 0.1\}$. The return differences $\varpi_{b,s}$ and $\Phi_{b,s}(\delta)$ are quoted in annualized BP, and an Omega ratio less than one indicates that ADH-HF achieves a higher value. The baseline portfolios are the S&P 500, an EW basket of DAX, FTSE 100, and S&P 500 futures contracts, and a balanced portfolio with 60% S&P 500 and 40% 10-year U.S. Treasury bond futures. The different hedging strategies are detailed in Section 3.2. The sample spans January 2000 through December 2019. The dynamic covariance models are estimated using an expanding window with a two-year initialization period.

fees are generally lower. However, since these fees are always greater than the 96 BP reported for the ADH-06 strategy, they remain economically large. The exception is the ROL-HF strategy, for which the relative gain is 35–60 BP, again demonstrating the value arising from using the precise flat-top realized kernel estimates of QC. The smaller switching fees for the balanced portfolio may readily be explained by the latter having substantially lower volatility than both equity portfolios. Hence, a risk-averse investor places a smaller premium on further volatility reductions, despite the $\varpi_{b,s}$ measure indicating that the risk-return gains are large.

Finally, the $\Omega_{b,s}(\varrho)$ estimates show that these conclusions remain robust when allowing higher-order return moments to impact the economic evaluation. The ADH-HF strategy, and its constrained version, consistently outperforms other approaches to currency hedging. In sum, the economic benefits achieved by the dynamic realized currency beta hedging

Table 6. FX trading strategies

Risk-return performance for FX trading strategies

	Summary statistics					Correlation			
	Mean	Std. Dev.	Skewness	Kurtosis	SR	RCB	MOM	CARRY	VALUE
RCB	1.95	9.15	1.35	7.97	0.21	1.00	0.21	−0.62	0.24
MOM	−1.24	7.39	0.52	6.23	−0.17	0.21	1.00	−0.10	−0.03
CARRY	3.38	8.01	−0.88	6.86	0.42	−0.62	−0.10	1.00	−0.60
VALUE	0.34	7.30	1.42	9.28	0.05	0.24	−0.03	−0.60	1.00

This table presents annualized risk-return performance for four different FX trading strategies using a monthly rebalancing frequency. These are carry, momentum, and value investments based on the G10 currency set, as well as the FX overlay of the ADH-HF RCB hedging strategy for the S&P 500 portfolio. Moreover, correlations between the strategies are reported. The sample spans January 2000 through December 2019. The dynamic covariance model is estimated using an expanding window with a two-year initialization period.

Figure 2. Exposure plots. This figure depicts cumulative (log-)returns to four different FX trading strategies using a monthly rebalancing frequency. These are carry, momentum, and value investments based on the G10 currency set, as well as the FX overlay of the ADH-HF RCB hedging strategy for the S&P 500 portfolio. The sample spans January 2000 through December 2019. The dynamic covariance model is estimated using an expanding window with a two-year initialization period.

strategies, especially when implemented using intradaily data, are economically large, and volatility reductions statistically significant.

5 Realized Currency Beta and Traditional Currency Investment Styles

The previous section demonstrates that the investor can achieve substantial gains in baseline portfolio performance by supplementing the latter with a tactical foreign exchange rate overlay based on the proposed realized currency beta (RCB) hedging procedure, especially, the dynamic ADH-HF approach. To synthesize and further explore these findings, this

section relates the performance of the zero net currency portfolio from the ADH-HF strategy to traditional currency investment styles, in particular, carry, momentum, and value strategies. Following, among others, Lustig et al. (2011), Menkhoff et al. (2012a, 2012b), and Asness et al. (2013), the carry trade is constructed by sorting on interest rate differentials, momentum on three-month excess currency returns, and the value trade by betting on mean-reversion against five-year average returns. Each of these strategies is implemented with standard rank-based weights and using the G10 set of currencies.

Table 6 reports summary statistics of the FX strategies, as well as correlations between them, with the currency overlay in the ADH-HF approach being designed to hedge the S&P 500 portfolio.[23] Moreover, their respective cumulative returns are depicted in Figure 2. In line with prior work, the carry and momentum investment strategies are very profitable until the 2008–2009 global financial crisis, when, specifically, carry trades exhibit a massive drawdown during the fall of 2008, culminating with a 12.4% loss in October. The strategy subsequently recovers, but has only delivered modest returns since 2010. This observed tail behavior of carry trades is consistent with prior findings in the literature, documenting that the strategy is exposed to "crash risk," for example, Brunnermeier et al. (2009), Burnside et al. (2011), and Menkhoff et al. (2012a). Momentum, on the contrary, acts as a hedge during the financial crisis, delivering a positive return of 11.5% in October 2008, but has performed abysmally since 2012. Finally, the value trading strategy similarly performed well during the financial crisis, but has been largely flat since 2010.

When traditional FX investment styles are compared to the ADH-HF overlay, the latter is also seen to provide protection during the financial crisis, returning 13.5% in October 2008. Interestingly, when comparing its performance to carry trades, keeping in mind the time-varying exposures of ADH-HF from Figure 1, the long positions in traditional carry funding currencies, such as CHF and JPY, suggest that the strong performance of ADH-HF is, at least partially, funded by carry traders unwinding their positions. In addition, the hedging strategy provides protection during other episodes of financial turmoil, such as those surrounding the downgrade of Greece's sovereign debt to junk bond status during the European sovereign debt crisis in April–May 2010, the Brexit vote in June 2016, and the "bloody Christmas" equity sell-off in 2018, delivering returns of 6.4, 5.0, and 4.5%, respectively, during these episodes. In contrast, momentum and value strategies return $-(1.4, 5.9, 2.7)\%$ and $(0.6, 1.5, 2.2)\%$ over the same time intervals. Hence, as intended, the ADH-HF overlay provides a robust hedge for an equity portfolio when protection is needed the most, and it performs better than existing FX strategies during such episodes. Moreover, the hedging strategy has lost less, on average, during bull market periods than it has gained during sell-offs, thus providing a positive return on average, albeit not as high as carry. Finally, the correlation between carry and the ADH-HF overlay is strongly negative, at –0.62 (cf. Table 6), suggesting that a profitable trading strategy could be constructed by combining the two. However, asset allocation among currency investment styles is beyond the scope of this article.

23 The results are very similar when applying the other two baseline portfolios and are, thus, omitted for brevity.

6 Conclusion

This article proposes a model for *discrete-time* currency hedging based on *continuous-time* movements in portfolio and foreign exchange rate returns. The vector of optimal currency exposures is shown to be given by the negative realized regression coefficients computed from a one-period conditional expectation of the intraperiod QC matrix for portfolio and foreign exchange rate returns, labeled the *realized currency betas*. The theoretical model, hence, facilitates the design of dynamic hedging strategies that depend exclusively on the evolution of the intraperiod QC matrix. This implies that interest differentials have *no* asymptotic impact on optimal currency hedging demands, and that an investor should sample observations as frequently as possible in fixed time intervals between portfolio rebalances to improve the accuracy of the QC estimates. Both implications contrast with prior theoretical results in the extant currency hedging literature, which assume that assets are observed at the same frequency as that at which the portfolio is being rebalanced. Moreover, since the proposed strategies only use information about the covariance between exchange rate and portfolio returns, not about local trends in the former, they are notably different from traditional currency investment styles, such as carry, momentum, and value.

The realized currency beta hedging strategies are implemented using modern, yet simple, nonparametric techniques to accurately measure the historical QC between assets and, subsequently, capture their dynamic evolution. Methodologically, this procedure addresses two general caveats in the literature. First, there has been a lack of dynamic modeling when computing optimal currency exposures, except when tied to slowly varying conditioning variables, such as past interest rate differentials. Second, previous work has been plagued by the use of forward-looking information when estimating optimal exposures, thus providing investors with the benefit of hindsight. Addressing both caveats is important for accurate assessments of intertemporal currency hedging demands and real-time investment decisions.

In an extensive empirical analysis, the use of the new hedging strategies, based on realized currency betas, produces novel results: (i) The optimal currency exposures display substantial time-variation, which is tied to important economic events, such as the 2008–2009 global financial crisis, the European sovereign debt crisis, and the "bloody Christmas" 2018 global stock sell-off. (ii) The proposed dynamic hedging strategies produce statistically significant, as well as economically substantial, volatility reductions for international equity portfolios and a balanced fixed income–equity portfolio, compared to either fully hedging currency exposure or applying existing approaches to (static) optimal hedging. (iii) These volatility reductions come without sacrificing returns, especially when implemented using intradaily data, delivering SRs 61% larger than key benchmarks. (iv) The estimated economic gains to the new hedging strategies are substantial, at 120–465 annual BP over full hedging—depending on baseline portfolio and investor risk-aversion—and 120–520 BP over existing static approaches. (v) The quality of the input QC measure seems to be more important for designing profitable realized currency beta hedging strategies than the dynamic model specification or an expansion of the hedging universe beyond the G06 currencies. (vi) The currency overlay behind the dynamic realized currency beta investment strategy is negatively correlated with the FX carry trade, and only modestly correlated with momentum and value investments. Interestingly, the empirical analysis suggests that carry traders, at least partially, fund the strong performance of the proposed dynamic strategy during the global financial crisis of 2008–2009.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

References

Akram, F., D. Rime, and L. Sarno. 2008. Arbitrage in the Foreign Exchange Market: Turning on the Microscope. *Journal of International Economics* 76: 237–253.

Andersen, T. G., and L. Benzoni. 2009. "Stochastic Volatility." In R. A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*. New York: Springer, pp. 8783-8815.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold. 2006. "Volatility and Correlation Forecasting." In G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*. Amsterdam: North Holland, pp. 777–878.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold. 2013. "Financial Risk Measurement for Financial Risk Management." In G. M. Constantinides, M. Harris, and R. M. Stulz (eds.), *Handbook of the Economics of Finance*. Amsterdam: North Holland, pp. 1127–1220.

Andersen, T. G., T. Bollerslev, and F. X. Diebold. 2010. "Parametric and Nonparametric Volatility Measurement." In Y. Ait-Sahalia and L. P. Hansen (eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland, pp. 67–138.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71: 579–626.

Anderson, R. W., and J.-P. Danthine. 1981. Cross Hedging. *Journal of Political Economy* 89: 1182–1196.

Andreou, E., and E. Ghysels. 2002. Rolling-Sample Volatility Estimators: Some New Theoretical, Simulation, and Empirical Results. *Journal of Business & Economic Statistics* 20: 363–377.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. Value and Momentum Everywhere. *The Journal of Finance* 68: 929–985.

Back, K. 1991. Asset Pricing for General Processes. *Journal of Mathematical Economics* 20: 371–395.

Baillie, R., and T. Bollerslev. 1989. Common Stochastic Trends in a System of Exchange Rates. *The Journal of Finance* 44: 167–181.

Barndorff-Nielsen, O. E., and N. Shephard. 2004. Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics. *Econometrica* 72: 885–925.

Bekaert, G., R. Hodrick, and X. Zhang. 2009. International Stock Return Comovements. *The Journal of Finance* 64: 2591–2626.

Bernardo, A. E., and O. Ledoit. 2000. Gain, Loss and Asset Pricing. *Journal of Political Economy* 108: 144–172.

Britten-Jones, M. 1999. The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *The Journal of Finance* 54: 655–671.

Brunnermeier, M. K., S. Nagel, and L. H. Pedersen. 2009. Carry Trades and Currency Crashes. *NBER Macroeconomics Annual* 23: 313–348.

Burnside, C., M. Eichenbaum, I. Kleshchelski, and S. Rebelo. 2011. Do Peso Problems Explain the Returns to the Carry Trade?. *Review of Financial Studies* 24: 853–891.

Campbell, J. Y., K. S. de Medeiros, and L. M. Viceira. 2010. Global Currency Hedging. *The Journal of Finance* 65: 87–121.

Caporin, M., M. Costola, G. Jannin, and B. Maillet. 2018. On the (ab)Use of Omega?. *Journal of Empirical Finance* 46: 11–33.

Chiriac, R., and V. Voev. 2011. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics* 28: 922–947.

Christoffersen, P. F., and F. X. Diebold. 2000. How Relevant is Volatility Forecasting for Financial Risk Management?. *Review of Economics and Statistics* 82: 12–22.

Christoffersen, P. F., V. R. Errunza, K. Jacobs, and H. Langlois. 2012. Is the Potential for International Diversification Disappearing? A Dynamic Copula Approach. *Review of Financial Studies* 25: 3711–3751.

de Roon, F. A., T. E. Nijman, and B. J. Werker. 2003. Currency Hedging for International Portfolios: The Usefulness of Mean-Variance Analysis. *Journal of Banking & Finance* 27: 327–349.

Della Corte, P., L. Sarno, and I. Tsiakas. 2009. An Economic Evaluation of Empirical Exchange Rate Models. *Review of Financial Studies* 22: 3491–3530.

Du, W., A. Tepper, and A. Verdelhan. 2018. Deviations from Covered Interest Parity. *The Journal of Finance* 73: 915–957.

Engle, R., and R. Colacito. 2006. Testing and Valuing Dynamic Correlations for Asset Allocation. *Journal of Business & Economic Statistics* 24: 238–253.

Errunza, V., K. Hogan, and M.-W. Hung. 1999. Can the Gains to International Diversification Be Achieved without Trading Abroad?. *The Journal of Finance* 54: 2075–2107.

Eun, C. S., W. Huang, and S. Lai. 2008. International Diversification with Large- and Small-Cap Stocks. *Journal of Financial and Quantitative Analysis* 43: 489–523.

Eun, C. S., S. Lai, F. A. de Roon, and Z. Zhang. 2010. International Diversification with Factor Funds. *Management Science* 56: 1500–1518.

Fama, E. F., and K. R. French. 2012. Size, Value, and Momentum in International Stock Returns. *Journal of Financial Economics* 105: 457–472.

Fleming, J., C. Kirby, and B. Ostdiek. 2001. The Economic Value of Volatility Timing. *The Journal of Finance* 56: 329–352.

Fleming, J., C. Kirby, and B. Ostdiek. 2003. The Economic Value of Volatility Timing Using "Realized" Volatility. *Journal of Financial Economics* 67: 473–509.

Foster, D. P., and D. B. Nelson. 1996. Continuous Record Asymptotics for Rolling Sample Variance Estimators. *Econometrica* 64: 139–174.

Glen, J., and P. Jorion. 1993. Currency Hedging for International Portfolios. *The Journal of Finance* 48: 1865–1886.

Grubel, H. G. 1968. Internationally Diversified Portfolios: Welfare Gains and Capital Flows. *The American Economic Review* 58: 1299–1314.

Han, Y. 2006. Asset Allocation with a High Dimensional Latent Factor Stochastic Volatility Model. *Review of Financial Studies* 19: 237–271.

Hansen, P. R., A. Lunde, and J. M. Nason. 2011. The Model Confidence Set. *Econometrica* 79: 453–497.

Jacod, J., and A. N. Shiryaev. 2003. *Limit Theorems for Stochastic Processes*, 2nd edition. New York: Springer.

Keating, C., and W. Shadwick. 2002. A Universal Performance Measure. *Journal of Performance Measures* 6: 59–84.

Kroencke, T. A., F. Schindler, and A. Schrimpf. 2014. International Diversification Benefits with Foreign Exchange Rate Styles. *Review of Finance* 18: 1847–1883.

Levy, H., and M. Sarnat. 1970. International Diversification of Investment Portfolios. *The American Economic Review* 60: 668–675.

Longin, F., and B. Solnik. 1995. Is the Correlation in International Equity Returns Constant: 1960-1990?. *Journal of International Money and Finance* 14: 3–26.

Lustig, H., N. Roussanov, and A. Verdelhan. 2011. Common Risk Factors in Currency Markets. *Review of Financial Studies* 24: 3731–3777.

Lustig, H., and A. Verdelhan. 2007. The Cross Section of Foreign Currency Risk Premia and Consumption Growth Risk. *American Economic Review* 97: 89–117.

Lyons, R. 2001. *The Microstructure Approach to Exchange Rates*. Cambridge, MA: MIT Press.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf. 2012a. Carry Trades and Global Foreign Exchange Volatility. *The Journal of Finance* 67: 681–718.

Menkhoff, L., L. Sarno, M. Schmeling, and A. Schrimpf. 2012b. Currency Momentum Strategies. *Journal of Financial Economics* 106: 660–684.

Moskowitz, T. J., Y. H. Ooi, and L. H. Pedersen. 2012. Time Series Momentum. *Journal of Financial Economics* 104: 228–250.

Opie, W., and J. Dark. 2015. Currency Overlay for Global Equity Portfolios: Cross-Hedging and Base Currency. *Journal of Futures Markets* 35: 186–200.

Schmittmann, J. M. 2010. "Currency Hedging for International Portfolios." *IMF Working papers*.

Varneskov, R. T. 2016. Flat-Top Realized Kernel Estimation of Quadratic Covariation with Non-Synchronous and Noisy Asset Prices. *Journal of Business & Economic Statistics* 34: 1–22.

Varneskov, R. T. 2017. Estimating the Quadratic Variation Spectrum of Noisy Asset Prices Using Generalized Flat-Top Realized Kernels. *Econometric Theory* 33: 1457–1501.

Varneskov, R. T., and V. Voev. 2013. The Role of Realized Ex-Post Covariance Measures and Dynamic Model Choice on the Quality of Covariance Forecasts. *Journal of Empirical Finance* 20: 83–95.

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 128–177
doi: 10.1093/jjfinec/nbaa036
Advance Access Publication Date: 6 January 2021
Article

A Descriptive Study of High-Frequency Trade and Quote Option Data*

Torben Andersen[1], Ilya Archakov[2], Leon Grund[2], Nikolaus Hautsch[2], Yifan Li[3], Sergey Nasekin[4], Ingmar Nolte[4], Manh Cuong Pham[4], Stephen Taylor[4], and Viktor Todorov[1]

[1]Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA, [2]Department of Statistics and Operations Research, University of Vienna, 1090 Vienna, Austria, [3]Alliance Manchester Business School, University of Manchester, Manchester M15 6PB, UK and [4]Department of Accounting and Finance, Lancaster University, Lancaster LA1 4YX, UK

Address correspondence to Ilya Archakov, Department of Statistics and Operations Research, University of Vienna, 1090 Vienna, Austria, and Manh Cuong Pham, Department of Accounting and Finance, Lancaster University, Lancaster LA1 4YX, UK or e-mail: ilya.archakov@univie.ac.at; m.c.pham@lancaster.ac.uk.

Received August 30, 2019; revised August 22, 2020; editorial decision August 26, 2020; accepted September 1, 2020

Abstract

This paper provides a guide to high-frequency option trade and quote data disseminated by the Options Price Reporting Authority (OPRA). We present a comprehensive overview of the U.S. option market, including details on market regulation and the trading processes for all 16 constituent option exchanges. We review the existing literature that utilizes high-frequency options data, summarizes the general structure of the OPRA dataset, and presents a thorough empirical description of the observed option trades and quotes for a selected sample of underlying assets that contains more than 25 billion records. We outline several types of irregular observations and provide recommendations for data filtering and cleaning. Finally, we illustrate the usefulness of the high-frequency option data with two empirical applications: option-implied variance estimation and risk-neutral density estimation. Both applications highlight the rich information content of the high-frequency OPRA data.

Key words: high-frequency data, market microstructure, options

JEL classification: C55, G10

* The authors thank the guest editor Kris Jacobs and two anonymous referees for their invaluable comments which greatly improved the quality of this paper. They would like to acknowledge the financial support from the ESRC-FWF bilateral grant titled "Bilateral Austria: Order Book Foundations of Price Risks and Liquidity: An Integrated Equity and Derivatives Markets Perspective", Grant Ref: ES/N014588/1 and the Austrian Science Fund (FWF): Research project: I-2762-G27.

The econometrics of option data has been a rapidly developing research area in recent years. Nonetheless, the full range of available data is underexploited, as empirical studies typically rely on closing (end-of-day) option prices. We advocate for a change in this regard. High-frequency option data have the potential to convey accurate real-time information regarding investors' expectations about a company, a sector, or even the entire market. Through the tight connection between option and underlying stock prices, intraday data provide a more comprehensive view of the realized and expected asset price dynamics, offering potential insights for short-term asset return predictability, intraday risk management, price discovery, information processing, and the role of liquidity. Finally, intraday option data allow for a more precise construction of popular model-free option-implied risk-neutral measures concerning the future return distribution and volatility. As such, they convey information regarding how strongly investors' expectations and risk appetites change in response to the intraday order flow and news arrivals.[1]

Among the vast literature on the U.S. options market, the use of high-frequency option prices is relatively rare.[2] There is a stream of literature employing high-frequency data to explore the intraday option price dynamics and provide inference on key quantities, including jumps, implied volatility surfaces, and risk-neutral densities, for example, Birru and Figlewski (2012), Andersen et al. (2015b), Audrino and Fengler (2015), Amaya, Bégin, and Gauthier (2018), Taylor, Tzeng, and Widdicks (2018), Kapetanios et al. (2019), and Dalderop (2020). However, these studies predominantly focus on options written on indices or index futures, while there is little work on the high-frequency dynamics of individual equity options.

Another severely under-researched area is the microstructure of equity option markets. The structure and organization of the U.S. market for exchange-traded options has undergone dramatic changes over recent decades. The number of exchange holding groups and trading venues has increased rapidly, the regulatory oversight has strengthened, infusing the markets with a higher degree of transparency and competition, the trading technology has developed at a rapid pace, experimentation with diverse incentives for market making and provision of order flow is rampant, and product innovation has been strong, with entirely new option categories gaining market share quickly. The implications of these developments for trading costs, price efficiency, liquidity, and market depth across exchanges and product categories are largely unexplored. A few recent studies are Muravyev and Pearson (2020) on trading costs, exploiting minute-by-minute option trades and quotes for stocks in the S&P 500 index, as well as Anand, Hua, and McCormick (2016) and Battalio,

1 In summary, intraday option prices are useful in improving our understanding of many of the issues explored by our friend, collaborator, and colleague Peter Christoffersen, including the valuation of options when volatility has multiple components (Christoffersen et al., 2008), the risk-neutral dynamics of volatility (Christoffersen, Jacobs, and Mimouni, 2010), the dynamic behavior of the implied volatility smirk (Christoffersen, Heston, and Jacobs, 2009), estimates of systematic equity risk provided by option prices (Chang et al., 2012), and the existence of liquidity premiums in option prices and quotes (Christoffersen et al., 2018).

2 See Section 2 for a comprehensive review.

Shkilko, and Van Ness (2016), who study the impact of the fee structure on the effective option bid–ask spread and total trading costs.

The paucity of research on the microstructure and high-frequency dynamics for the full cross-section of securities and option venues is due to a variety of factors. One primary reason is the large amount of data storage and processing resources required to analyze even a short sample comprising all options for a limited number of underlying assets. This is particularly striking for actively traded securities. For example, from January to August 2015, for the Apple stock, we observe an average of 271,641 trades and 3,283,067 quotes per day, implying a quote-to-trade ratio of 12.1. This activity is greatly surpassed by the option market featuring, on average, 77,849 trades and 226,822,053 quotes *per day* across all 1169 option classes written on Apple, suggesting an option quote-to-trade ratio of almost 3000. Second, the maturity structure evolves every day, with maturities dropping and new option contracts entering the sample, as others expire. Hence, the panel is unbalanced and it contains many thinly traded options. Third, the market environment is constantly shifting, with new venues appearing, exchange mergers eliminating existing ones, and trading protocols undergoing frequent modification. It can be difficult to generate a sufficiently large sample under stable conditions to obtain good empirical estimates of relevant structural quantities in this setting, but high-frequency data should help. Fourth, no single source exists for option trades and quotes, reported in a homogenous manner, over the full period of exchange-based options trading in the U.S. Fifth, obtaining a complete record of the market activity at the highest frequency has, at least until very recently, come at a high financial cost. As a result, scholars have exploited a plethora of *ad hoc* arrangements to obtain partial coverage of the market activity over limited sample periods.

Our main objective is to provide an overview of the current U.S. exchange-based option market, with an emphasis on the pitfalls and opportunities associated with high-frequency data. We rely on the most comprehensive database—the CBOE OPRA Data (Bulk) package—provided by the Options Price Reporting Authority (OPRA), which collects and disseminates intraday trade and quote data at a millisecond precision from all option exchanges operating in the United States. It covers all option classes written on more than 3500 equities, more than 500 exchange-traded products (ETPs), and about 50 index-driven assets. Hence, we seek to provide scholars with guidance on how to process and utilize such data, provide an explorative overview, and present some initial illustrations regarding the data's potential and benefits.

Therefore, we contribute to the literature by providing (i) an overview of the institutional and regulatory settings of the fragmented option market in the United States, based on all 16 U.S. security exchanges eligible for options listing and trading, (ii) a selective, yet fairly extensive, descriptive analysis of the intraday trades and quotes for options written not only on equity indices, but also individual stocks and other ETPs, and (iii) a few illustrative applications, implemented to demonstrate the feasibility of constructing high-frequency option-based measures through standard procedures, and then assess their advantages and drawbacks relative to end-of-day option data.

As a consequence, we do not pursue any specific research question in depth, but review relevant market institutions and features. Along the way, we identify opportunities for new research, that are opening up with the increasing availability of databases covering the option market activity at the tick-by-tick level, coupled with the rapid advances in processing power and the declining data storage costs.

We first briefly review the history and competitive forces that have shaped the market evolution since the start of organized option trading in the early 1970s. We then outline the most critical regulatory initiatives determining the transparency of trading, price formation, and market maker quoting obligations. The latter are especially important given the dominance of quotes relative to trades in option markets.

We next explore the quality of the raw (bulk) data. We identify the extent of potential faulty recordings, other data errors, non-informative or irregular quotations, outliers, and the occurrence of records with identical time stamps. To provide practical guidance, we develop a filtering algorithm in the spirit of the cleaning algorithm proposed by Barndorff-Nielsen et al. (2009) for high-frequency equity data.

To keep the analysis manageable, we focus on data originating from 12 representative underlying equities and 5 exchange traded funds, with parts of the study concentrating on the trading during January 2015. The latter still leaves us with approximately 27 billion trades and quotes. We summarize the trading and quoting activities for each underlying asset, with further categorization based on option maturity and moneyness. We also explore the liquidity characteristics through bid–ask spread measures. The latter are critical for assessing the signal-to-noise ratio associated with the option quotes. Finally, we check for arbitrage violations associated with a basic put-call parity relation. These occur at non-trivial frequencies and are observed from quotes at all exchange venues, albeit at somewhat different intensities.

A unique feature of the new OPRA database is that it allows for a study of the quoting and trading of an instrument across all exchange venues. This enables us to compare the cross-exchange trade and quote flows and to assess potential price leadership, determined by the degree to which the exchange participates in the National Best Bid and Offer (NBBO) quoting pair. We find that CBOE, AMEX, and ARCA most often match the NBBO, but participation is otherwise quite evenly distributed across all active exchanges, lending credence to the hypothesis of an integrated national options market.

Finally, we provide a pair of illustrations using OPRA data to gauge the feasibility and reliability of standard techniques for constructing popular option-implied measures at very high frequencies. The first application involves estimation of the risk-neutral return variance. We construct these measures second-by-second from option prices written on SPY (an exchange traded fund) and GOOG (Google stock). We verify that they display substantial, genuine intraday variation, so they add significantly to the information content provided by typical end-of-day measures extracted from, say, OptionMetrics data. However, we also find that the series suffer from significant serial correlation, indicating a non-trivial impact of noise. We find that such effects only vanish for sampling frequencies of about 1 min or lower.

In the second application, we estimate the risk-neutral return density (RND) from intraday option prices. We compare the estimated RND curves for all underlying 30 min before and after the news release from the FOMC meeting in March 2015. We document a uniform reduction in implied volatility along with a dampening of the left tail of the distribution after the FOMC statement. The point is that such analyses are perfectly feasible across a wide cross-section of equity options, allowing for future studies investigating heterogeneity in the response across stocks in different sectors or with distinct characteristics.

The rest of the paper is organized as follows. Section 1 provides an overview of the U.S. option exchange trading industry. Section 2 reviews the relevant literature that uses high-frequency options data. Section 3 contains an overview of the OPRA dataset, the characterization of its special records, and develops suggestive data filtering rules. A thorough description of the observed trade and quote record for the selected data sample are presented in Section 4. Section 5 provides our empirical illustrations using intraday option prices to estimate the risk-neutral variance and density of the underlying asset returns. Section 6 concludes the paper. All additional materials are relegated to the Online Appendix.

1 OPTION TRADING IN THE UNITED STATES

1.1 The Evolution of the U.S. Options Market

Exchange-based options trading in the U.S. began in April 1973 with the foundation of the Chicago Board Options Exchange (CBOE). During 1973–1999, three additional exchanges opened and continued operating as independent trading venues for options till 1999, namely the American Stock Exchange (AMEX), the Philadelphia Stock Exchange (PHLX), and the Pacific Exchange (PCX). These four exchanges were all floor-based with either an open outcry or a specialist structure. Over this period, the market was highly fragmented, as actively traded options, almost exclusively, were listed on a single exchange, so the trading activities were governed by the listing options exchange only, see, for example, Battalio, Hatch, and Jennings (2004).

This fragmentation was targeted by an options listing campaign in August 1999, when prior exchange-exclusive options began to be listed at competing exchanges. This campaign sharpened the competition among exchanges so that, in short order, 37% of all equity option volume were for contracts traded on multiple exchanges (De Fontnouvelle, Fishe, and Harris, 2003). Mayhew (2002) and De Fontnouvelle, Fishe, and Harris (2003) find the enhanced competition to improve market quality in terms of smaller quoted or effective spreads.

Inspired by the heightened exchange competition, in 2000, the Security and Exchange Commission (SEC) approved the Plan for the Purpose of Creating and Operating an Intermarket Options Linkage[3] (the "Linkage Plan") and Firm Quote and Trade-Through Disclosure Rules for Options[4] to facilitate the creation of a national market. The Linkage Plan is essentially a central routing system operated by the Options Clearing Corporation (OCC) for the participating exchanges to route order flows with the aim to limit trade-throughs and execute at the National Best Bid and Offer (NBBO) price. Comparing market quality before and after the Linkage Plan, Battalio, Hatch, and Jennings (2004) find the quality of the options market to be substantially improved, with the number of crossed quotes reduced by 85%, a 7% decrease in the trade-through rates, and an overall reduction in the effective spread of over 60%, suggesting the Linkage Plan was a major driver in transforming the fragmented options market into a national market system.

The year 2000 also witnessed the foundation of the International Securities Exchange (ISE)—a fully electronic options venue without a physical trading floor. This innovative design presented a challenge to the traditional floor-based options exchanges. According to

3 See, for example, https://www.sec.gov/rules/sro/34-43086.htm (accessed 23 October 2020).

4 https://www.sec.gov/rules/final/34-43591.htm (accessed 23 October 2020).

Simaan and Wu (2007), ISE generated more informative and executable option quotes with a smaller bid–ask spread compared with its rivals. These advantages catapulted ISE into the leading exchange in terms of trading volume by 2003, and the electronic-based market structure was adopted by CBOE in 2001, PCX in 2003, and the new Boston Options Exchange (BOX) in 2004.

The development of electronic option trading platforms and the associated boom in market activity in the mid-2000s created a problem for the Linkage Plan, as the centralized routing system was not designed to handle the elevation in order flow. This problem was compounded by the introduction of the Penny Quoting Pilot Program[5] (the "Penny Plot") in 2007, which increased the possibility of trade-throughs and locked/crossed markets due to a smaller tick size. Based on the Regulation National Market System (NMS) for the equity market in 2005, in 2009, the SEC approved the Options Order Protection and Locked/Crossed Market Plan[6] (the "Decentralized Linkage Plan"). This plan augments the former central routing channel with the Intermarket Sweep Order, which allows market makers to route their orders in an efficient and decentralized manner. The intention was to eliminate locked/crossed markets and allow for more efficient price protection following the Regulation NMS.

In 2007, the newly formed NYSE Arca exchange introduced a make-take fee structure to the option classes affected by the Penny Pilot. This led to direct competition with the traditional payment for the order flow (PFOF) model used by other exchanges. Anand, Hua, and McCormick (2016) find that NYSE Arca's shift to a make-take fee structure reduced the execution costs for liquidity demanders and improved the quotes posted by market makers. This is partially confirmed by Battalio, Shkilko, and Van Ness (2016), although they note that the PFOF may lead to lower effective transaction costs for low-priced options. Subsequently, the make-take model was adopted by various exchanges, including Nasdaq, BOX, BATS, ISE, and PHLX in 2010.

The number of option venues has doubled over the past decade in parallel with multiple exchange mergers. As of January 2020, 16 option exchanges operate in the United States, with five holding companies owning one or more of these. They include Nasdaq (PHLX, NOM, BX Options, ISE, GEMX, MRX), CBOE Holdings (CBOE, C2, BATS BZX, BATS EDGX), Intercontinental Exchange (NYSE AMEX, NYSE Arca), Miami International Holdings (MIAX, MIAX Pearl, MIAX Emerald), and TMX Group (BOX). Table 1 summarizes major events that influenced the evolution of option trading venues in the United States. A detailed description of all exchanges is provided in Table A.1 of Online Appendix A.

Figure 1 depicts the total annual number of option contracts traded on various exchanges along with the corresponding market shares for 1973–2016. In the two decades leading up to 2000, the trading activity was fairly stable, with CBOE being the dominant venue. The foundation of ISE in 2000 and the introduction of electronic trading triggered unprecedented growth, tripling the overall volume by 2010 and leading to a sharply increased market share for ISE. Finally, after 2010, the trading activity has stabilized, while also becoming more evenly distributed across exchanges.

5 https://www.sec.gov/news/press/2007/2007-10.htm (accessed 23 October 2020).

6 https://www.sec.gov/rules/sro/nms/2009/34-60405.pdf (accessed 23 October 2020).

Table 1. The evolution of options exchanges in the United States

Date	Event	Number of venues
April 1973	The **Chicago Board Options Exchange (CBOE)** was launched.	1
January 1975	Options trading initiated at the American Stock Exchange (AMEX)	2
June 1975	Options trading initiated at the PHLX	3
April 1976	Options trading initiated at the PCX	4
December 1976	Options trading initiated at the Midwest Stock Exchange (MSE)	5
June 1980	Options business at MSE was consolidated with CBOE	4
June 1985	Options trading initiated at the New York Stock Exchange (NYSE)	5
April 1997	Options business at the NYSE was consolidated with CBOE	4
May 2000	The ISE was launched	5
February 2004	The **BOX** was launched.	6
September 2005	The Archipelago Exchange (ArcaEx) acquired PCX.	6
February 2006	The NYSE Group acquired ArcaEx to form **NYSE Arca Options**	6
July 2007	Nasdaq acquired PHLX to form **Nasdaq PHLX**	6
March 2008	**Nasdaq Options Market (NOM)** was launched	7
October 2008	The NYSE Group acquired AMEX to form NYSE AMEX Options	7
February 2010	BATS Options was launched	8
October 2010	**CBOE C2** was launched	9
June 2012	Options trading initiated at **Nasdaq BX Options (NOBO)**	10
July 2013	ISE Gemini was launched	11
December 2013	The **MIAX Options Exchange** was launched	12
November 2015	BATS EDGX Options was launched	13
February 2016	ISE Mercury was launched	14
June 2016	Nasdaq acquired ISE to form **Nasdaq ISE**, **Nasdaq GEMX** (former ISE Gemini), and **Nasdaq MRX** (former ISE Mercury)	14
February 2017	CBOE acquired BATS to form **CBOE BZX** (former BATS Options) and **CBOE EDGX**	14
February 2017	**MIAX Pearl** was launched	15
August 2017	NYSE AMEX Options was renamed to **NYSE American Options**	15
March 2019	**MIAX Emerald** was launched	16

The active options exchanges as of February 2020 are in bold. The information in the table is partially based on Mayhew (2002).

1.2 Market Regulation

The option exchanges are authorized by SEC to act jointly, as parties to a number of NMS plans, including the aforementioned decentralized linkage plan. These plans centralize requirements across exchanges to ensure cross-exchange protection, transparency, surveillance, standardization, and audit trails.

The "Plan for Reporting of Consolidated Options Last Sale Reports and Quotation Information"[7] is one of the NMS plans aimed at reporting trading information from all U.S. option exchanges. The OPRA is registered as a securities information processor,

7 In accordance with Section 11A of the Securities Exchange Act of 1934.

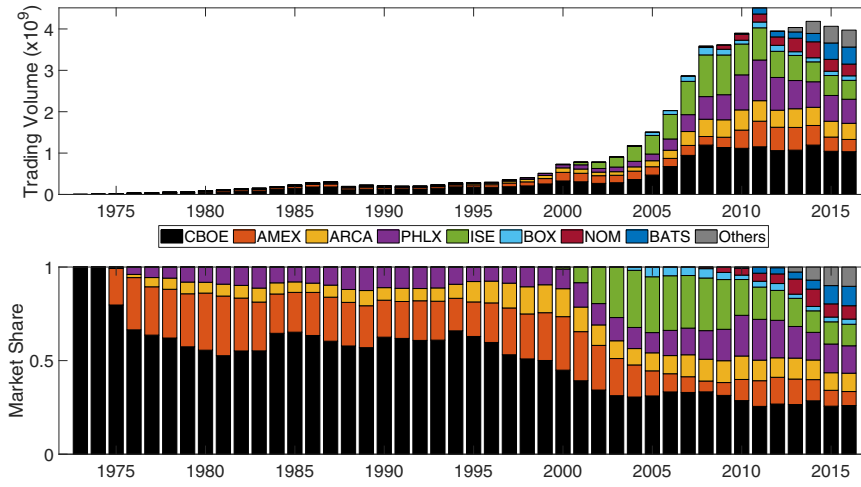

Figure 1. Total annual option contract volume traded on different exchanges. The "Others" category contains C2, EDGX, MIAX, MRX, NOBO, and GEMX. We exclude NYSE, MSE, and NASDAQ, as their respective option trading activities amassed only minor market shares during brief periods prior to 2000. The data stem from the CBOE Annual Market Statistics.

responsible for the implementation of this plan, and is regulated by a committee comprising all participating exchanges.

OPRA currently processes option trading data from all 16 U.S. option exchange markets. OPRA requires participants to report information on last sale and current quotes in accordance with Rule 602 of Regulation NMS (including prices, quotation sizes, and some regulatory auditing information). The Securities Industry Automation Corporation (SIAC) provides technological infrastructure for collection, consolidation, and dissemination of this real-time information. OPRA provides market data to professional (directly or through vendors) and non-professional subscribers (only through vendors) for a fee.

There are three additional mandatory plans for the exchanges. The Options Regulatory Surveillance Authority (ORSA) Plan was adopted in 2006. It seeks to deter insider trading. The Options Listing Procedures (OLP) Plan was introduced in 2006 "to facilitate the listing and trading of standardized option contracts on each of the exchanges." The Consolidated Audit Trail (CAT) was filed in 2014 to collect all orders and identify them as cancellations, modifications, or executions for the exchange-listed equities and options across all U.S. markets.

1.3 Options Trading and Market Maker Obligations

The core trading session in the U.S. option markets begins at 8:30 and lasts until 15:00 Central Time (CT) every business day. Index- and ETP options have an extended session ending at 15:15 CT. In addition, some exchanges (BATS BZX and BATS EDGX) provide a premarket trading session, initiated up to 2 h prior to the regular market open.

The OCC summarizes the most important product-specific information, sets daily position limits (250,000 contracts for most liquid stocks), and requires minimum customer margins—up to 120% of the aggregate contract volume for writers of uncovered options.

The most heavily traded options are written on single stocks, cash indices, and ETF/ETPs, but many other types of underlying are available, for example, commodities, interest rates, and foreign exchange. A standard equity or ETF/ETP option contract is an American-type option covering 100 shares of the underlying, and exercise of the contract results in physical delivery of the underlying shares. In case of stock splits or dividend payments, the number of shares and exercise prices are adjusted accordingly. Index options are, in contrast, mostly European-style and settled in cash. Regular options usually expire at the close of trading on the third Friday of each month. In 2005, CBOE issued weekly options that expire every Friday[8] for some indices and ETFs, and expanded it to cover individual securities in 2012. Monday and Wednesday-expiring weekly options were launched in 2016 for selected indices and ETFs.[9]

The option market is a *hybrid quote-driven* market, where market makers are responsible for providing continuous bid and offer quotes. Often, there are multiple types of market makers that differ in privileges and responsibilities. More senior categories (e.g., lead or primary market makers) are granted allocation priority in the relevant option classes, but are subject to stricter capital requirements, quoting obligations, and other responsibilities. Regular market makers may also be registered as preferred or directed market makers with certain privileges in executing preferenced and directed orders. Since the type and role of market makers differ across exchanges and over time, we only discuss some of their general obligations. We refer interested readers to Mayhew (2002) and Simaan and Wu (2007) for a more detailed account.

There are several market-wide obligations for option market makers. First, in February 2001, the SEC introduced a market-wide firm quote obligation through an amendment of the Quote Rule (Securities Exchange Act Rule 11Ac1-1), which was previously applied only to the equity market. This rule requires the market makers to post firm quotes that are valid for order executions of at least one contract. Before that, option exchanges imposed their own firm quote requirements on market makers independently of each other. Second, in 2010, the SEC proposed an amendment to the local exchange rules that prohibit market maker stub quotes, that is, quotes that are far away from the prevailing market. Stub quotes might be posted, when market makers attempt to fulfill quoting obligations without an actual intent to trade. Stub quotes were viewed as a contributing factor to the Flash Crash on May 6, 2010. The new rule requires quotes to be within a certain percentage band around the NBBO (or the consolidated last sale, if the NBBO is not available). These requirements seek to make options trading less risky for investors and prevent transactions from being executed at irrational prices.

By the Quote Rule, market makers must provide continuously updated two-sided quotes throughout the trading day. Each option exchange (or a self-regulatory organization) imposes additional obligations on its market makers. In general, these quoting obligations are in force irrespective of the prevailing market conditions. Therefore, during episodes of stress, market makers are supposed to maintain liquidity, absorbing the impact of shocks

8 Weekly options are not issued when an existing option of the same specification also expires on the same Friday.

9 Underlyings with weekly options available can be found in http://www.cboe.com/products/weeklys-options/available-weeklys.

on individual investors.[10] The requirement of continuous quoting is especially important for option markets, because an appreciable fraction of the securities is thinly traded.

Most exchanges require market makers to quote at least 90% of the time during the trading day, with compliance assessed on a monthly basis. Moreover, the quote size should exceed a minimum number of contracts, usually determined on a class-by-class basis, and may vary with market maker type. Moreover, market makers must quote continuously in some minimum fraction of the option classes and series to which they are assigned. These fractions range from 60% to 100% across exchanges and market maker type. More detailed information on quotation requirements is collected in Table A.2 of Online Appendix A.

The minimum tick size—the smallest possible price increment—depends on the price level of an option. For those traded below $3, the minimum tick constitutes $0.01 for the option classes participating in the option penny pilot program and $0.05 for other classes. For options traded above $3, the minimum tick is $0.05 for the classes from the penny pilot program and $0.10 for the other series. The options written on many market-wide ETFs (namely, QQQQ, IWM, and SPY) and option-related products (XSP and VIXW) represent exceptions with a minimal increment of $0.01 for all corresponding option series.

2 LITERATURE REVIEW ON HIGH-FREQUENCY OPTIONS DATA

This section provides a brief and selective review of the existing literature using U.S. high-frequency options data and discusses the differences across the data sources with emphasis on the distinction between the OPRA bulk dataset and those used in prior studies. Toward that purpose, Table 2 references a number of papers along with their associated datasets, sampling periods, and option classes used.

The first widely adopted intraday U.S. stock options data stem from the Berkeley Options Data Base (BODB), which collected each transaction and bid/ask update for every option series on the CBOE from the Market Data Report, time-stamped to the second. The BODB only covers CBOE trades and quotes until 1997, as the systematic collection of BODB data seems to end by December 1996.[11] However, some authors may still, subsequently, have acquired this type of data directly from the CBOE.[12]

After the termination of BODB, scholars focused on an older version of the OPRA dataset for intraday stock options data, stemming from the early 2000s, labeled OPRA (old) in Table 2. This edition of OPRA provides a complete record of option trades and best available quotes for all exchanges, time-stamped to-the-second, but depth information is unavailable. This dataset appears to have been discontinued later in the 2000s. We have been unable to establish the exact sample period covered.

The old OPRA is considerably less granular than the up-to-date, to-the-millisecond OPRA dataset exploited in this article, and labeled OPRA (new) in Table 2. New OPRA

10 See, for example, https://www.sec.gov/comments/s7-05-15/s70515-34.pdf (accessed 23 October 2020) and Nagel (2012).

11 See, for example, https://libguides.stanford.edu/az.php?a=b and https://catalog.princeton.edu/catalog/2593696 (accessed 23 October 2020).

12 Bollen and Whaley (2004) use such data covering 01/1995–12/2000 for the 20 most heavily traded stock options.

Table 2. List of papers using high-frequency options data from the U.S. market

Authors	Dataset	Sampling Period	Sampled Options Classes
Bhattacharya (1987)	BODB	06/1977–08/1978	Options on the 32 top stocks by volume
Stephan and Whaley (1990)	BODB	01/1986–03/1986	All call options
Vijh (1990)	BODB	03/1985–04/1985	Options on NYSE-listed stocks
Chan, Chung, and Johnson (1993)	BODB	01/1986–03/1986	All options
Sheikh and Ronn (1994)	BODB	01/1986–09/1987	30 most heavily traded stock options
Chan, Chung, and Johnson (1995)	BODB	01/1986–03/1986	Options on NYSE-listed stocks
Mayhew, Sarin, and Shastri (1995)	BODB	09/1985–06/1986, 01/1988–10/1988	Short-dated options on approximately 110 stocks
Bakshi, Cao, and Chen (1997)	BODB	06/1988–05/1991	Options on the S&P 500 index
Easley, O'Hara, and Srinivas (1998)	BODB	10/1990–11/1990	Options on the 50 top stocks by volume
Lee and Yi (2001)	BODB	01/1980–12/1990	Call options on NYSE-listed stocks
Chan, Chung, and Fong (2002)	BODB	01/1995–03/1995	Options on the 60 top stocks by volume
Mayhew (2002)	BODB	01/1986–08/1997	All stock options traded on CBOE
Pan (2002)	BODB	01/1989–12/1996	Options on the S&P 500 index
Chakravarty, Gulen, and Mayhew (2004)	BODB	01/1988–12/1992	Sixty most heavily traded stock options
Cao, Chen, and Griffin (2005)	BODB	01/1986–12/1994	Options on firms involved in merger and acquisition activities
George and Longstaff (1993)	CBOE	01/1989–12/1989	Options on the S&P 100 Index
Bollen and Whaley (2004)	CBOE	06/1988–12/2000	Options on the S&P 500 index and 20 individual stocks
De Fontnouvelle, Fishe, and Harris (2003)	OPRA (old)	08/1999, 08/2000	28 multilisted option classes
Battalio, Hatch, and Jennings (2004)	OPRA (old)	06/2000, 01/2002	71 (615) option classes in 2000 (2002)
Harris and Mayhew (2005)	OPRA (old)	01/2003	Options on 451 stocks
Battalio and Schultz (2006)	OPRA (old)	01/2000–06/2000	Options on up to 49 stocks
Holowczak, Simaan, and Wu (2006)	OPRA (old)	05/2002–07/2002	Options on the 40 most actively traded stocks
Anand and Chakravarty (2007)	OPRA (old)	07/1999–08/1999	Options on 100 sample firms
Simaan and Wu (2007)	OPRA (old)	01/2002	Options on the 50 top stocks by volume
Battalio and Schultz (2011)	OPRA (new)	08/2006–10/2008	Options on stocks subject to the short sale ban and a matched sample

(continued)

Table 2. (continued)

Authors	Dataset	Sampling Period	Sampled Options Classes
Birru and Figlewski (2012)	OPRA (new)	09/2006–10/2006, 09/2007–10/2007, 09/2008–11/2008	December-expiry options on S&P 500 Index
Muravyev, Pearson, and Paul Broussard (2013)	OPRA (new)	04/2003–10/2006	Options on 36 liquid U.S. stocks and 3 ETFs
Cakici, Goswami, and Tan (2014)	OPRA (new)	05/2010	Options on S&P 500 and S&P 100 constituents
Holowczak, Hu, and Wu (2014)	OPRA (new?)	02/2006–12/2006	Options on QQQQ
Hu (2014)	OPRA (new?)	04/2008–08/2010	Options on all individual stocks
Mishra and Daigler (2014)	OPRA (new?)	10/2008–12/2008, 10/2009–12/2009	Options on SPX and SPY
Chatrath et al. (2015)	OPRA (new)	01/2011–05/2012	Options on S&P 500 Index
Anand, Hua, and McCormick (2016)	OPRA (new)	01/2007–12/2010, 11/2012–01/2013	Options traded on NYSE Arca and a matched sample
Muravyev (2016)	OPRA (new)	04/2003–10/2006	Options on 39 most actively traded stocks (including four ETFs)
Amaya, Bégin, and Gauthier (2018)	OPRA (new?)	07/2004–12/2012	Options on S&P 500 Index
Zhang (2018)	OPRA (new)	01/1996–01/2015	Options on S&P 500 Index and its constituents, and sector ETFs
Muravyev and Pearson (2020)	OPRA (new)	04/2003–10/2006	Options on 39 most actively traded stocks (including two ETFs)
	LiveVol (?)	01/2004–04/2013	
Simon (2013)	LiveVol	05/2005–04/2010	Options on SPY ETF
Battalio, Shkilko, and Van Ness (2016)	LiveVol	05/2010–06/2010	3233 options classes on all stocks and ETFs
Christoffersen et al. (2018)	LiveVol	01/2004–12/2012	Options on S&P 500 constituents
Battalio, Figlewski, and Neal (2020)	LiveVol	03/2010	Options on 2945 stocks

We report the number of option classes before any filtering. BODB stands for the Berkeley Options Data Base. For datasets with unidentifiable data source, we insert a question mark.

also includes depth information at the best quotes for each exchange. The new OPRA dates back to July 2004, but is discontinued after September 2019. Moreover, not all papers using new OPRA acquire data from CBOE, for example, Battalio and Schultz (2011) cites a market maker as their source, Hu (2014) uses data from Trade Alert LLC, while Muravyev, Pearson, and Paul Broussard (2013) and Muravyev (2016) rely on data from NANEX, time-stamped every 25 ms. The OPRA data in Anand, Hua, and McCormick (2016) is preprocessed by the Baruch College Options Data Warehouse, and the source for Zhang (2018) is Thomson Reuters Tick History (TRTH). These third-party vendors receive feeds directly from OPRA, but apply various data cleaning and aggregation procedures which, inevitably, imply some (unknown) loss of information. For example, TRTH processes the raw quote messages from OPRA and provides time-stamped NBBO quotes, but

exchange identifiers for the NBBO quotes and exchange-level best quotes are not provided by TRTH.

As recent alternatives to new OPRA, LiveVol,[13] also based at the CBOE, offers two separate intraday datasets. One contains to-the-minute NBBO quotes and trades for all exchanges, obtained by aggregating all underlying individual OPRA entries. A second dataset provides all (tick-by-tick) option transaction prices and volume along with the concurrent best bid–ask quotes for the option and the underlying (time and sales data). These datasets have dramatically reduced granularity relative to OPRA new, but have the advantage of a more manageable size, and the former includes the disseminated NBBO quotes.

In summary, OPRA new is more granular than all of the alternative datasets referenced in Table 2. Moreover, along with LiveVol and TRTH, it is the main source of information regarding recent intraday option market activity. Hence, our study of the OPRA new sample, covering 01/2015–08/2015, provides a detailed look at the most granular option dataset available to scholars.

3 DATA OVERVIEW, SPECIAL RECORDS, AND DATA FILTERING

In this section, we provide a broad overview of the bulk OPRA data, select a working sample, and inspect the associated trade and quote records. Next, we categorize trades and quotes, which might be irrelevant, or even detrimental, for certain types of analyses. Furthermore, we explore the frequency with which such potentially problematic records appear in the bulk data. Finally, we explore the extent to which intraday option quotes recorded by OPRA imply violations of a basic no-arbitrage put–call parity relation.

3.1 Data Overview

The CBOE OPRA Data (Bulk) package covers all transactions and top-level quotes disseminated from all U.S. option exchanges on a millisecond basis in accordance with the OPRA Plan. Each record reflects either a quote or trade event realization for one of the available contracts [identified with the underlying, expiration date (tenor), strike price and put versus call type] on one of the U.S. option markets.

Each transaction record displays the price and corresponding trading volume. Each quote record contains top-level bid and ask prices along with the quoted amounts, implying that each such record reflects an update of a bid–ask pair (a change in the quoted prices or amounts) relative to the preceding quote record for a given exchange market. In addition, each option quote or trade record contains the most recent first-level quotes for the underlying.[14] A more detailed description of the content and structure of the OPRA dataset is provided in Online Appendix B.

Our dataset spans the first eight months of 2015, containing a total of 167 trading days. Altogether, we identify 3686 equities, 566 exchange-traded funds (ETF) or exchange-

13 https://datashop.CBOE.com/option-quotes-intervals (accessed 23 October 2020).

14 This applies only when the underlying is a tradable instrument. For example, for SPX options, such quotes are not available, while records for SPY options, written on a tradable ETF, contain the most recent top quotes of the underlying. For an in-depth comparison between SPX and SPY options, see Mishra and Daigler (2014).

traded notes (ETN), and 47 indices as underlying instruments. In addition, 335 underlyings have a non-standard deliverable, resulting in multiple listed option symbols for the same underlying entity.[15] On average, 160 contracts are listed per option symbol on a daily basis but, for the most liquid underlyings, there may be up to 4000 different contract variations. In terms of tenor, 45% of the contracts may be classified as standard (equity, ETF, or index options), while weekly (20%), quarterly (25%), and long-term equity anticipation securities (LEAPS, 10%) options are listed in sizable proportions as well. In contrast, Mini Options rarely appear and are only found for five of the most traded equities and ETFs.[16]

Between January and August 2015, OPRA recorded 1.22 trillion quotes and nearly 159 million trades, as the trading volume exceeded 2.76 billion contracts with a total notional value of $7.95 billion. Table 3 summarizes the average daily quote and trade activity for the three main asset classes, with the most active constituents within each class being AAPL, SPY, and SPX. These underlyings have multiple option classes, but for illustrative purposes we only report statistics for the standard categories.[17]

To provide a representative overview of the OPRA data, while keeping it reasonably succinct, we select a small subset of the underlyings. This sample of 12 stocks and 5 ETFs from different sectors mirrors the general heterogeneity in terms of the (average) number of records (see Table 4). We focus on equities and ETFs to study the cross-exchange patterns, which are absent for index options. Furthermore, we exclude option classes such as Mini and Jumbo options, or corporate-action adjusted ones.[18]

3.2 Potentially Irrelevant or Faulty Observations

Depending on the analysis, some OPRA records might be irrelevant, redundant, or even introduce undesirable noise through data errors or market microstructure peculiarities. We classify such special records in line with the criteria used for algorithms developed for cleaning the TAQ data in the prior literature (Brownlees and Gallo, 2006; Barndorff-Nielsen et al., 2009). Our classification contains, however, several categories specific to high-frequency options data.

3.2.1 Classification of special records

We identify six categories of OPRA records that, depending on the context, may be suppressed. We characterize these special records in Table 5, along with the detailed rules for authentication. For each group, we introduce a filter that applies separately for trade and quote records from a given exchange.

The F1 category consists of all observations recorded before the start of the regular trading session (8:30 CT), or after the close (15:00 CT for stock options and 15:15 CT for ETF/ETP options). Although some exchanges (e.g., BATS) accept early quotation, such records

15 Often, non-standard deliverables are Mini and Jumbo options, but can also be, for example, corporate-action affected stocks.

16 There is trading in some non-standard S&P 500 index contracts, such as *binary* (BSZ) and *range* (SRO) options. Another *binary* option exists for the S&P 500 Volatility Index (BVZ). These contracts are only listed and traded on the CBOE.

17 For example, the mini options AAPL7 and SPY7 were excluded.

18 However, by including AAPL and SPY, two of the most liquid assets among the 4562 option class symbols, we still cover more than 10% of the entire OPRA quote data, see Table 3.

Table 3. OPRA average daily statistics from January to August 2015

	Asset classes			Specific examples		
	Equity	ETF/ETN	Index	AAPL	SPY	SPX
Underlyings	3686	566	47			
Option symbols	3814	585	53			
Option classes	620,123	156,162	33,690	1169	3685	2544
Quotes	5,021,787,050	1,965,243,131	321,813,772	226,822,053	520,188,594	10,183,398
Trades	678,120	222,143	50,896	77,849	75,509	10,994
Volume	8,871,502	6,024,324	1,672,754	1,029,155	2,521,188	493,663
Notional	$19,523,973	$9,016,319	$19,080,347	$3,436,161	$4,534,401	$13,133,924
MPS	214,635	83,994	13,755	9697	22,234	436
QPC	8098	12,585	9552	194,031	141,164	4003
QPT	7405	8847	6323	2914	6889	926
QPV	566	326	192	220	206	21
QPN	257	218	17	66	115	0.78

MPS refers to messages per second, while QPC, QPT, QPV, and QPN are the number of quote messages divided by option classes, trade messages, traded number of contracts, and notional trading volume in dollar terms, respectively.

are often noisy and may induce misalignments in terms of timing, if observations from multiple exchange markets are considered jointly.

The F2 category includes entries that likely involve data error. In particular, we identify trade records with a zero transaction price or zero size. Likewise, we filter quotes associated with negative spreads and entries with zero offer price or size. Note that, in contrast to the stock market cleaning algorithm from Barndorff-Nielsen et al. (2009), we do not eliminate quotes with zero bid prices (and positive offer prices). This occurs regularly for deep out-of-the-money options and is typically not associated with an error.

The F3 category represents trades and quotes with specific conditions, indicative of potentially irregular features of the given entry. This occurs for records of cancelled trades and non-firm quotes.

The F4 group contains quote records of minimum size. Such quotes might be less informative about the latent equilibrium option price. The presumption is that market makers experiment, when uncertain about the fair price, by posting minimally sized quotes at a distinct price level—simply to uncover latent demand with minimal exposure. Alternatively, due to the obligation to continuously maintain quotes throughout the trading day (see Table A.2 in Online Appendix A), market makers may at times post minimum sized quotes to satisfy requirements without providing genuine liquidity. Thus, we flag quotes when the size is less than or equal to a single contract at both the bid and offer side.[19]

19 We do not remove quotes, which feature minimal size at only one side. Such scenarios often materialize when one side of the market dominates the quoting activity. Specifically, for deep out-of-the-money options ask orders tend to dominate, while, conversely, bid orders prevail for in-the-money options.

Table 4. Sample of underlying assets for the descriptive analysis

Sector	Asset	Ticker	Name	Industry	Average records (per day)
Consumer	Equity	PG	The Procter & Gamble Company	Personal Products	22,194,052
staples	Equity	WMT	Wal-Mart Stores, Inc.	Discount, Variety Stores	20,171,655
	Equity	K	Kellogg Company	Processed & Packaged Goods	5,461,349
	ETF	XLP	Consumer Staples Select Sector SPDR Fund		5,517,801
Energy	Equity	XOM	Exxon Mobil Corporation	Major Integrated Oil & Gas	31,764,057
	Equity	CVX	Chevron Corporation	Major Integrated Oil & Gas	28,095,240
	Equity	MPC	Marathon Petroleum Corporation	Oil & Gas Refining & Marketing	4,389,222
	ETF	XLE	Energy Select Sector SPDR Fund		42,901,922
Financial	Equity	JPM	JPMorgan Chase & Co.	Money Center Banks	32,209,032
	Equity	BAC	Bank of America Corp	Money Center Banks	24,726,836
	Equity	BLK	BlackRock, Inc.	Asset Management	1,701,449
	ETF	XLF	Financial Select Sector SPDR Fund		7,465,856
Technology	Equity	AAPL	Apple, Inc.	Electronic Equipment	226,899,902
	Equity	GOOG	Alphabet, Inc.	Internet Information Providers	61,812,973
	Equity	EA	Electronic Arts, Inc.	Multimedia & Graphics Software	11,184,655
	ETF	XLK	Technology Select Sector SPDR Fund		6,963,341
Global	ETF	SPY	SPDR S&P 500 ETF		520,264,103

We caution that, since only the top bid and ask quotes are available in (new) OPRA, we cannot always identify whether a quote is of minimal size. For example, whenever a bid or ask quote changes, we observe a different set of contracts available at the new level. Since this figure is unobserved prior to the quote shift, it is impossible to determine the exact size of the new quote. As a result, we allocate only those minimally sized quotes, that can be identified with certainty, to category F4.

The F5 group contains outliers. For trades, we focus on abnormally high or low transaction prices relative to the current spread. For quotes, we identify entries with excessive spreads, extraordinarily high or low mid-quotes, and entries for which the ratio between ask and bid is unusually high. Our detection rules are similar to those of Barndorff-Nielsen et al. (2009) for trades and quotes. Such outliers may introduce unwarranted irregularities into the data structure, complicating the analysis. In particular, outliers may reflect data errors or market microstructure artifacts, such as stub quotes.

Table 5. Classification of special trade and quote OPRA records according to six groups

Group	Description	Trades	Quotes
F1	Records outside the regular trading hours	Entries with timestamps outside the normal trading period (from 8:30 until 15:00 CT for underlying stocks and until 15:15 CT for underlying ETF/ETP)	
F2	Records with possible misrecordings and data errors	Entries with zero transaction price or size (b) Entries with a negative spread	(a) Entries with zero offer price or size
F3	Records with irregular conditions	Entries which correspond to canceled transactions (observations with condition codes "A", "C", "E", "G", and "O")	Entries which correspond to non-firm quotes (observations with condition code "F")
F4	Records corresponding to non-informative quotes		Entries for which both bid and offer sizes do not exceed 1 contract
F5	Possibly outlying records	Entries for which a transaction price is either lower than the current bid price minus the current spread, or higher than the current offer price plus the current spread (b) Entries for which the mid-quote is by more than 10 mean absolute deviations apart from the rolling centered median (that is based on a rolling window of 50 observations) (c) Entries for which the ratio of offer price to bid price exceeds 5 when the bid price is non-zero.	(a) Entries for which the spread exceeds 50 median spreads on that day
F6	Records with identical timestamps	At the millisecond frequency, can be replaced with a single entry with median/mean/volume-weighted transaction price and the total size. For lower frequencies, the observations with the latest millisecond time stamp can be used.	At the millisecond frequency, can be replaced with a single entry with median/mean/volume-weighted bid and offer prices and the total sizes. For lower frequencies, the observations with the latest millisecond time stamp can be used.

Records with identical time-stamps are collected in category F6. Often, multiple quotes or transactions are recorded at the same millisecond. When treating the observed option prices as a time series, it is convenient to replace such entries with a single record. This can be done in several ways.

First, OPRA captures the exact execution time for each record, so the final entry within a given millisecond may be used as the single price observation in lieu of the multiple records with the same time stamp. Obviously, this approach has drawbacks, as it discards

the information contained in the other records and, in particular, fails to take advantage of the possibility of constructing a more robust observation from the available records. An alternative is to develop a procedure that aggregates the "simultaneous" records within a given millisecond by computing the median, mean, or volume-weighted average of the trade/quote prices and summing up individual trade/quote sizes. Specifically, for time series analysis, where the frequency is considerably lower than a millisecond, for example, a second or minute, it is sensible to consider an approach that mitigates the distortions induced by potential outliers or other microstructure noise effects associated with the reliance on a single end-of-interval observation.

We note that the frequency of records in group F2 and, to a lesser extent, F5 might provide a signal regarding the overall data quality, as such observations are more likely to be associated with data errors.

Finally, we stress that the filters are intended solely as a tool for option selection in the context of a given research objective. They can be ignored or imposed independently of each other. Moreover, they apply separately to the trade and quote records from a given exchange. If a specific study needs to guard against data errors, deal with multiple records at a single timestamp, or remove noisy and outlying observations, the filters in Table 5 provide a guide for identifying the relevant records in the dataset.[20]

3.2.2 Summary statistics of special records

We inspect trade and quote records from our selected raw OPRA option sample throughout all trading days in January 2015 for contracts expiring on February 20, 2015. This sample consists of 2.77 billion quote and 730,000 trade records. Tables A.3 and A.4 in Online Appendix A provide detailed aggregated statistics on the presence of records in the special categories identified above across underlyings and exchange markets, respectively. Figures 2 and 3 summarize key aspects of the findings. They report the average fraction of daily records belonging to one of the categories F1–F6 (white bars) and the fraction belonging to multiple categories (colored bars). Figure 2 depicts the fraction of special option records for each underlying asset. Figure 3 displays the fraction of special option records for each exchange venue.

Quote records are far more numerous than trade records. For example, in January 2015, SPY options expiring on February 20 feature, on average, more than 66 million daily quotes across the exchanges[21] compared with less than 16,500 transactions. The average number of special entries for SPY options exceeds 8.4 million quotes (more than 12% of all SPY quote records) and about 2000 trades (again, more than 12% of all SPY trade records) per trading day.

Depending on the underlying asset, 5–25% of the trade records have non-unique millisecond time stamps (category F6), which may reflect the execution of larger aggressive orders hitting several smaller limit orders in the book simultaneously. In fact, this group constitutes more than 99% of the special trade records. Importantly, the categories F1, F2, F3, and F5 are largely absent for trade data. This suggests that the trade records are

20 Studies focusing on specific topics will often apply additional filters. For example, Christoffersen et al. (2018) impose a positive daily volume requirement and check for violations of minimal tick size rules implied by the option quotes.

21 It implies an average of about 240,000 quotes for the 282 distinct contract specifications.

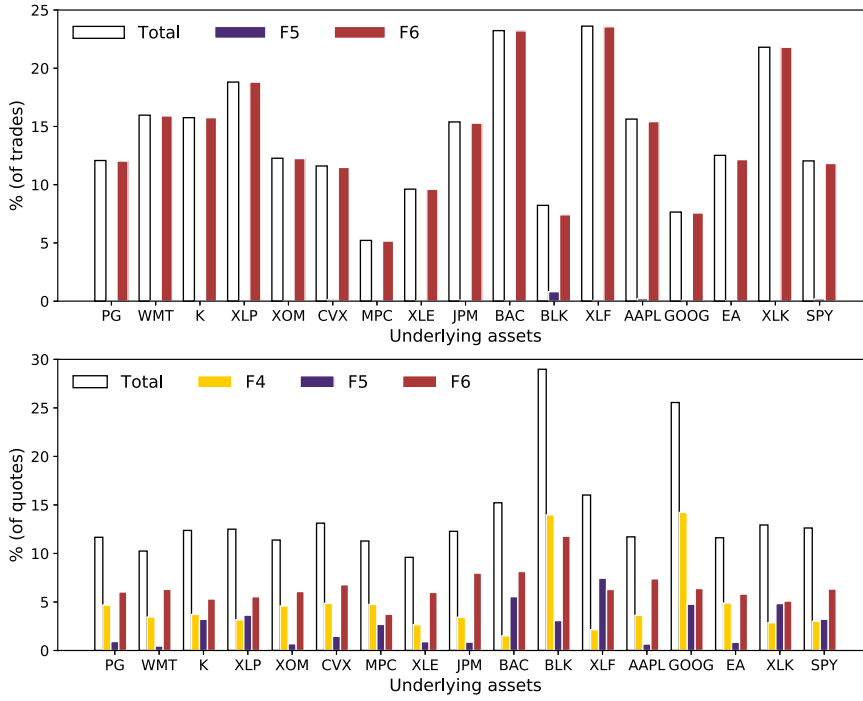

Figure 2. Average daily percentage of special option trade (top panel) and quote (bottom panel) records across selected underlying assets. White bars correspond to the total fraction of special records in the raw data (all categories, F1–F6). Yellow bars (available for quotes only) correspond to the fraction of records with minimal quote size (category F4). Blue bars correspond to the fraction of potentially outlying records (category F5). Red bars correspond to the fraction of records with the same millisecond time stamps (category F6). The results cover data for all options traded in January 2015, expiring on February 20, 2015, observed across all available exchange markets.

remarkably clean and free of extreme outliers. As seen from Figure 3, trades with identical timestamps are especially prevalent on the NYSE Amex (exchange symbol A), BOX (B), and MIAX Options Exchanges (M), reaching 30–35% of all trade records on these platforms. In contrast, for Nasdaq BX (T) and C2 (W), the fraction of such trade records is less than 1.5%.

In total, between 10% and 30% of the quote records for a given option class are deemed special and fall into one of six categories. Quote records outside regular trading hours (F1) or appearing erroneous (F2) are rare (less than 0.1%). Non-firm quotes (F3) are also infrequent, never exceeding 0.5% of the total quote records. Minimal two-sized quotes (F4) are common on all exchanges, exceeding 15% of all quotes for the BOX exchange (B). Quote records representing potential outliers (F5) amount for up to 7.5% of the total for certain underlyings, and are about equally common across all markets. Figure 2 shows that the fraction of quotes with identical millisecond stamps (F6) is substantial, varying between 3% and 12% of all entries across underlyings. For example, for Apple options, such quotes amount to almost 2 million entries per day. Although group F6 contributes substantially to the amount of special records, its relative contribution is less important than for trades.

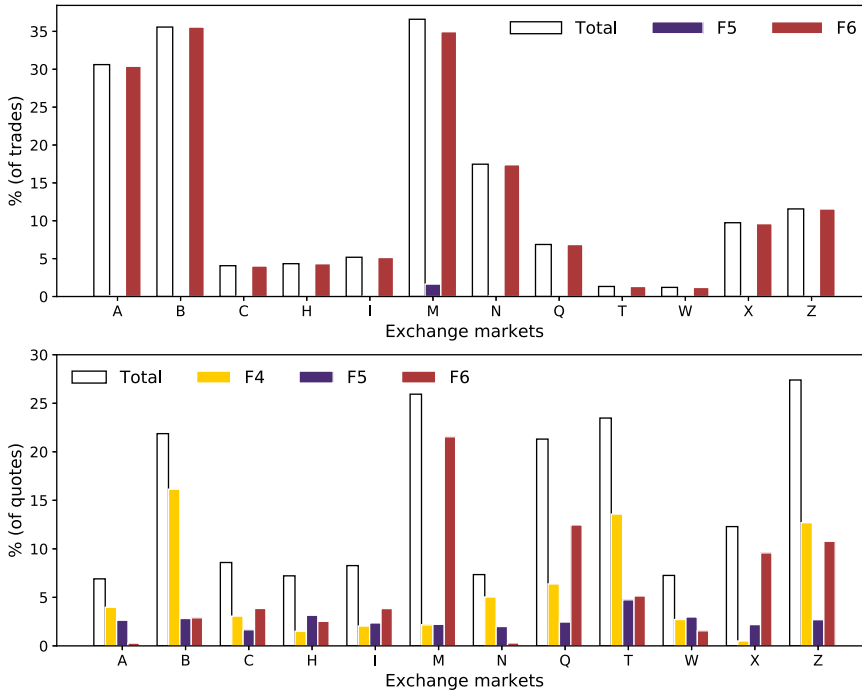

Figure 3. Average daily percentage of special option trade (top panel) and quote (bottom panel) records across 12 option exchange markets. White bars correspond to the total fraction of special records in the raw data (all categories, F1–F6). Yellow bars (available for quotes only) correspond to the fraction of records with the minimal quoting size (category F4). Blue bars correspond to the fraction of potentially outlying records (category F5). Red bars correspond to the fraction of records with the same millisecond time stamps (category F6). The results are based on the data for all option contracts traded in January 2015, which expire on February 20, 2015, observed on all available exchange markets.

Thus, we conclude that OPRA quote records are mostly void of obvious data errors and invalid observations, but contain a non-trivial amount of potential outliers, which may require some attention, depending on the research objective.

3.3 Intraday Deviations from Put–Call Parity

Identifying records violating no-arbitrage principles is important for some applications, as it suggests the presence of frictions, market failures, faulty data entries, or non-synchronous recording of option and underlying asset prices. Consequently, the frequency of such violations at a given venue may be indicative of the relative exchange efficiency and the reliability of the associated option price and quote record.

We hasten to add that minor deviations from the put-call parity may not offer actual arbitrage opportunities. Whether the violation is economically relevant hinges on the proxy for the risk-free interest rate, representing a cost of capital for the option trader, and certain

shadow costs associated, for example, with margin requirements and inventory control.[22] As such, one may think in terms of "apparent" violations instead. Nonetheless, near arbitrage violations may also be informative, as they reflect scenarios with high call option quotes relative to the corresponding put options. In particular, the measures capturing the put–call parity deviations potentially can be relevant for asset pricing more broadly. For example, Cremers and Weinbaum (2010) find that such end-of-day metrics are indicative of price pressures, which provide significant predictability for future asset returns.

Theory predicts that the price of European options must be convex and monotone functions—increasing for put and decreasing for calls—of the strike price, see, for example, Breeden and Litzenberger (1978). For brevity, we focus on one specific relation—the put–call parity—which establishes a no-arbitrage pricing relationship between put and call options at a given strike jointly with the price of the underlying asset.[23]

For American options, put–call parity cannot be represented as an equality. Instead, certain inequalities apply, including the following[24]:

$$C_t + Ke^{-r_t\tau} \leq P_t + S_t,$$

where C_t and P_t are call and put option prices at time t, respectively, K is the strike price, S_t is the underlying asset price, r_t denotes the risk-free rate, and τ is the tenor of the option pair. Consequently, put–call parity is violated if

$$C_t^{\mathrm{B}} > P_t^{\mathrm{A}} + S_t^{\mathrm{A}} - Ke^{-r_t\tau},$$

where the superscripts A and B refer to ask and bid prices, respectively.

For all underlying assets in our sample, we identify violations of this inequality using intraday option quote records separately for each of the 12 exchanges. We restrict our attention to data from 20 trading days in January 2015 and consider only options expiring on February 20, 2015. Consequently, the tenor ranges from seven to three weeks. For each trading day, underlying security and strike price, we inspect all put–call quote pairs, checking the put–call inequality record-by-record.

Table A.5 in Online Appendix A presents aggregate results from this analysis, spanning all 17 underlying assets for each of 12 option venues that were active at the time. The results reveal strong heterogeneity in the (apparent) violations across options for different underlyings. Quotes for JPM, BAC, and SPY options display the highest rate of violations,

22 For example, when using the 3-month Libor instead of the Treasury rate in our analysis, we detect significantly lower numbers of such violations.

23 The filters described in the previous section might also identify records that violate general no-arbitrage principles. To identify all records violating a given restriction, prices over a cross-section of options contracts (possibly along with the underlying prices) may need to be analyzed jointly. This type of comprehensive analysis necessitates additional assumptions for the construction of option cross-sections from non-synchronous intraday prices and specific criteria for identifying records that violate the specific arbitrage condition in question. In general, such studies differ substantively from the analysis of special records in the previous section, based only on the data from a single contract.

24 There is a closely related upper bound on the put plus asset price. However, given the illustrative nature of this exercise, we ignore that constraint, which involves data on the dividend yield as well.

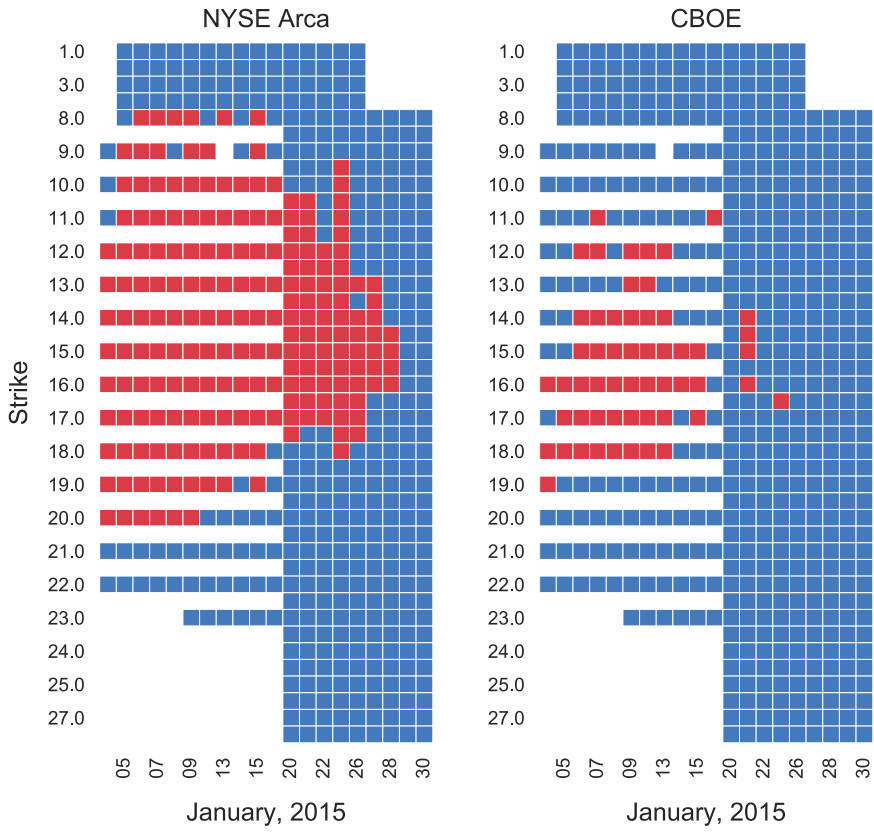

Figure 4. An illustration of the put–call parity violations for BAC option quotes posted at NYSE Arca (left panel) and CBOE (right panel) in January 2015. A square represents BAC options expiring on February 20, 2015, if the corresponding strike price (vertical axis) were available for trade on the specified date (horizontal axis). Red color implies that the put–call parity has been violated for more than 1 min for a given strike price (in dollars) on a given date. Otherwise, the blue color is used.

whereas we detect none at all for XOM, CVX, and MPC in January 2015. The variation of put–call parity violations across exchanges is much less pronounced, reinforcing the point that such violations are linked closely to features of options for specific underlying asset.

We illustrate how the (apparent) violations vary over time and depend on both moneyness and exchange venue by providing a more granular visualization in Figure 4. Each small square reflects a particular trading day and strike price in January 2015. The color is red, if put–call parity was violated for a combined period exceeding 1 min; otherwise, the color is blue. To conserve space, we display results only for the BAC option quotes recorded on two exchanges, NYSE Arca and CBOE. This choice is useful in exemplifying the type of economic events and exchange-specific features that can be explored through high-frequency options data. To put the results in context, we note that the financial sector performed poorly over the first half of the month, with the BAC stock displaying heightened volatility and suffering a cumulative loss of approximately 14% over that period.

Several patterns stand out in Figure 4. First, there is a clustering of the violations in the first half of the month, when volatility was elevated. Second, they center on the ATM strike price, which was close to \$18 at the start of the month and between \$14 and \$16 over the last two weeks. Third, they are asymmetric across the strikes, as there are more violations for lower strikes. Fourth, they are almost absent for the deep ITM and deep OTM contract pairs. Fifth, for BAC during this month, the violations are much more prevalent on NYSE Arca than CBOE.

For a different, but related perspective, Figure 5 displays the frequency with which we observe a violation of put–call parity for any one of the quoted BAC or JPM options within a given 10-min interval for a specific exchange in January 2015 with February 20 expiry. Here, two observations are striking. First, the violations are concentrated on a few venues, NYSE Arca, Nasdaq Options Market (NOM), and BATS plus, in the case of BAC, also ISE Gemini. Nonetheless, violations are observed occasionally for almost every venue. Second, the violations seem to be less frequent around the opening of the trading day, when the underlying assets typically display high volatility and wide bid–ask spreads.

The above observations raise numerous research questions that fall outside the scope of the current paper. One, is the volatility of the underlying a primary driver in the put–call parity violations? Two, does the direction of the violations possess short-term predictive

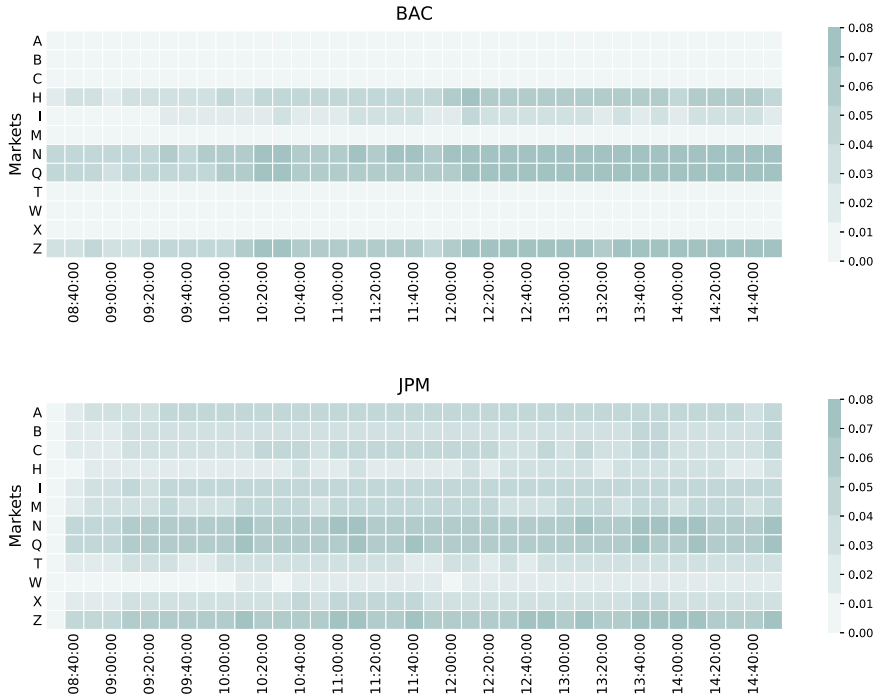

Figure 5. Intraday frequency of put–call parity violations for BAC and JPM options. The figure displays the frequency of put–call parity violations for BAC and JPM option quotes across all exchange venues in January 2015, for options which expire on February 20, 2015. The violations are compiled for 10-min intervals across the trading day. The color coding reflects the frequency of observed violations within the given interval and exchange across the full month.

power for the asset return? Three, do effective bid–ask spread measures (using trade prices relative to mid-quotes) also display arbitrage violations? Four, do the violations cluster around the ATM strike because the posted spreads are particularly narrow, in an economically meaningful sense, for this (liquid) part of the strike range? Five, why do we observe large discrepancies in violations across exchange venues? Six, does the option bid–ask spread display strong intraday patterns, possibly explaining why there may be fewer violations during morning trading? In Section 4, we present some findings related to the three latter points, so, at that stage, we shed a bit more light on a few of these issues, even if we do not provide comprehensive evidence.

In summary, instances of put–call parity violation are observed across all exchanges, they do at times cluster across consecutive trading days, they are more frequent on certain exchanges, they are more prevalent for some underlyings at a given point in time, and they vary in intensity across the trading day. Although such observations often are filtered out, they may potentially bring forth new and interesting market microstructure issues, and even broader asset pricing questions. However, we also recall the opposite argument: the violations may be economically inconsequential, as other (shadow) costs may render exploitation unprofitable, or they arise solely because the cost of capital is understated. Irrespective, a thorough analysis, expanding on our approach above, requires a comprehensive dataset like the bulk OPRA package, as one needs to monitor the top-level quotes across all active exchanges.

4 AN EXPLORATIVE ANALYSIS OF OPRA TRADES AND QUOTES

4.1 Basic Trade and Quote Statistics

Table 6 reports the daily number of trades and quotes for option contracts written on the 17 underlying assets in January 2015, before and after applying all the filtering algorithms in Section 3.2. On average, 14.53% of the records are eliminated due to this (aggressive) data cleaning. The number of option quotes and trades varies greatly across underlyings, reflecting largely the liquidity of the latter. The most actively quoted (and traded) option contracts are those written on the ETF SPY and Apple, with on average, respectively, more than 628 million and 279 million clean quotes daily. On the most active trading day, the SPY quotes exceed 872 million after filtering and close to 1 billion prior to cleaning.

The average order-to-trade ratio ranges between 3036.9 (for BAC) and 22,433.6 (for EA), implying a dramatic excess of quotes relative to trades in option markets. This is an order of magnitude larger than for equity markets, where order-to-trade ratios rarely exceed 100, even for algorithmic and high-frequency traders, see, for example, Hagströmer and Nordén (2013) and Brogaard et al. (2015).

4.2 Trade and Quote Records by Tenor and Moneyness

The usual expiration day for a standard option contract—following conventions adopted when exchange-based option markets were initially established—is the third Friday of the month. Given the diverse economic incentives for option trading, the subsequent successful introduction of quarterly and yearly option contracts for most underlyings is not surprising. In addition, over the past decade, weekly options, or weeklies, have increased dramatically in importance. For instance, the trading volume of S&P 500 weeklies (SPXW) grew from

Table 6. Descriptive statistics for daily OPRA data records in January 2015

Ticker	Avg	Standard deviation	Median	Min	Max
Number of records (before cleaning)					
SPY	700,838,009	139,414,541	694,355,725	409,607,244	984,010,543
AAPL	301,575,942	78,014,816	310,310,810	159,094,428	417,942,352
GOOG	73,543,546	17,064,430	71,363,687	45,479,752	113,190,636
EA	15,184,625	4,269,292	14,629,507	9,744,870	23,388,291
XLK	9,489,613	2,406,511	8,729,688	5,399,868	13,105,051
XOM	36,198,297	7,548,864	34,021,556	24,735,171	49,712,308
CVX	28,728,799	5,085,071	27,206,324	21,143,683	38,714,736
MPC	5,215,386	1,290,710	5,066,729	3,245,001	8,314,240
XLE	59,890,142	11,537,807	58,010,337	45,196,390	84,261,758
PG	23,723,427	8,134,120	24,989,843	9,543,493	38,230,503
WMT	22,802,939	6,288,729	23,184,688	12,224,635	33,970,789
K	745,353	120,562	718,706	574,871	981,853
XLP	8,778,407	2,341,311	8,477,310	4,851,901	13,133,224
JPM	43,001,824	8,747,138	41,790,218	25,718,345	59,398,468
BAC	36,051,422	6,725,070	34,030,353	23,281,419	49,911,031
BLK	3,166,838	724,910	3,139,884	1,968,863	5,340,371
XLF	11,930,495	2,642,094	11,100,126	6,549,215	17,999,368
Number of trades (after cleaning)					
SPY	80,476	15,585	80,004	44,787	104,560
AAPL	78,298	27,043	72,670	45,928	162,313
GOOG	8078	3932	6637	4947	20,738
EA	607	643	424	185	2898
XLK	462	125	456	224	706
XOM	3846	1028	3820	1847	5809
CVX	2979	1084	2828	1568	6149
MPC	405	261	301	143	936
XLE	3635	1223	3277	1933	6542
PG	1728	742	1513	912	3714
WMT	1699	623	1515	858	2911
K	127	59	122	52	243
XLP	375	179	337	164	826
JPM	4117	1941	3443	1617	9552
BAC	9895	3724	8869	3649	19,100
BLK	157	79	115	55	350
XLF	1388	449	1321	685	2299
Number of quotes (after cleaning)					
SPY	628,611,886	125,112,503	622,338,806	368,646,694	872,278,526
AAPL	279,290,504	72,431,205	286,858,000	147,174,936	389,048,310
GOOG	64,487,534	14,818,791	61,416,002	40,321,726	97,324,007
EA	13,617,224	3,903,640	13,364,134	8,347,573	20,947,630
XLK	7,987,132	2,323,915	7,488,202	3,747,030	11,382,152
XOM	32,616,857	7,434,977	29,583,046	22,156,109	46,440,632
CVX	25,518,930	4,757,349	24,582,347	19,131,646	34,812,831

(continued)

Table 6. (continued)

Ticker	Avg	Standard deviation	Median	Min	Max
MPC	3,958,422	1,572,582	4,098,120	1,293,294	7,343,634
XLE	54,090,705	11,359,835	51,518,109	37,783,596	77,700,703
PG	21,348,654	7,476,082	22,057,529	8,222,172	34,932,632
WMT	20,543,501	5,899,002	20,700,550	10,856,256	31,134,228
K	508,302	136,216	465,697	318,438	838,467
XLP	7,325,981	2,196,516	7,268,446	3,309,087	11,548,186
JPM	38,322,047	7,838,284	366,82,359	22,522,639	51,724,816
BAC	30,050,269	5,886,635	28,184,448	19,449,233	43,635,724
BLK	2,532,815	699,342	2,534,352	1,336,222	4,508,035
XLF	9,503,469	2,421,343	9,194,951	4,186,745	14,354,507

"Trades-only" contracts are not considered for the calculation.

about 10% in early 2010 to nearly 30% in mid-2015 (Andersen, Fusari, and Todorov, 2017) .

According to current OCC contract specifications, regular options expire on the third Friday each month, weeklies on any other near-term Friday (up to five consecutive weeks), quarterlies on the financial quarter-end (not necessarily Fridays), and LEAPS are characterized by tenors greater than 12 months. For the first eight months of 2015, we classify about 45% of the contracts in our sample as regular (i.e., equity, ETF, or index options), but weekly (20%), quarterly (25%), and LEAPS (10%) options are also listed in considerable proportions. To convey how quotes and volume are distributed across maturities, we categorize options as ultra short-, short-, medium-, or long-term, depending on tenor

$$\text{Time to expiry} := \begin{cases} \text{ultrashort} - \text{term} & \text{for } T - t \leq 7, \\ \text{short} - \text{term} & \text{for } T - t \in (7, 30], \\ \text{medium} - \text{term} & \text{for } T - t \in (30, 90], \\ \text{long} - \text{term} & \text{for } T - t > 90, \end{cases}$$

where t and T denote the current and expiration date, respectively, measured in calendar days.

Figure 6 displays the average daily proportion of quotes and trading volume by tenor. For most option classes, the trading volume is relatively more concentrated among the ultra short-term contracts compared to the corresponding quoting activity. For example, listed options on GOOG can be traded for 12 separate expiration dates (14 for AAPL and 24 for SPY), but contracts for the nearest maturity account for 51.5% of all trades (48.1% for AAPL and 40.0% for SPY) relative to about 15% of the quotes (22% for AAPL and 8% for SPY). Hence, not only is the weekly maturity profile increasing in popularity and trading activity, but less liquid option classes without weeklies, for example, BLK, MPC, and K with 46.2%, 17.0%, and 25.1%, respectively, are also traded more intensively in the week prior to expiration.

We now turn toward the option trade and quote activity as a function of moneyness. Figure 7 tabulates the percentage of option contracts traded across different degrees of moneyness between January 2 and February 18, with expiration February 20, 2015. Across the securities, the vast majority of the option trading is in close-to-the-money options. For

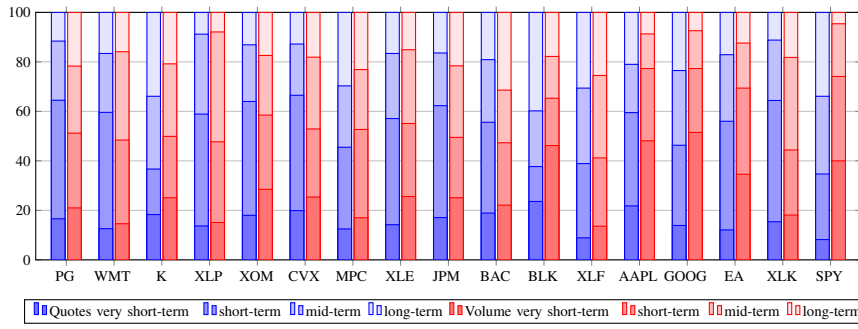

Figure 6. Average daily proportions of quote and trading volume by tenor (in %). The two stacked bar charts represent the proportion of quotes (left) and volume of traded option contracts (right) for each underlying between January and August 2015. The bars are partitioned in segments by *time-to-expiry*, as ultra short-term (bottom), short-term, medium-term, and long-term (top).

the individual stocks, the trading volume for options with positive moneyness tends to exceed that of options with negative moneyness, when the options are within two (Black-Scholes) "sigmas" of the ATM strike. However, the reverse is true, when considering moneyness beyond two "sigmas," where the options with "downside" strikes command the highest volume, almost uniformly across all the underlyings in our sample, and often by a wide margin. The equity indices are clearly different, as we observe a very strong asymmetry toward excess trading in options with below ATM strikes, irrespective of the degree of moneyness. This is, of course, consistent with a strong motive for downside hedging by fund managers who hold net long positions in equities.

Figure C.3 in Online Appendix C confirms that these findings apply qualitatively for the total option trading volume as well, while the quote activity is much more balanced for "up" and "down" strikes, as seen in Figure C.4, even though a slight tilt toward the downside strikes remains.

4.3 Activity across Exchange Platforms

A distinct feature of the OPRA data is the availability of trade and quote records disseminated by all of the individual U.S. option exchanges. Figure 8 shows the relative proportion of quoting and trading volume across exchange holding groups and individual exchanges. In general, the pattern for the trading activity across exchanges is in line with the evidence from Figure 1.[25]

Within each holding group, the longest-serving venues account for the majority of the trading, as seen from the bottom right chart in Figure 8. The same is generally true for quotes, which typically are more than twice as frequent for these exchanges than for the secondary venues within each group. However, NYSE is an exception, as the AMEX (A) exchange was launched three years after ARCA (N), but has double the number of quotation messages by 2015. Both markets offer floor and complex trading, but they differ in

25 This figure is based on market activity in January 2015, when ISE and BATS were still independent from NASDAQ and CBOE, respectively. Furthermore, the option exchanges MCRY, EDGX, MPRL, and EMLD had not yet been launched.

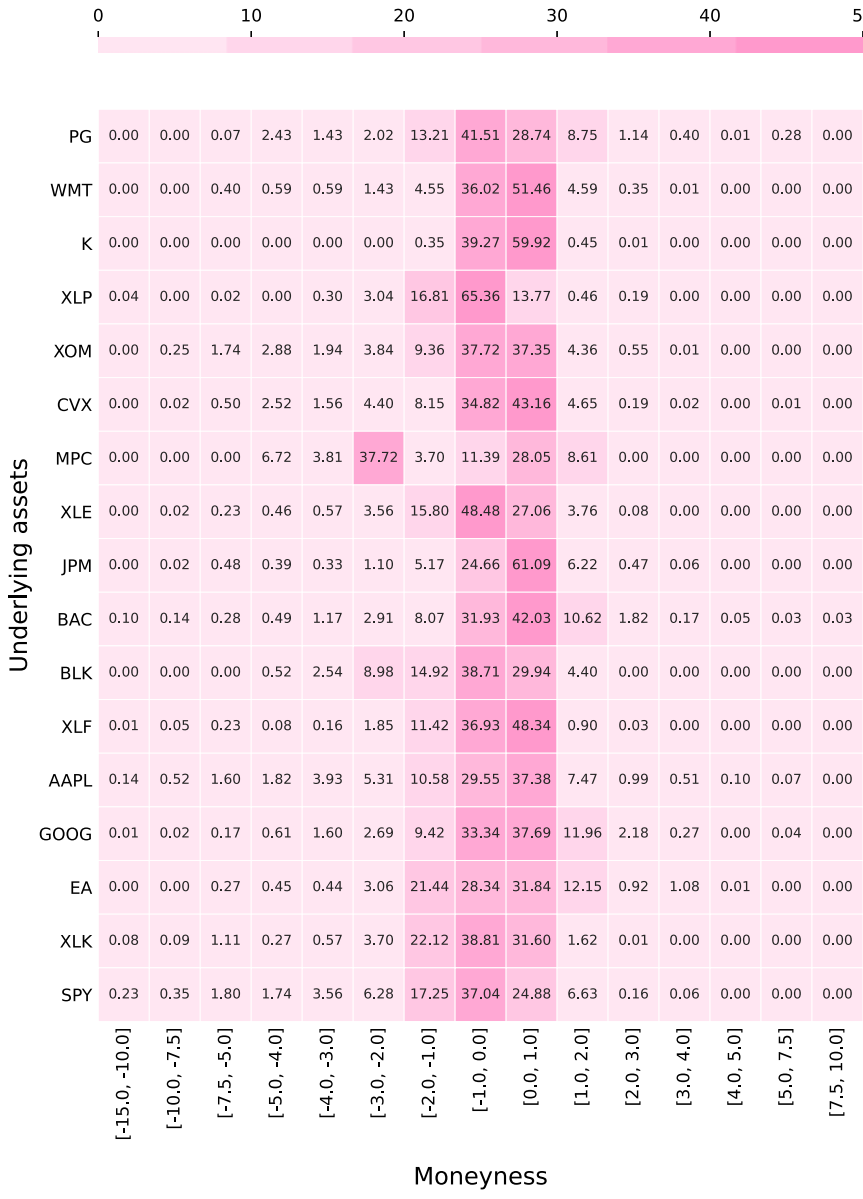

Figure 7. Proportions of contracts traded (in %) by moneyness for the selected option classes. The results are obtained using the OPRA quote records for put and call option contracts traded between January 2 and February 18, 2015, which expire on February 20, 2015. The moneyness is defined as $m = \log(K/F_t)/(\sigma_t\sqrt{\tau})$.

their pricing and allocation structures, as documented in Table A.1 of Online Appendix A. AMEX utilizes the classical *customer-priority* model with a *pro-rata* allocation, which encourages deeper liquidity. In contrast, ARCA focuses on price efficiency, exploiting

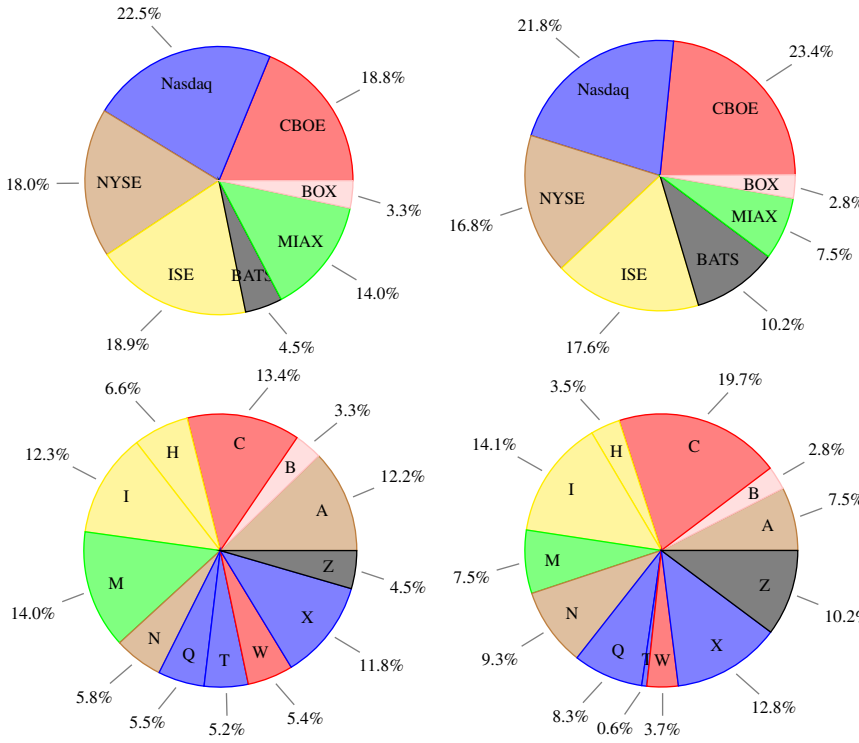

Figure 8. Quotation and trading volume market shares per exchange. Pie charts on the left-hand side reflect the quotation distribution among holding groups (top) and individual exchanges (bottom) for our sample of underlying assets. Analogously, the trading volumes are indicated on the right-hand side. The identity of the individual venues associated with letter codes can be gleaned from the header in Table 7.

maker-taker and *price-time* procedures. In this case, the liquidity-inducing approach is more quote-rich, but generates a lower trading volume relative to the alternative. It is evident, however, that the pricing structure is not the only factor accounting for the heterogeneity in quotation and trading activity across markets. The MIAX exchange, which also follows the *maker-taker* model, had the highest number of quote updates in 2015 among all option marketplaces in the United States for this set of underlyings.

Given our relatively short sample and the high variability of market shares in options trading, we cannot identify a clear trend in exchange competitiveness. From Figure 9, the leading groups NASDAQ and CBOE started out with a market share of approximately 25% each, ahead of ISE and NYSE with about 15%. The remaining exchange holdings, BATS, MIAX, and BOX, sported significantly lower daily volumes in 2015. By August, however, the four largest exchange holdings went head-to-head with market shares of around 20% each, suggesting a stronger degree of competition, but also a shift in trading interest toward the smaller exchanges. In addition, there was remarkable growth in the volume at BATS in August 2015, when it gained an additional 5% market share, possibly in

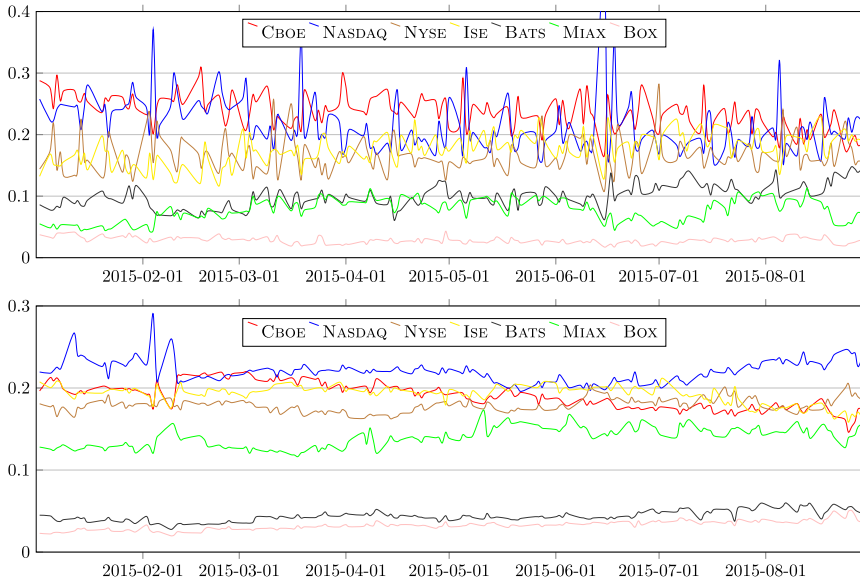

Figure 9. Market shares for the daily contract trading volume (top plot) and quotation activity (bottom plot) per exchange holding group for our option sample.

response to a bout of market stress.[26] By posting limit orders in the pre-market session, offered solely by BATS, traders obtain time priority for the opening call auction and, if not executed at that stage, within the following regular trading session.

In contrast, the relative number of quotations is fairly stable across exchanges. This is not unexpected. As the price on the underlying asset fluctuates, option prices change as well, necessitating adjustments to a wide range of quotes. Hence, if the exchanges maintain a menu of options with similar characteristics across this period, the frequency of their quote updates is naturally aligned with the volatility of the underlying assets, irrespective of the actual transactions they facilitate. Consequently, the relative quotation shares may well remain stable, unless there is an underlying shift in the market structure. Such changes are often associated with an innovation on one or more venues in the type of marketed options, in the incentive structure for giving and taking liquidity, and in technology-driven changes in trading costs, along with affiliated shifts in trading strategies. In the absence of major developments in these areas, the quote frequency for a given type of option is likely relatively stable.

Tables 7 and 8 provide the market shares of trading volume and quoting activity across the exchange platforms for the option contracts written on our sample of underlying securities. The most striking feature is the wide dispersion in trading across the venues, with the largest market share being the 22% obtained by MIAX for BAC options. Overall, the dominant venue is CBOE, but BATS and MIAX are not very far behind, followed by PHLX,

26 Notable events include the Greek default on June 30, 2015, and a market crash in China, where the Shanghai Composite shed 38% between June 12 and August 24.

Table 7. Shares of trading volumes per exchange for selected underlyings

Code	A	B	C	H	I	M	N	Q	T	W	X	Z
Exchange	AMEX	BOX	CBOE	GEMX	ISE	MIAX	ARCA	NOM	NOBO	C2	PHLX	BATS
PG	0.08	0.07	0.18	0.03	0.04	0.10	0.13	0.10	0.01	0.02	0.12	0.13
WMT	0.06	0.07	0.14	0.03	0.12	0.11	0.09	0.09	0.01	0.03	0.12	0.13
K	0.08	0.08	0.12	0.04	0.10	0.11	0.16	0.08	0.02	-	0.10	0.11
XLP	0.07	0.06	0.12	0.03	0.07	0.13	0.11	0.08	0.02	0.02	0.09	0.19
XOM	0.06	0.06	0.18	0.02	0.05	0.09	0.10	0.14	0.01	0.02	0.09	0.17
CVX	0.07	0.05	0.15	0.02	0.07	0.14	0.09	0.15	0.01	0.02	0.10	0.13
MPC	0.08	0.06	0.15	0.04	0.11	0.08	0.12	0.10	0.01	-	0.10	0.16
XLE	0.08	0.06	0.20	0.03	0.09	0.10	0.09	0.10	0.01	0.03	0.09	0.12
JPM	0.06	0.08	0.17	0.02	0.06	0.09	0.10	0.14	0.01	0.01	0.15	0.11
BA C	0.06	0.07	0.15	0.03	0.08	0.22	0.10	0.08	0.01	0.01	0.08	0.11
BLK	0.05	0.03	0.19	0.04	0.12	0.06	0.10	0.06	0.02	-	0.13	0.20
XLF	0.10	0.08	0.13	0.04	0.06	0.18	0.07	0.06	0.01	0.04	0.12	0.11
AAPL	0.07	0.07	0.18	0.03	0.08	0.13	0.09	0.11	0.01	0.02	0.08	0.15
GOOG	0.06	0.04	0.14	0.06	0.09	0.06	0.12	0.09	0.01	0.02	0.14	0.17
EA	0.06	0.05	0.12	0.02	0.06	0.05	0.12	0.15	0.01	0.04	0.19	0.14
XLK	0.07	0.06	0.12	0.03	0.07	0.16	0.08	0.07	0.02	0.02	0.13	0.16
SPY	0.07	0.07	0.18	0.03	0.15	0.16	0.07	0.07	0.01	0.03	0.07	0.10

Highlighting: $x \geq 0.15$, $0.15 > x \geq 0.1$, and $0.1 > x \geq 0.05$

Table 8. Shares of aggregate quotation activity per exchange for selected underlying securities

Code	A	B	C	H	I	M	N	Q	T	W	X	Z
Exchange	AMEX	BOX	CBOE	GEMX	ISE	MIAX	ARCA	NOM	NOBO	C2	PHLX	BATS
PG	0.16	0.04	0.15	0.09	0.12	0.14	0.07	0.03	0.03	0.03	0.10	0.04
WMT	0.15	0.04	0.13	0.06	0.13	0.14	0.05	0.04	0.05	0.05	0.13	0.04
K	0.16	0.03	0.16	0.08	0.13	0.11	0.05	0.04	0.05	-	0.13	0.05
XLP	0.13	0.03	0.14	0.05	0.19	0.14	0.05	0.04	0.05	0.05	0.10	0.03
XOM	0.18	0.03	0.15	0.06	0.13	0.14	0.05	0.04	0.05	0.03	0.11	0.04
CVX	0.14	0.04	0.15	0.04	0.12	0.15	0.06	0.05	0.05	0.04	0.13	0.04
MPC	0.25	0.02	0.17	0.07	0.10	0.12	0.04	0.03	0.03	-	0.15	0.03
XLE	0.15	0.03	0.14	0.07	0.15	0.13	0.05	0.04	0.05	0.05	0.10	0.03
JPM	0.14	0.03	0.14	0.07	0.15	0.14	0.05	0.04	0.04	0.04	0.12	0.04
BAC	0.10	0.03	0.13	0.11	0.13	0.12	0.05	0.05	0.07	0.03	0.13	0.04
BLK	0.14	0.04	0.15	0.06	0.09	0.15	0.06	0.05	0.07	-	0.15	0.04
XLF	0.12	0.03	0.13	0.10	0.15	0.14	0.04	0.04	0.05	0.06	0.10	0.03
AAPL	0.13	0.03	0.16	0.07	0.11	0.15	0.04	0.04	0.05	0.06	0.11	0.04
GOOG	0.15	0.04	0.11	0.08	0.09	0.13	0.08	0.06	0.07	0.03	0.12	0.06
EA	0.16	0.03	0.16	0.05	0.13	0.10	0.06	0.05	0.04	0.03	0.13	0.05
XLK	0.11	0.02	0.13	0.14	0.15	0.13	0.04	0.04	0.04	0.05	0.10	0.03
SPY	0.11	0.04	0.12	0.06	0.13	0.14	0.06	0.06	0.05	0.06	0.12	0.05

Highlighting: $x \geq 0.15$, $0.15 > x \geq 0.1$, and $0.1 > x \geq 0.05$.

NOM, ARCA, and ISE. In fact, all exchanges display non-trivial participation, except for the marginal activity on OBO, C2, and, to some extent, GEMX. The quoting activity on the individual venues is correlated with the trading volume, but AMEX, ISE, and PHLX are relatively more active in this dimension, while the quoting activity on BATS, NOM, and

Table 9. Heatmap on NBBO participation per exchange and underlying

Code	A	B	C	H	I	M	N	Q	T	W	X	Z
Exchange	AMEX	BOX	CBOE	GEMX	ISE	MIAX	ARCA	NOM	NOBO	C2	PHLX	BATS
PG	0.56	0.62	0.64	0.47	0.55	0.65	0.69	0.63	0.63	0.43	0.54	0.64
WMT	0.57	0.56	0.64	0.59	0.62	0.62	0.65	0.58	0.61	0.50	0.61	0.55
K	0.52	0.44	0.43	0.37	0.40	0.47	0.52	0.45	0.47	0.00	0.44	0.47
XLP	0.53	0.47	0.47	0.42	0.41	0.47	0.55	0.47	0.47	0.33	0.45	0.47
XOM	0.61	0.50	0.57	0.60	0.56	0.58	0.57	0.56	0.54	0.44	0.64	0.53
CVX	0.63	0.43	0.54	0.48	0.49	0.55	0.55	0.51	0.55	0.30	0.62	0.50
MPC	0.58	0.58	0.56	0.49	0.54	0.61	0.62	0.59	0.59	0.00	0.55	0.59
XLE	0.68	0.67	0.69	0.66	0.64	0.69	0.71	0.69	0.68	0.48	0.70	0.62
JPM	0.67	0.68	0.70	0.69	0.62	0.71	0.71	0.72	0.70	0.57	0.66	0.71
BA C	0.38	0.36	0.41	0.29	0.36	0.40	0.45	0.44	0.40	0.31	0.41	0.45
BLK	0.61	0.49	0.57	0.38	0.37	0.61	0.66	0.53	0.49	0.00	0.64	0.47
XLF	0.63	0.51	0.53	0.51	0.46	0.52	0.63	0.52	0.52	0.42	0.49	0.52
AAPL	0.76	0.58	0.74	0.62	0.69	0.72	0.73	0.71	0.72	0.66	0.69	0.70
GOOG	0.66	0.46	0.62	0.67	0.42	0.54	0.57	0.58	0.69	0.20	0.69	0.53
EA	0.51	0.53	0.54	0.38	0.51	0.51	0.60	0.50	0.53	0.45	0.47	0.52
XLK	0.57	0.51	0.54	0.51	0.53	0.54	0.57	0.51	0.55	0.46	0.53	0.52
SPY	0.47	0.36	0.50	0.40	0.45	0.44	0.47	0.42	0.41	0.39	0.49	0.41

Percentages of average time each exchange displays the two-sided NBBO prices, calculated from all option contracts within a given class. For multiple messages per minute, only the last quoted levels per exchange and asset were considered. Quotation before 8:30 a.m., after 3:00 p.m. and those having either zero bid or ask sizes were neglected. Highlights: green $\geq$ 0.6, 0.6 > light green $\geq$ 0.5 and yellow $\leq$ 0.4.

ARCA is very subdued compared to their share of trading volume. Clearly, there is a strong heterogeneity in the mode of operation across exchanges, altering the relative quoting versus trading frequency substantially in some cases. As an aside, we note that the venue-specific trade and quote frequencies are not reliable indicators of the put–call parity violations identified in Section 3.3.

The participation of the different markets in the NBBO is of substantial interest, as this figure may be suggestive of the relative price efficiency of the venues. For each exchange, underlying asset, and second-to-second stamp, we compare the BBO of a given venue against the NBBO. We only consider the core trading session, from 8:30 am until 3:00 pm CT, to avoid the results being trivially tilted in favor of BATS—the only exchange group offering a form of pre-trading session.[27]

Table 9 indicates that C2 and GEMX are the least competitive venues with respect to NBBO participation for most underlyings and, not surprisingly, these exchanges have less than half the quote updates of the market leaders. However, two other "small" exchanges, BATS and BOX, with even less quotation messages (Figure 8) exhibit significantly higher rates of NBBO participation. Nonetheless, the general impression from Table 9 is that the overall price quality for first-level quotations is remarkably similar across all 12 options exchanges. Undoubtedly, this is an important factor behind the highly dispersed option

27 For example, the value "0.50" in Table 9 for CBOE and underlying SPY means that CBOE market makers match or improve on the best bid-offer quotes for SPY options across all other venues for 3.25 h per day (i.e., 50% of the daily trading period) on average.

trading volumes, and it is consistent with the notion of a competitive national options market.

These tentative conclusions do not, however, clarify why we see large differences in the rate of apparent arbitrage violations across venues, as illustrated for BAC options quoted at CBOE and NYSE Arca in Section 3.3. From Tables 7 to 9, we find NYSE Arca to be an active venue for BAC option trading, but more marginal in its quoting intensity. Moreover, it is one of the most competitive venues in terms of matching the NBBO quotes for BAC options. Hence, while CBOE is more actively engaged in BAC option trading and quotation, NYSE Arca is, on average, providing equal or better top-level quotes. This suggests that, during turbulent periods, a slow quote update frequency may be an important contributor to the arbitrage opportunities identified in Section 3.3.[28]

4.4 Quoted Spreads

We explore the size of the spreads in option quotes across different markets, stocks, tenor, and moneyness. We focus on the relative spread, defined as

$$\mathrm{RS}_t = \frac{Q_t^{\mathrm{A}} - Q_t^{\mathrm{B}}}{MQ_t},$$

where MQ_t is a mid-quote price at time t. To keep the analysis manageable, we restrict our attention to option contracts traded from January 2 to February 18 with expiry February 20, 2015. Thus, the option tenor ranges from a couple of days to seven weeks. We remove entries belonging to the filtering groups F1, F2, F3, and F5, so our results reflect only regular quotes and mitigate the impact of outliers.

Numerous factors may help rationalize the size and variability of the option bid–ask spread. We convey our main findings through a few illustrative figures, while deferring extensive tabulations to Online Appendix A and additional illustrative evidence to Online Appendix D.

First, for equity indices, it is known that put options tend to be more liquid and have lower spreads than call options, but corresponding stylized facts for individual equity options are less well established. Figure 10 displays the relative quoted spread differential between calls and puts at the identical degree of moneyness. The figure renders the asymmetry in the option spread across the full set of underlying assets transparent. The discrepancy is particularly dramatic for OTM spreads, with only positive, and often very large, entries appearing for moneyness in excess of unity. In contrast, for ITM options, the spreads are quite closely aligned and, if anything, they are slightly smaller for call options, as indicated by the negative entries on the left-hand side of the figure. Consequently, the evidence on spreads is qualitatively quite similar for the individual equity and equity index options.

Next, Figure 11 displays the average intraday spreads for put options over three distinct tenor categories, covering the same period as above.[29] The left column depicts an "L-shaped" pattern in the spread over the trading day for three separate equity options.

28 A robust exploration of this conjecture requires an elaborate empirical analysis and is outside the scope of this paper.

29 The corresponding figure for call options conveys the identical qualitative results, see Online Appendix D.

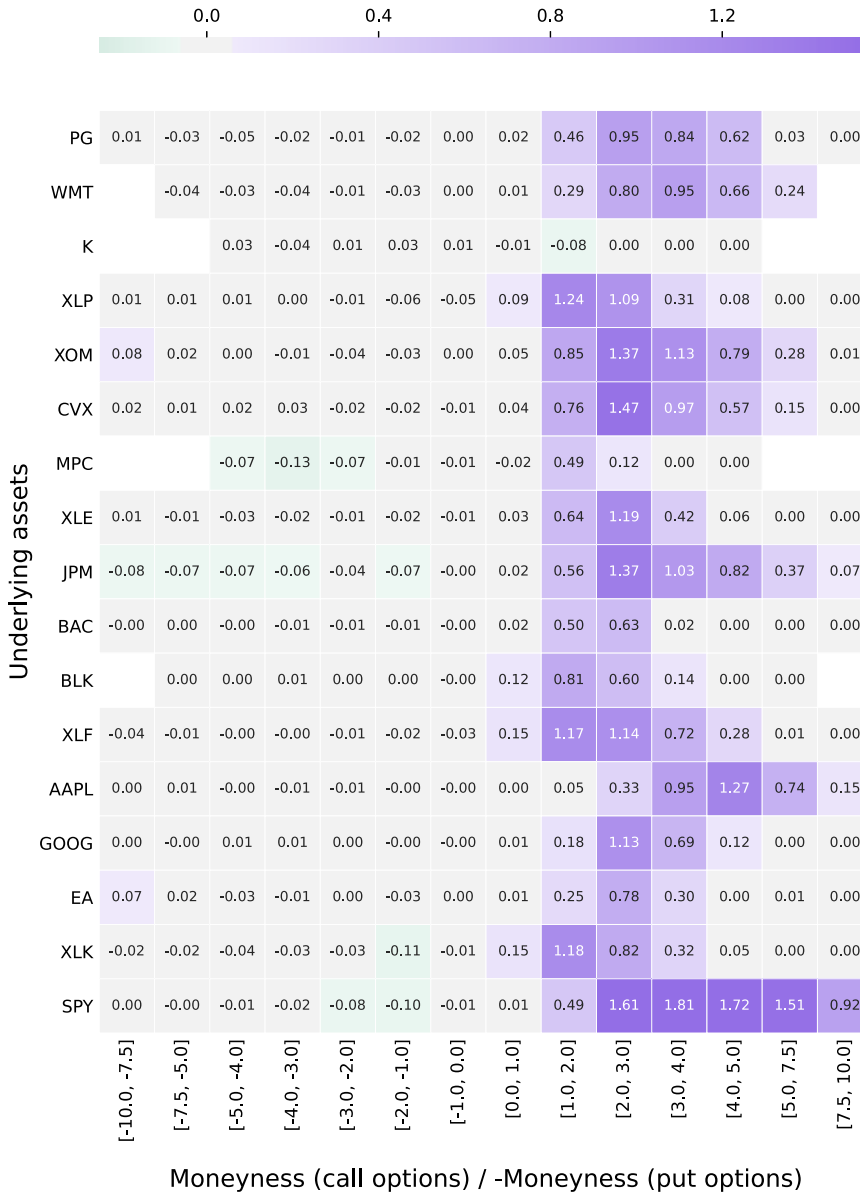

Figure 10. Difference in average relative spreads between call and put options across moneyness for each underlying asset in our sample. The results are based on OPRA quote records for put and call options between January 2 and February 18, 2015, that expire February 20, 2015. Moneyness is defined as $m = \log(K/F_t)/(\sigma_t\sqrt{\tau})$ with m multiplied by -1 for put options.

These are representative of the results for our individual stock options. In contrast, the option spreads for the two equity indices in the right column are approximately "U-shaped," while the Apple spread configuration in the middle panel looks like a mixture of the two

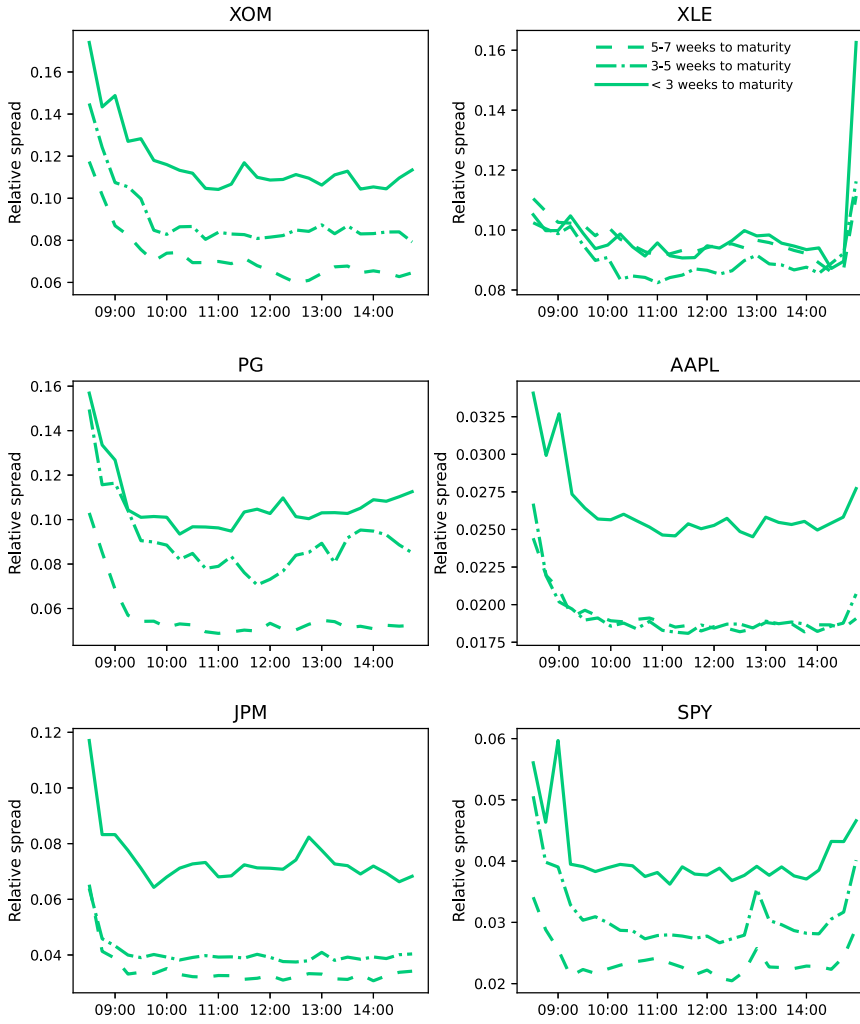

Figure 11. Relative spread measures (average over 15 min intra-daily intervals) computed from the OPRA quote records for put option contracts traded in January and February, 2015, which expire on February 20, 2015. The reported results are obtained using only those local 15-min intervals, where moneyness is in the range $-1 \leq m \leq 1$. The moneyness is defined as $m = \log(K/F_t)/(\sigma_t\sqrt{\tau})$.

patterns. Apple is also noteworthy by having the smallest relative option spread, by far, among all the underlying securities in our sample, followed by the SPY ETF. Interestingly, the energy ETF, XLE, has the largest uptick in relative spread toward the close of trading, while XOM, a large constituent of this energy ETF, shows no signs of an elevation at the end of the session. This divergence in trading costs mimics the discrepancies often observed in the volatility of the underlying equity indices and ETFs relative to individual stocks toward the market close. Finally, we note that the relative spread is almost monotonically

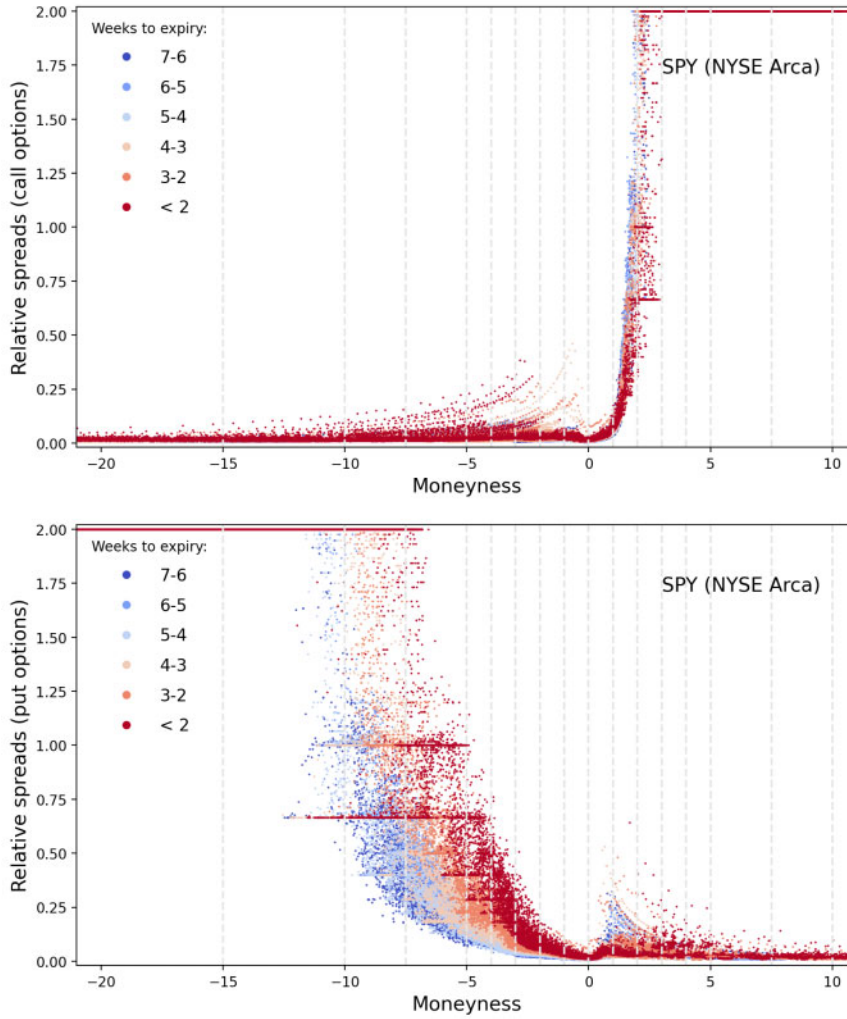

Figure 12. Relative spread measures (average within 15-min intraday intervals observed between 9:00 and 10:00 CT) of SPY options across maturity groups computed from the OPRA quote records for call (upper plot) and put (bottom plot) option contracts traded on NYSE Arca exchange between January 2 and February 18, 2015, which expire on February 20, 2015. The moneyness is defined as $m = \log(K/F_t)/(\sigma_t\sqrt{\tau})$.

increasing as time-to-maturity shrinks, which aligns well with the fact that option implied volatilities increase dramatically, as the tenor approaches zero.

The actual relationship between tenor and spreads is a bit more involved than conveyed above. Figure 12 displays the relative spreads for a set of SPY options observed on NYSE Arca exchange in January 2015.

Figure 12 exemplifies the strong asymmetry in spreads between puts and calls. In the top panel, OTM call options hit the maximum relative spread of 2 for moneyness 2–3. In

the bottom panel, a large fraction of the OTM put options feature much lower relative spreads for moneyness well beyond −5. It is also evident that certain spread values dominate for the short-maturity options, visible as straight horizontal red lines for both OTM put and calls. This stems from identical posted bid and ask quotes for a string of options across adjacent strikes, which is indicative of clustering at certain ticks.

Further instances of non-monotonicity in spreads as a function of tenor may be gleaned from Figure 12, where moderately ITM put options with long maturities attain the largest relative spreads (the blue dots lie above the red for low positive m values in the bottom panel). This feature is corroborated more broadly for our option sample in Tables A.6–A.7 (Online Appendix A), which capture the relation between moneyness and spreads along with Figures 7, C.3, and C.4 (Online Appendix C), that relate moneyness to trade and quote activity. We conjecture this stems from opposing effects of liquidity and tenor on the spread.

In summary, the relative option trading costs, as represented by the quoted bid–ask spread, are highly heterogeneous, varying substantially from one underlying to another, even within the same sector, and also showing non-trivial deviations across exchanges. A second factor is the option tenor, since relative spreads tend to increase, as the time value of the options shrinks in line with maturity. Nonetheless, the most important determinant is the identity of the underlying and the associated liquidity of the options market. Specifically, the actively traded options written on SPY and AAPL have dramatically narrower spreads than the less liquid options, irrespective of the exchange market and controlling for option tenor. Moreover, as short-dated options often are quite liquid, we do observe a non-monotonic tenor-spread or volume-spread relation across a subset of the maturity spectrum.

5 Empirical Applications

This section explores whether a couple of commonly used option-implied measures may be generated in a meaningful manner from high-frequency option prices exploiting standard techniques. If shortcomings become evident, it serves as motivation for future work on generating more robust intraday measurement procedures. It is beyond the scope of the present paper to pursue any such comprehensive remedies.

The first application involves second-by-second model-free estimation of the risk-neutral variance for an underlying security based on the intraday cross-section of option prices. The second focuses on estimation of the risk-neutral asset return distribution from high-frequency option prices.

5.1 Intraday Risk-Neutral Return Variation Measures

The expected future return volatility is a critical input to numerous financial and economic applications. As is well known, the cross-section of European-style option prices, covering the full range of strikes for a given expiration date, enables model-free computation of the expected return variation for the underlying security return under the risk-neutral probability measure, see, for example, Carr and Madan (1999) and Britten-Jones and Neuberger (2000). This model-free implied-volatility (MFIV) measure contains both a predictive and a risk premium component, as it reflects the expected future variance as well as the risk pricing for a security with a payoff equal to the future *realized* return variation.

In practice, the MFIV can only be approximated, as we do not observe "true spot" option prices, but rather bid–ask quotes. Likewise, the set of strikes is discrete, finite, and does not cover the entire positive real axis. As such, there are legitimate questions about the reliability of high-frequency MFIV measures.

The typical option data used in the academic literature consists of end-of-day cross-sections of prices or quotes (e.g., OptionMetrics data). This precludes investigation of the implied variance dynamics in real time, the reaction of volatility expectations to specific events within the trading day, the intraday co-dynamics of implied variances across multiple assets, and so on. Furthermore, the quality of a MFIV measure, based on a single cross-section of option prices, depends on the degree of noise or data error at the observation time. Intuitively, we obtain a more robust measure using multiple cross-sections from adjacent seconds, much in the spirit of pre-averaging for the measurement of the return variation from intraday log asset prices, see, for example, Jacod et al. (2009).

For the broader U.S. market, we already have a popular high-frequency return variation measure. The VIX index disseminated by the CBOE, capturing the expected risk-neutral volatility of the S&P 500 index over a 30-day period, is computed on a continuous basis throughout the trading day and released at a 15-s frequency, providing a real-time benchmark volatility indicator. Unfortunately, the intraday VIX series is not a reliable real-time MFIV measure. Andersen, Bondarenko, and Gonzalez-Perez (2015a) document significant spurious outliers in the high-frequency VIX index, due largely to the random tail truncation of the OTM option prices. In addition, it is subject to a non-trivial delay of 15–45 s stemming from random variation in the processing and dissemination speed. As a consequence, direct use of the high-frequency VIX series can result in severely distorted inference, especially during periods of market stress, when accurate real-time measures, arguably, are most needed. A number of studies recognize the potential distortion arising from the tail truncation inherent in the VIX computation and provide alternative recipes that amend the tails in different ways using the observed cross-section of option quotes. However, these procedures are almost invariably implemented at the daily frequency, and not on an intraday basis.

The highly accurate time-stamps in the OPRA data allow for the construction of MFIV series for a large number of stocks and ETFs traded on the U.S. equity market across a range of time horizons and at almost arbitrarily high frequencies. Of course, it remains an open question whether lack of liquidity or other distinct market microstructure features will render such measures excessively noisy or biased.

For our empirical illustration, following Carr and Wu (2009), we rely on a log-linear extrapolation of option prices in the strike domain for the tails of the return distribution rather than using tail truncation, as applied in the official VIX computation, or a robust corridor-based measure, as suggested in Andersen and Bondarenko (2007). Online Appendix E.1 provides a detailed description of our option portfolio design.

Our illustration focuses on MFIV's extracted at the 1-s frequency using intraday quotes for American-style SPY and GOOG options. The former provide an alternative way to construct a high-frequency S&P 500 volatility measure, while the GOOG options speak to the possibility of generating reliable intraday MFIV estimates for individual stocks. The choice of very liquid options provides us with a near best-case scenario to assess the properties of such measures at extremely high frequencies. We use options with only 3 days till expiration. This generates a return variation measure that is closely related to the spot volatility

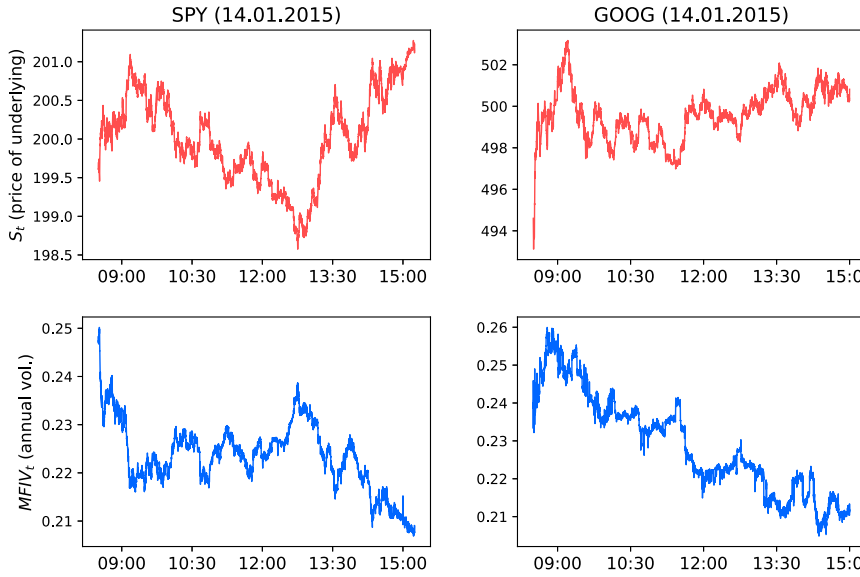

Figure 13. On the top plots, intraday prices of SPY (left side) and GOOG (right side) observed on January 14, 2015. On the bottom plots, intraday MFIV series constructed from SPY and GOOG options on January 14, 2015 (expiring January 17, 2015). Real-time MFIV is calculated on a second-by-second basis and normalized to annual volatility units.

of the underlying series, given the very short tenor.[30] Moreover, it ensures that the price differential between American- and European-style options truly is negligible.

Figure 13 plots the intraday prices for SPY and GOOG (top panel) and corresponding second-by-second annualized MFIV estimates (bottom panel). Several features stand out. First, the MFIV exhibits an appreciable amount of variation within the trading day, allowing for direct identification of shifts in the value across 10–20 min intervals. Nonetheless, the series are quite choppy, indicating a fair degree of measurement error, inducing negative serial correlation over short horizons. This suggests it may be worthwhile to construct more robust measures by smoothing suitably across adjacent second-by-second MFIV measures. Second, for both SPY and GOOG, the extracted MFIV is very volatile immediately after the market open. This is consistent with the elevated return volatility and high bid–ask spreads for the underlying securities in the morning. It likely reflects (option) price discovery, when overnight information and newly arriving orders are absorbed into the market. Finally, we note that the leverage effect, defined as a negative return-volatility relation, is evident, but clearly more pronounced for the SPY index than GOOG. The realized correlation between MFIV and the underlying price, computed from 2-min increments, equals −0.85 for SPY and −0.36 for GOOG. Existing studies of the high-frequency leverage effect from option data has focused exclusively on equity indices, see, for example,

30 There is a gap between spot (diffusive) volatility and the above option-based measure due to the risk-neutral jump variation, which can differ significantly from its statistical counterpart. This component of the option-based variation measure can be further removed in a nonparametric way using the approach of Todorov (2019).

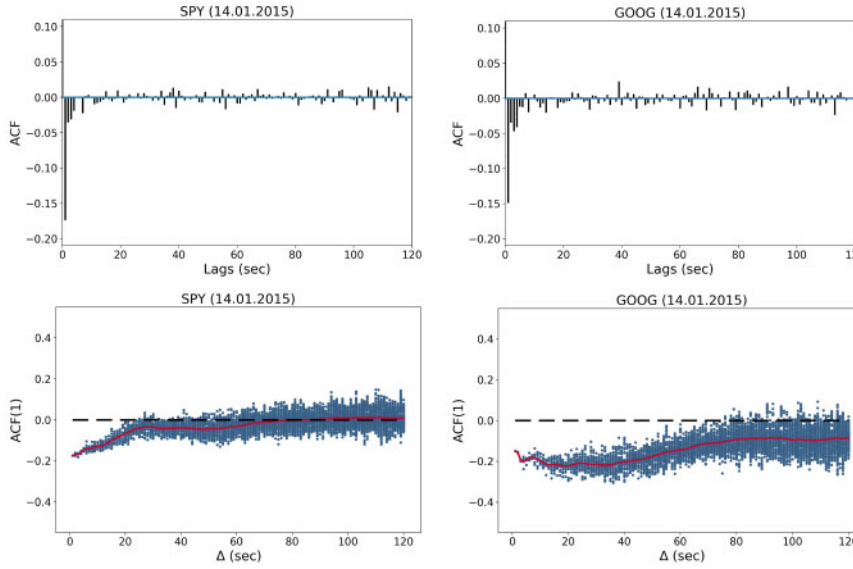

Figure 14. Autocorrelation functions for MFIV series constructed with SPY options (left side) and with GOOG options (right side) on January 14, 2015 (expired on January 17, 2015). Top plots show the autocorrelations as functions of lags constructed for the second-by-second increments of intraday MFIV (with the maximum lag of 120 s). Bottom plots picture the first-order serial correlations as functions of a sampling frequency (Δ) calculated for the increments of intraday MFIV obtained at the corresponding frequency. We consider Δ ranging from 1 to 120 s with a second step. Blue dots represent the first-order autocorrelations computed for a given Δ on multiple sampling "grids" achieved by shifting the initial MFIV observation by 1 s (thus, for $\Delta = 1$ s we have one "grid" and for $\Delta = 120$ s we have 120 "grids"). Solid red line is an average autocorrelation across all "grids" for a given Δ.

Andersen, Bondarenko, and Gonzalez-Perez (2015a) and Kalnina and Xiu (2017), but Figure 13 points to the feasibility of extending such studies to individual equity series.

To gauge the degree of noise in our intraday series, Figure 14 displays the serial correlation pattern for MFIV increments at different lags and sampling frequencies. The top panels present the empirical autocorrelations across lags ranging from 1 to 120 s. The bottom panels depict the first-order autocorrelation of MFIV increments for sampling frequencies $\Delta = 1, \ldots, 120$ seconds across different sampling grids, obtained by initiating the computation of the autocorrelations at each possible 1-s grid point $\Delta = 1, \ldots, 120$. The first observation of the initial grid is set to 8:35:00 CT and consecutive observations follow with a step of Δ seconds.[31] Hence, for each sampling frequency Δ, we have Δ serial correlation estimates (blue dots) generated from Δ distinct sampling grids. The red solid line corresponds to the average first-order autocorrelations calculated across all grids for each sampling frequency Δ.

The MFIV increments for both SPY and GOOG exhibit substantial negative autocorrelations for ultra-short lags (3–4 s) while, for higher lags, the autocorrelations change signs

31 We exclude MFIV estimates for the first 5 min of trading to avoid distortions related to market opening effects, including a lack of mid-quotes for some strikes and excessively volatile mid-quote revisions due to active price discovery.

randomly and become smaller in magnitude. However, the bottom plots of Figure 14 also suggest important discrepancies. The negative first-order serial dependence in the SPY MFIV increments vanishes, on average, for sampling frequencies of 70 s or lower, whereas, for GOOG, they remain sizable, even as Δ reaches 120 s. This is consistent with the SPY MFIV series being less impacted by measurement error, which may be due to the lower spreads and moderately richer set of observations available for the OTM SPY relative to GOOG options, with cross-sections of 182 and 155 strikes for SPY and GOOG, respectively.

The illustration highlights the feasibility of generating, but also the need for additional scrutiny of, high-frequency option-implied return variation measures. In Figures E.9–E.14 in Online Appendix E.2, we provide the corresponding January 14, 2015, MFIV measures for another six securities in our sample. The negative autocorrelations vanish fairly rapidly, within 50 s, for AAPL, but for the remaining ETFs and stocks, we observe non-zero correlations throughout the 120-s horizon. Our tentative conclusion is that MFIV measures obtained at the 1–2 min frequency using standard procedures may be reliable, when based on quotes from the most active option markets. In general, however, caution is warranted, and sampling at distinctly lower frequencies is advised for less liquid markets.[32]

5.2 High-Frequency Risk-Neutral Density Extraction

Another popular source of information extracted from the cross-section of option prices is the risk-neutral density (RND) of the underlying asset price over the period until option expiry. Since this measure seeks to provide a detailed picture of the risk-neutral distribution, while the corresponding MFIV only reflects the return variation, the RND is likely to be estimated with a greater degree of imprecision. However, they may still provide useful insights across lower intraday sampling intervals, and they may be particularly informative regarding the response to specific events across the cross-section of underlying assets.

In this section, we examine the intraday changes in the RND of both the market index ETF and several individual stocks before and after a Federal Open Market Committee (FOMC) meeting. The FOMC meetings are closely related to the equity risk premium and can have a significant impact on the prices of both stocks [e.g., Bernanke and Kuttner (2005), Savor and Wilson (2013), Lucca and Moench (2015) among others] and the underlying options (Andersen, Fusari, and Todorov, 2017). By exploiting high-frequency options data, we document immediate market-wide and idiosyncratic RND responses to scheduled macroeconomic news announcements, which is otherwise not feasible using only end-of-day options data.

Firstly, we introduce some notations. The price of a call option written on an underlying asset at time t with a strike price K and expiration date T is given by

$$C(K) = e^{-r_f\tau}\mathbb{E}^{\mathbb{Q}}\Big(\max(S_T - K, 0)\Big) = e^{-r_f\tau}\int_{S_T=K}^{\infty}(S_T - K)f_{\mathbb{Q}}(S_T)dS_T,$$

32 We leave for future work the formal analysis of microstructure noise in option panels observed at high frequency.

where S_T is the price of the underlying asset at expiry, $\tau = T - t$ is the time to maturity of the call option, r_f is the risk-free interest rate, and $f_{\mathbb{Q}}(S)$ denotes the RND of the asset price.

The object of interest is the RND $f_{\mathbb{Q}}(x)$. We adopt the well-established mixture of lognormals (MLN) approach, advocated by Ritchey (1990), Melick and Thomas (1997), and Liu et al. (2007). It implies that the call option price equals the weighted average of the option prices implied by the individual lognormal densities, which may be calculated in closed form via the Black (1976) formula. Improved estimates are obtained, asymptotically, by employing a larger number of distinct lognormal densities. A detailed description of the MLN methodology is provided in Online Appendix F.1.

For all securities with sufficiently liquid option markets, we estimate the RND before and after the FOMC announcement, released at 13:00 CT on March 18, 2015. To this end, for each underlying, we construct cross-sections containing the last observed option quote for each strike 30 min before and after the announcement. We use the shortest dated options with tenor exceeding three calendar days, leading to cross-sections with 9 days to expiry for this trading day.[33]

We follow standard procedures (e.g., Andersen et al., 2015a; Song and Xiu, 2016) to remove potentially erroneous quote entries.[34] Furthermore, to avoid excessively noisy estimates, we proceed only if the option cross-section has at least seven strikes after the data cleaning.[35] For brevity, we only discuss results for the most liquid stock and ETF in each sector, relegating the remaining results to Online Appendix F.3.[36]

Figure 15 depicts the price series for seven underlyings over the 1-h window straddling the FOMC announcement. At 13:00 CT, it was announced that the federal funds rate would remain at the 0% to 1/4% target range and the stance of monetary policy would be reassessed only after further market indicators became available. This decision triggered an immediate price jump for all seven underlyings, ultimately generating 1-h returns ranging from 1.20% to 1.51%.

Figure 16 depicts the implied RND for the return, $r_T = S_T/F_t - 1$, obtained from a mixture of $M = 3$ lognormal densities, using SPY options with 9 days to maturity at 12:30 and 13:30 CT on March 18, 2015.[37] Both RND curves exhibit pronounced negative skewness, consistent with the extant literature. This notwithstanding, the pertinent question is whether the RND changes shape in response to the FOMC announcement. In fact, from the log-density plot, we do identify a thinning of the left SPY tail and we find both the ATM Black–Scholes implied volatility and the implied risk-neutral return dispersion, $\hat{\sigma}_{\mathrm{R}}$, to have

33 We exclude K (Kellogg Company) and BLK (BlackRock), because their shortest-dated options have a 30-day tenor—much longer than the available tenor for the remaining underlyings.

34 A detailed description of our data preparation steps and estimation procedure are available in Online Appendix F.2.

35 This leads to the removal of BAC (Bank of America) and XLF (Financial Sector ETF) from this analysis.

36 We only report results for three sectors, Consumer Staples, Energy, and Technology, as all financial sector assets, except for JPM (JPMorgan & Chase), are excluded by our liquidity filter.

37 These RNDs are readily derived from $f_{\mathbb{Q}}(S_T)$ using a Jacobian transformation.

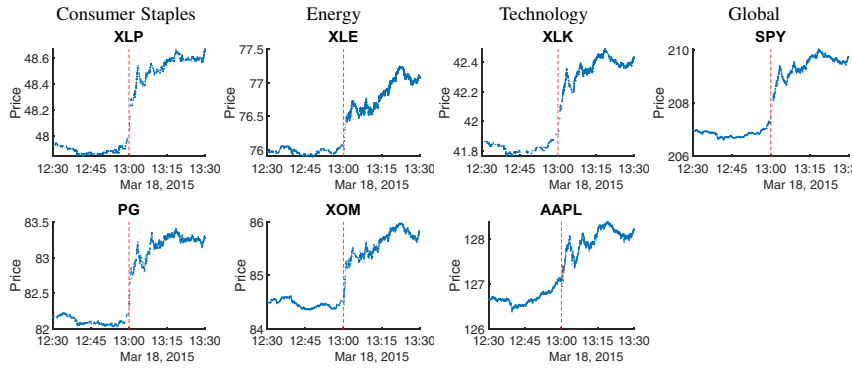

Figure 15. Time series plots of prices of different underlyings during 30 min before and after the FOMC announcement at 14:00 Eastern time (i.e., 13:00 CT) on March 18, 2015. Time in each plot is CT.

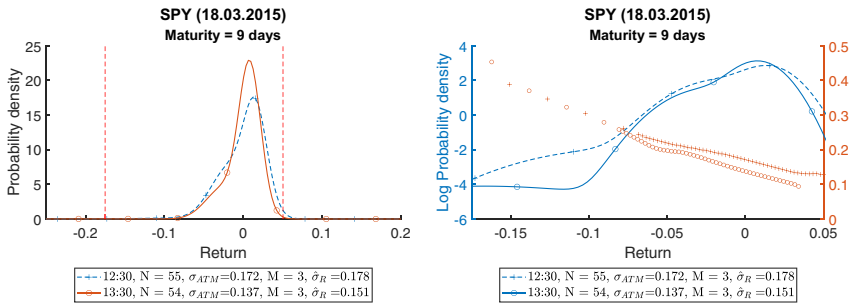

Figure 16. The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on SPY in the Global sector at 30 min before and after the FOMC announcement at 14:00 Eastern time (i.e., 13:00 CT) on March 18, 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black–Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black–Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT.

dropped.[38] This suggests that market participants, in response to the policy statement, have lowered their expectations regarding near-term unfavorable events impacting the SPY.

Of course, the statistical significance of the above shift in the RND is hard to assess. The ability to estimate concurrent shifts in the RND curves for a number of diverse assets is one way to gauge the robustness of this finding.

38 We define the dispersion of the RND curve (σ_{R}) as the annualized risk-neutral standard deviation of return $\frac{1}{\sqrt{\tau}}\sigma_{r_T}$, which equals $\frac{1}{\sqrt{\tau}F_t}\sigma_{S_T} = \frac{1}{\sqrt{\tau}F_t}\sqrt{[\sum_{i=1}^{M} w_i F_i^2 \exp(\sigma_i^2\tau)] - F_t^2}$, as implied by Equations (F.1) and (F.2) in Online Appendix F.1.

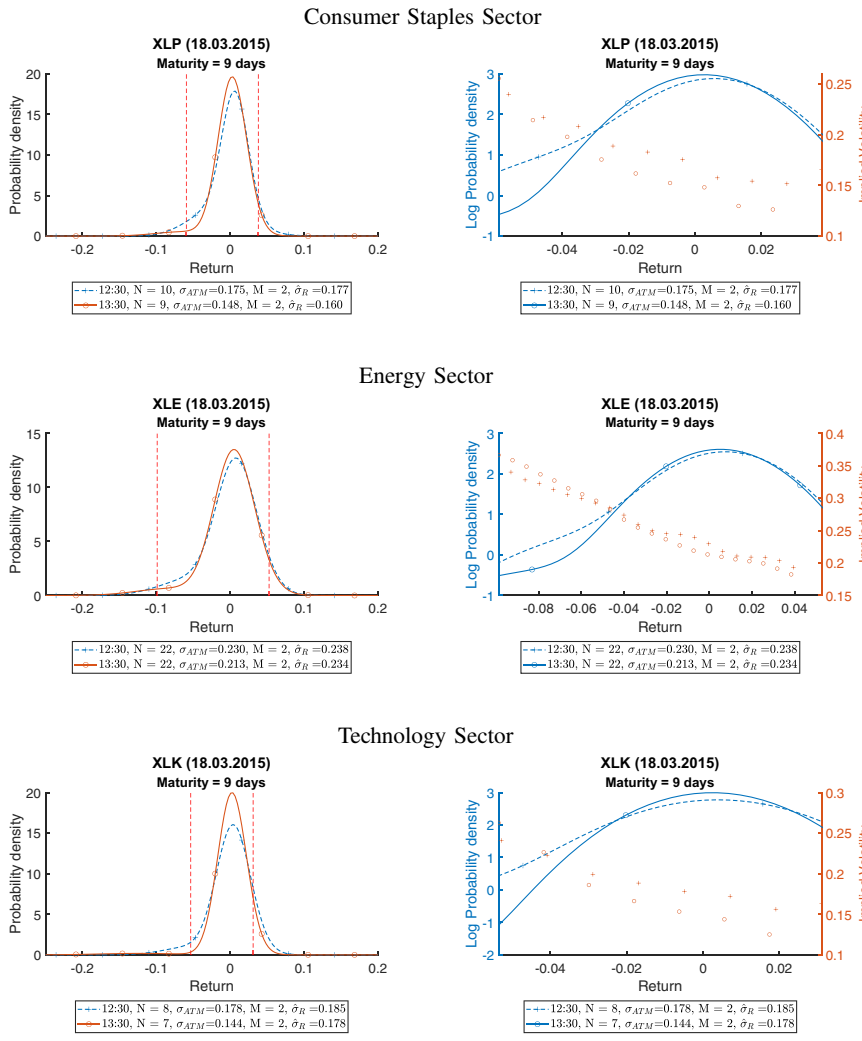

Figure 17. The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on ETFs of different sectors at 30 min before and after the FOMC announcement at 14:00 Eastern time (i.e., 13:00 CT) on March 18, 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black–Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_R$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black–Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT.

Figures 17 and 18 reveal how the RND for three ETFs and the most liquid stock within each of these sectors react to the announcement. In analogy to the SPY, the estimated RND (left plots) for all these securities become less dispersed (have lower standard deviation $\hat{\sigma}_{\text{R}}$) and experience a downward shift in ATM Black–Scholes implied volatilities after the

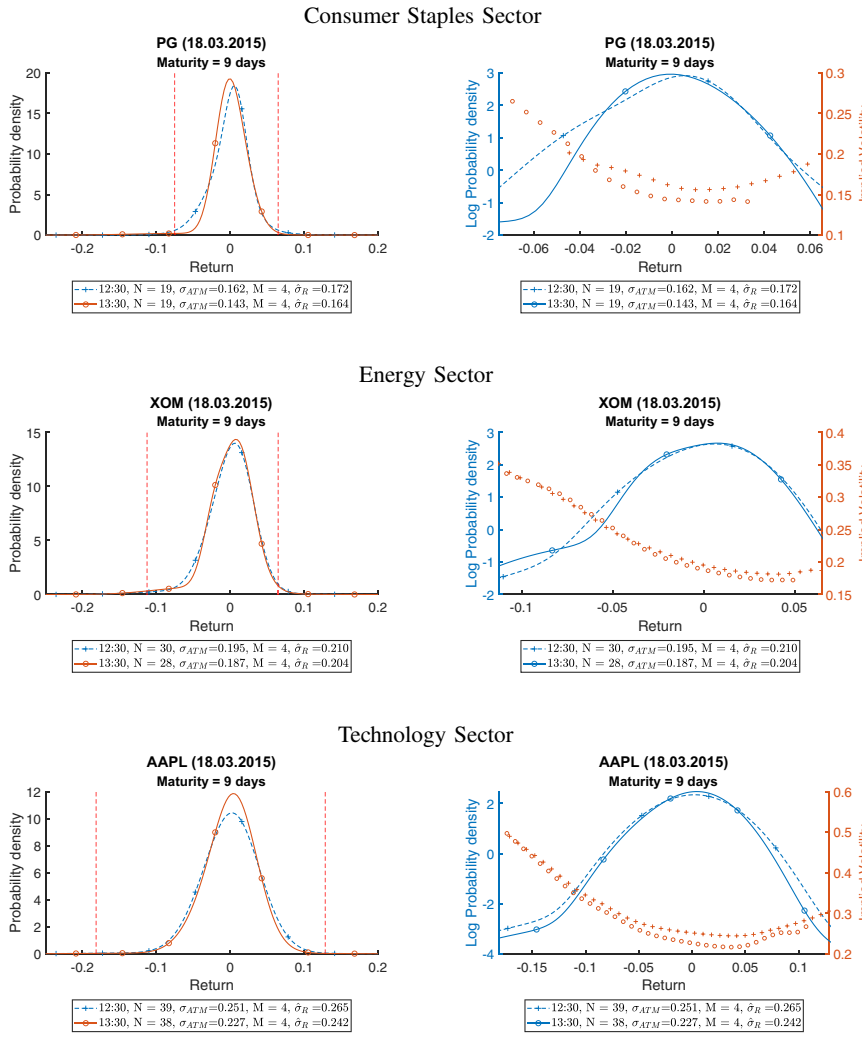

Figure 18. The left plots depict the estimated RNDs, as a function of return $S_T/F_t - 1$, for the shortest time to maturity obtained from intraday OPRA data for options written on individual stocks in different sectors at 30 min before and after the FOMC announcement at 14:00 Eastern time (i.e., 13:00 CT) on 18 March 2015. Each RND curve is estimated from a mixture of M lognormal distributions. In each subplot, the number of strikes (N), the ATM Black–Scholes implied volatility (σ_{ATM}), the number of lognormal densities in each mixture (M), and the estimated annualized standard deviation ($\hat{\sigma}_{\text{R}}$) of each RND curve are reported. Vertical red dashed lines indicate the observed return ($K/F_t - 1$) range. Right plots show the logarithm of the RNDs (left axis) and the Black–Scholes implied volatility (right axis) over the observed return range. Time in each plot is CT.

FOMC statement. This is consistent with the policy decision reducing the uncertainty about the future price and volatility of not only the equity index, but also each of the sector ETFs and individual stocks. In addition, consistent with SPY, the left tail of the estimated RNDs

of the sector ETFs shrinks after the FOMC announcement. This is generally also true for the individual stocks, but the estimation uncertainty concerning the tail shape is occasionally quite severe due to the limited number of OTM strikes, so the evidence is at times ambiguous.

Figures F.15 and F.16 in Online Appendix F.3 provide the corresponding evidence for the remaining securities in our sample. The findings are qualitatively similar, although we once again encounter a few cases where the shift in the RND is ambiguous. Nonetheless, across all the underlying assets explored, we find the left RND tail to either shrink or remain largely unaltered, while the measures of dispersion and volatility almost uniformly drop. In other words, the high-frequency options data indicate a reduction in the risk-neutral risk measures both for the aggregate equity index and across the various sectors and individual stocks. For other types of economic news, the reaction across different sectors and individual stocks is likely to be more heterogenous, which should facilitate interpretation regarding the shift in the perception of underlying risks and associated risk premiums. Since many other factors impact the RND across the full trading day, inference based solely on end-of-day option implied measures is much less efficient.

We conclude that the construction of meaningful intraday RND estimates is feasible for the assets with liquid option trading, as long as the sampling frequency is moderate. At the same time, it is evident that further research into the construction of robust and reliable extraction procedures is warranted for less liquid option markets. These issues notwithstanding, it is clear that the increasing availability and richness of high-frequency, short-dated option quotes enhances our ability to explore the economic and financial effects of various types of news events through their differential impact on a large option cross-section.

6 Conclusion

This study provides a detailed description of high-frequency trade and quote data for options traded in the United States. It reviews the current structure of the U.S. market by characterizing the 16 constituent option exchanges, summarizing the market regulatory plans governing the option trading, and discussing specific market maker quoting obligations that are pertinent to the functioning of the markets.

Our data are provided by OPRA in accordance with the "Plan for Reporting of Consolidated Options Last Sale Reports and Quotation Information." It contains more than 150 million trade and 1.2 trillion quote records at a millisecond resolution for all option classes written on individual equities, stock indices, and ETPs traded in the U.S. during the first eight months of 2015. Our dataset is more comprehensive than the alternative high-frequency option datasets employed in the limited number of existing studies, which, typically focus strictly on index options. We provide a detailed assessment of the quality of our dataset, and develop a general filtering algorithm for data cleaning the spirit of the Barndorff-Nielsen et al. (2009) algorithm for tick-by-tick data on equities.

Based on a representative sample in January 2015, we find a very small fraction of erroneous and irregular records, suggesting that the OPRA records are of high quality. An analysis of various liquidity measures confirms our expectation that options written on more liquid underlyings generally have tighter spreads. In addition, a cross-exchange investigation suggests that in 2015, CBOE, AMEX, and ARCA were the more competitive exchanges, participating most frequently in the NBBO quoting pair.

We characterize the trade and quote intensities across the securities and exchange venues as a function of option tenor and moneyness. Likewise, we explore the association between the size of the quoted option bid–ask spreads and the moneyness, tenor, and type of option (call or put).

Finally, we present two illustrative applications—the estimation of the risk-neutral return variance and the RND—using intraday OPRA data. We confirm that such measures can be constructed and yield information beyond what can be gleaned from end-of-day option data, but we also point toward limitations associated with the measures at the very highest frequencies.

The various findings serve as inspiration for new studies addressing either market microstructure or asset pricing questions related to the option markets. The literature on high-frequency equity data is voluminous. We expect the much richer, but also less manageable, option data to provide a stimulating basis for novel work in the future, both through exploration within the option space itself, but also in terms of the interaction between the option and equity markets at a fine resolution. In order for this to materialize, however, the rather complex market structure and regulatory environment should be recognized, so that appropriate hypotheses and empirical procedures can be designed.

Overall, we hope this overview will serve as inspiration to explore the OPRA dataset to its fullest. We are convinced it holds the key to progress on many market microstructure and high-frequency asset pricing questions.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

References

Amaya, D., J.-F. Bégin, and G. Gauthier. 2018. Extracting Latent States from High Frequency Option Prices. Available at SSRN https://ssrn.com/abstract=2975355 (accessed 23 October 2020).

Anand, A., and S. Chakravarty. 2007. Stealth Trading in Options Markets. *Journal of Financial and Quantitative Analysis* 42: 167–187.

Anand, A., J. Hua, and T. McCormick. 2016. Make-Take Structure and Market Quality: Evidence from the U.S. Options Markets. *Management Science* 62: 3271–3290.

Andersen, T. G., and O. Bondarenko. 2007. "Construction and Interpretation of Model-Free Implied Volatility." NBER Working Papers 13449, National Bureau of Economic Research, Inc.

Andersen, T. G., O. Bondarenko, and M. T. Gonzalez-Perez. 2015a. Exploring Return Dynamics via Corridor Implied Volatility. *Review of Financial Studies* 28: 2902–2945.

Andersen, T. G., O. Bondarenko, V. Todorov, and G. Tauchen. 2015b. The Fine Structure of Equity-Index Option Dynamics. *Journal of Econometrics* 187: 532–546.

Andersen, T. G., N. Fusari, and V. Todorov. 2017. Short-Term Market Risks Implied by Weekly Options. *The Journal of Finance* 72: 1335–1386.

Audrino, F., and M. R. Fengler. 2015. Are Classical Option Pricing Models Consistent with Observed Option Second-Order Moments? Evidence from High-Frequency Data. *Journal of Banking & Finance* 61: 46–63.

Bakshi, G., C. Cao, and Z. Chen. 1997. Empirical Performance of Alternative Option Pricing Models. *The Journal of Finance* 52: 2003–2049.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard. 2009. Realized Kernels in Practice: Trades and Quotes. *The Econometrics Journal* 12: C1–C32.

Battalio, R., S. Figlewski, and R. Neal. 2020. Option Investor Rationality Revisited: The Role of Exercise Boundary Violations. *Financial Analysts Journal* 76: 82–99.

Battalio, R., B. Hatch, and R. Jennings. 2004. Toward a National Market System for U.S. Exchange-Listed Equity Options. *The Journal of Finance* 59: 933–962.

Battalio, R., and P. Schultz. 2006. Options and the Bubble. *The Journal of Finance* 61: 2071–2102.

Battalio, R., and P. Schultz. 2011. Regulatory Uncertainty and Market Liquidity: The 2008 Short Sale Ban's Impact on Equity Option Markets. *The Journal of Finance* 66: 2013–2053.

Battalio, R., A. Shkilko, and R. Van Ness. 2016. To Pay or Be Paid? The Impact of Taker Fees and Order Flow Inducements on Trading Costs in U.S. Options Markets. *Journal of Financial and Quantitative Analysis* 51: 1637–1662.

Bernanke, B. S., and K. N. Kuttner. 2005. What Explains the Stock Market's Reaction to Federal Reserve Policy?. *The Journal of Finance* 60: 1221–1257.

Bhattacharya, M. 1987. Price Changes of Related Securities: The Case of Call Options and Stocks. *The Journal of Financial and Quantitative Analysis* 22: 1–15.

Birru, J., and S. Figlewski. 2012. Anatomy of a Meltdown: The Risk Neutral Density for the S&P 500 in the Fall of 2008. *Journal of Financial Markets* 15: 151–180.

Black, F. 1976. The Pricing of Commodity Contracts. *Journal of Financial Economics* 3: 167–179.

Bollen, N. P., and R. E. Whaley. 2004. Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? *The Journal of Finance* 59: 711–753.

Breeden, D. T., and R. H. Litzenberger. 1978. Prices of State-Contingent Claims Implicit in Option Prices. *The Journal of Business* 51: 621–651.

Britten-Jones, M., and A. Neuberger. 2000. Option Prices, Implied Price Processes, and Stochastic Volatility. *The Journal of Finance* 55: 839–866.

Brogaard, J., B. Hagströmer, L. Norden, and R. Riordan. 2015. Trading Fast and Slow: Colocation and Liquidity. *Review of Financial Studies* 28: 3407–3443.

Brownlees, C., and G. Gallo. 2006. Financial Econometric Analysis at Ultra-High Frequency: Data Handling Concerns. *Computational Statistics & Data Analysis* 51: 2232–2245.

Cakici, N., G. Goswami, and S. Tan. 2014. Options Resilience during Extreme Volatility: Evidence from the Market Events of May 2010. *Journal of Banking & Finance* 49: 262–274.

Cao, C., Z. Chen, and J. M. Griffin. 2005. Informational Content of Option Volume Prior to Takeovers. *The Journal of Business* 78: 1073–1109.

Carr, P., and D. B. Madan. 1999. Option Valuation Using the Fast Fourier Transform. *The Journal of Computational Finance* 2: 61–73.

Carr, P., and L. Wu. 2009. Variance Risk Premiums. *Review of Financial Studies* 2: 1311–1341.

Chakravarty, S., H. Gulen, and S. Mayhew. 2004. Informed Trading in Stock and Option Markets. *The Journal of Finance* 59: 1235–1257.

Chan, K., Y. P. Chung, and W.-M. Fong. 2002. The Informational Role of Stock and Option Volume. *Review of Financial Studies* 15: 1049–1075.

Chan, K., Y. P. Chung, and H. Johnson. 1993. Why Option Prices Lag Stock Prices: A Trading-Based Explanation. *The Journal of Finance* 48: 1957–1967.

Chan, K., Y. P. Chung, and H. Johnson. 1995. The Intraday Behavior of Bid–Ask Spreads for NYSE Stocks and CBOE Options. *The Journal of Financial and Quantitative Analysis* 30: 329–346.

Chang, B.-Y., P. Christoffersen, K. Jacobs, and G. Vainberg. 2012. Option-Implied Measures of Equity Risk. *Review of Finance* 16: 385–428.

Chatrath, A., R. A. Christie-David, H. Miao, and S. Ramchander. 2015. Short-Term Options: Clienteles, Market Segmentation, and Event Trading. *Journal of Banking & Finance* 61: 237–250.

Christoffersen, P., R. Goyenko, K. Jacobs, and M. Karoui. 2018. Illiquidity Premia in the Equity Options Market. *The Review of Financial Studies* 31: 811–851.

Christoffersen, P., S. Heston, and K. Jacobs. 2009. The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well. *Management Science* 55: 1914–1932.

Christoffersen, P., K. Jacobs, and K. Mimouni. 2010. Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns, and Option Prices. *Review of Financial Studies* 23: 3141–3189.

Christoffersen, P., K. Jacobs, C. Ornthanalai, and Y. Wang. 2008. Option Valuation with Long-Run and Short-Run Volatility Components. *Journal of Financial Economics* 90: 272–297.

Cremers, K., and D. Weinbaum. 2010. Deviations from Put–Call Parity and Return Predictability. *Journal of Financial and Quantitative Analysis* 45: 335–367.

Dalderop, J. 2020. Nonparametric Filtering of Conditional State-Price Densities. *Journal of Econometrics* 214: 295–325.

De Fontnouvelle, P., R. P. H. Fishe, and J. H. Harris. 2003. The Behavior of Bid–Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999. *The Journal of Finance* 58: 2437–2463.

Easley, D., M. O'Hara, and P. S. Srinivas. 1998. Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *The Journal of Finance* 53: 431–465.

George, T. J., and F. A. Longstaff. 1993. Bid–Ask Spreads and Trading Activity in the S & P 100 Index Options Market. *The Journal of Financial and Quantitative Analysis* 28: 381–397.

Hagströmer, B., and L. Nordén. 2013. The Diversity of High-Frequency Traders. *Journal of Financial Markets* 16: 741–770.

Harris, L. E., and S. Mayhew. 2005. "Market Fragmentation Metrics." Working Paper, University of Southern California.

Holowczak, R., J. Hu, and L. Wu. 2014. Aggregating Information in Option Transactions. *The Journal of Derivatives* 21: 9–23.

Holowczak, R., Y. E. Simaan, and L. Wu. 2006. Price Discovery in the U.S. Stock and Stock Options Markets: A Portfolio Approach. *Review of Derivatives Research* 9: 37–65.

Hu, J. 2014. Does Option Trading Convey Stock Price Information? *Journal of Financial Economics* 111: 625–645.

Jacod, J., Y. Li, P. Mykland, M. Podolskij, and M. Vetter. 2009. Microstructure Noise in the Continuous Case: The Pre-averaging Approach. *Stochastic Processes and Their Applications* 119: 2249–2276.

Kalnina, I., and D. Xiu. 2017. Nonparametric Estimation of the Leverage Effect: A Trade-off between Robustness and Efficiency. *Journal of the American Statistical Association* 112: 384–396.

Kapetanios, G., E. Konstantinidi, M. Neumann, and G. Skiadopoulos. 2019. Jumps in Option Prices and Their Determinants: Real-Time Evidence from the E-Mini S&P 500 Options Market. *Journal of Financial Markets* 46: 100506.

Lee, J., and C. H. Yi. 2001. Trade Size and Information-Motivated Trading in the Options and Stock Markets. *The Journal of Financial and Quantitative Analysis* 36: 485–501.

Liu, X., M. B. Shackleton, S. J. Taylor, and X. Xu. 2007. Closed-Form Transformations from Risk-Neutral to Real-World Distributions. *Journal of Banking & Finance* 31: 1501–1520.

Lucca, D. O., and E. Moench. 2015. The Pre-FOMC Announcement Drift. *The Journal of Finance* 70: 329–371.

Mayhew, S. 2002. Competition, Market Structure, and Bid–Ask Spreads in Stock Option Markets. *The Journal of Finance* 57: 931–958.

Mayhew, S., A. Sarin, and K. Shastri. 1995. The Allocation of Informed Trading across Related Markets: An Analysis of the Impact of Changes in Equity-Option Margin Requirements. *The Journal of Finance* 50: 1635–1653.

Melick, W. R., and C. P. Thomas. 1997. Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis. *The Journal of Financial and Quantitative Analysis* 32: 91–115.

Mishra, S., and R. T. Daigler. 2014. Intraday Trading and Bid–Ask Spread Characteristics for SPX and SPY Options. *The Journal of Derivatives* 21: 70–84.

Muravyev, D. 2016. Order Flow and Expected Option Returns. *The Journal of Finance* 71: 673–708.

Muravyev, D., and N. D. Pearson. 2020. Options Trading Costs Are Lower than You Think. *The Review of Financial Studies* 16: 527–566.

Muravyev, D., N. D. Pearson, and J. Paul Broussard. 2013. Is There Price Discovery in Equity Options? *Journal of Financial Economics* 107: 259–283.

Nagel, S. 2012. Evaporating Liquidity. *Review of Financial Studies* 25: 2005–2039.

Pan, J. 2002. The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics* 63: 3–50.

Ritchey, R. J. 1990. Call Option Valuation for Discrete Normal Mixtures. *Journal of Financial Research* 13: 285–296.

Savor, P., and M. Wilson. 2013. How Much Do Investors Care about Macroeconomic Risk? Evidence from Scheduled Economic Announcements. *Journal of Financial and Quantitative Analysis* 48: 343–375.

Sheikh, A. M., and E. I. Ronn. 1994. A Characterization of the Daily and Intraday Behavior of Returns on Options. *The Journal of Finance* 49: 557–579.

Simaan, Y. E., and L. Wu. 2007. Price Discovery in the U.S. Stock Options Market. *The Journal of Derivatives* 15: 20–38.

Simon, D. P. 2013. The Intraday and Overnight Behavior of SPY Options and Adjusted Delta Hedging. *Journal of Futures Markets* 33: 443–468.

Song, Z., and D. Xiu. 2016. A Tale of Two Option Markets: Pricing Kernels and Volatility Risk. *Journal of Econometrics* 190: 176–196.

Stephan, J. A., and R. E. Whaley. 1990. Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets. *The Journal of Finance* 45: 191–220.

Taylor, S. J., C.-F. Tzeng, and M. Widdicks. 2018. Information about Price and Volatility Jumps Inferred from Options Prices. *Journal of Futures Markets* 38: 1206–1226.

Todorov, V. 2019. Nonparametric Spot Volatility from Options. *The Annals of Applied Probability* 29: 3590–3636.

Vijh, A. M. 1990. Liquidity of the CBOE Equity Options. *Journal of Finance* 45: 1157–1179.

Zhang, Y. 2018. Essays in Risk Management and Asset Pricing with High Frequency Option Panels. PhD thesis, Durham University.

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 178–201
doi: 10.1093/jjfinec/nbaa031
Advance Access Publication Date: 2 December 2020
Article

Dynamics of Equity Factor Returns and Asset Pricing

Stoyan V. Stoyanov[1] and Francesco A. Fabozzi[2]

[1]Charles Schwab and [2]Stevens Institute of Technology

Address correspondence to Stoyan V. Stoyanov, Charles Schwab, 150 S. Wacker Dr. Ste 4, Chicago, IL 60606, USA, or e-mail: stoyan.stoyanov@schwab.com

Received May 19, 2019; revised August 11, 2020; editorial decision August 17, 2020; accepted August 17, 2020

Abstract

In empirical equity asset pricing, the stochastic discount factor (SDF) is implicitly modeled as a linear function of equity factors and is influenced by the empirical properties of the factor returns. We investigate the pricing error introduced by a misspecified SDF which ignores each of the following established empirical phenomena: autocorrelation, dynamics of covariances, dynamics of correlations, and heavy tails for the conditional factor return distribution. We consider near-linear SDFs and nonlinear specifications characterized by a high degree of risk aversion. We find that assuming constant covariances or constant correlations can significantly overprice certain equity portfolios at all risk-aversion levels and that ignoring fat tails can lead to large pricing errors for some derivative assets for highly nonlinear SDFs.

Key words: equity factors, volatility clustering, correlation dynamics, tail thickness, asset pricing

JEL classification: C32, G12

The most common model in empirical asset pricing is the linear factor model which identifies a linear expected return–beta relationship between the excess returns of risky assets and the expected returns of the primitive assets, also known as factors. The classical formulation of an asset pricing model, however, expresses current prices as an expectation of the discounted future payoffs in which the discount factor is a random variable that summarizes investor preferences over different states of the world. It can be shown that the linear expected return–beta model is equivalent to a stochastic discount factor (SDF) which is a linear function of the primitive assets, see, for example, Cochrane (2005).

In this article, we investigate the relative importance of certain empirical properties of established equity factors on asset prices. Because the SDF is implicitly or explicitly modeled as a function of the primitive assets, a misspecified econometric model introduces a pricing error. We consider the market, size, value, investment, and profitability factors of the five-factor Fama–French model developed by Fama and French (2005) extended with

the momentum factor as proposed by Carhart (1997). Using an autoregressive (AR) dynamic conditional correlation (DCC) GARCH model, we estimate the pricing error arising from an SDF which is misspecified by ignoring each of the following empirical phenomena captured by the econometric model: autocorrelation, dynamics of covariances, dynamics of correlations, and the fat tail of the conditional distribution of factor returns.

There are related studies in the empirical literature. Christoffersen and Langlois (2013) investigate the effect of tail behavior, the dynamics of volatility and correlations, and tail dependence of the market, size, value, and momentum factors on portfolio construction. Employing a DCC GARCH model with various parametric models including the Gaussian, Student's t, and skewed Student's t copulas, they find that volatility and correlation dynamics in factor returns are persistent. Christoffersen and Langlois (2013) conclude that the heavy tails, the correlation dynamics, and the tail dependence of factor returns represent phenomena of economic significance for investment management.

Even though linear models in which SDF is a linear function of the equity factors are standard in the asset pricing literature, econometric techniques for dealing with nonlinear models have been developed, see (Cochrane, 2005, Section 9.3). For example, Almeida and Garcia (2016) consider a class of SDFs that arise by imposing moment conditions on the space of admissible SDFs by means of a class of convex discrepancy functionals. The dual formulation can be related to a portfolio problem with a hyperbolic absolute risk-aversion (HARA) utility the solution of which is a nonlinear function of the factors. Highly nonlinear SDFs are associated with a higher degree of risk aversion. Several models in the asset pricing literature, such as Hansen and Jagannathan (1991), Snow (1991), and Černý (2003), arise as special cases. An empirical application of the nonlinear model is provided in Almeida, Ardison, and Garcia (2020).

Our article contributes to the literature by ranking the relative importance of the empirical properties of equity factors for near-linear and highly nonlinear SDFs with a high degree of risk aversion. The asset pricing model is based on Almeida and Garcia (2016) and the pricing error is estimated using the upper bound of Hansen and Jagannathan (1997). In our case, the upper bound can be interpreted as the maximal possible difference between the Sharpe ratios computed using the admissible SDF and a misspecified SDF across all risky assets. We focus on near-linear SDFs because of their relevance for the practice of finance but we also consider highly nonlinear specifications. We use monthly returns of the five Fama–French and the momentum factors from July 1963 to September 2017.

The main findings can be summarized as follows. First, we find that correlation dynamics are of first-order importance. Ignoring the autocorrelation of the factor returns leads to much smaller pricing errors on a relative basis. Moreover, assuming the conditional distribution of factor returns is Gaussian leads to relatively small pricing errors for near-linear SDFs but can become more pronounced for highly nonlinear SDFs.

Second, we assess the pricing error of the same types of misspecification on portfolios whose systematic exposure is a linear function of the five Fama–French factors and momentum. We find that, with some exceptions, the mispricing of the factors on a stand-alone basis leads to insignificant pricing errors in general. However, some linear combinations of factor exposures can be significantly overpriced by the misspecified SDF. This holds for all levels of risk aversion when the misspecification includes constant covariances and constant correlations and for some levels of risk aversion when autocorrelation is ignored and when the conditional excess return distribution is incorrectly assumed to be Gaussian.

Our results complement Christoffersen and Langlois (2013), Bali, Engle, and Tang (2016), and Engle (2016) by emphasizing the importance of correlation dynamics in asset pricing with linear SDFs. Also, our results suggest that underestimating the heavy tail of the conditional distribution of factor returns can introduce a more pronounced pricing error for high levels of risk aversion especially when considering derivative claims.

The article is organized in the following way. The statistical methodology is described in Section 1. Sections 2 and 3 discuss the data and the empirical results, respectively. Our conclusions are summarized in Section 4.

1 Methodology

Consistent with the asset pricing literature, we assume that transactions take place at two time instants, t and $T > t$. At the time instant t, financial assets are purchased and at time T the payoffs are received; the payoffs are viewed as random variables as of time t. The prices of financial assets at time t are then represented as the expected discounted value of the future payoffs conditional on the information at time t in which the payoffs are discounted state-by-state by means of an SDF.

More formally, let $(\Omega, \mathfrak{F}, \mathbb{P})$ denote a probability space and $p_{i,t}$ denote the price of a financial asset at time t in a set of assets under consideration. Denote the gross return of an asset by $R_{i,T} = p_{i,T}/p_{i,t}$ and suppose that there are K primitive assets (or basic assets) with gross returns $X_T = (X_{1,T}, \ldots, X_{K,T})$.

The fundamental asset pricing equation can be written in the following form in terms of excess returns rather than prices,

$$\mathbb{E}[R_{i,T} - R_{f,T}] = -\frac{\operatorname{cov}(M_T, R_{i,T} - R_{f,T})}{\mathbb{E}[M_T]}, \tag{1}$$

where M_T is the SDF at time T, $R_{f,T} = 1/\mathbb{E}[M_T]$ denotes the risk-free rate, see, for example, Campbell (2000) and Cochrane (2005).

In empirical asset pricing, the most common models are linear expected return–beta models such as CAPM, APT, etc. In these models, the expression in Equation (1) is not explicitly utilized but they are known to be equivalent to an SDF which is a linear function of the primitive assets, see Cochrane (2005). Consider the following linear factor model

$$R_{i,T} - R_{f,T} = (X_T - R_{f,T})' \beta_i + \epsilon_i \tag{2}$$

where $\beta_i = (\beta_{1,i}, \ldots, \beta_{K,i})$ denotes the beta exposures of the i-th risky asset and ϵ_i is the corresponding residual. Equation (1) can be represented (see Cochrane, 2005; Engle, 2016) as

$$\mathbb{E}[R_{i,T} - R_{f,T}] = -\frac{\operatorname{cov}(-\frac{1}{R_{f,T}}\mathbb{E}[X_T - R_{f,T}]C_X^{-1}X_T, R_{i,T} - R_{f,T})}{\mathbb{E}[M_T]},$$

where C_X is the covariance matrix of the excess returns $X_T - R_{f,T}$ computed at time t, implying the following expression for the SDF

$$M_T = \frac{1}{R_{f,T}} - \frac{1}{R_{f,T}}\mathbb{E}[X_T - R_{f,T}]C_X^{-1}(X_T - \mathbb{E}[X_T]), \tag{3}$$

which is a linear function of the primitive assets. Even though the parameter C_X is of first-

order importance, other empirical characteristics can be relevant. For example, if the multivariate distribution of X_T is heavy tailed, an econometrician may choose an estimation procedure that recognizes this fact.[1]

Furthermore, as noted by (Cochrane, 2005, Section 9.3), the asset pricing model need not be linear. In nonlinear models, characteristics of the distribution of X_T other than C_X may become significant. For example, if the SDF is a nonlinear function of the primitive assets $M_T = M_T(X_T)$, then in the classical two-period model the properties of M_T depend on the entire joint distribution of the random vector X_T. Therefore, neglecting a relevant empirical fact results in a misspecified random vector X_T^* and a misspecified SDF, $M_T^* = M_T^*(X_T^*)$.

In empirical asset pricing, it is uncommon to work with an ex-ante formulation which requires a full specification of the joint distribution of the basic assets and the set of risky assets in order for the right-hand side of Equation (1) to be determined completely. In empirical research, it is common to use time series averages to estimate unconditionally the corresponding expectations, which means that the researcher takes advantage of the corresponding empirical joint distribution. Regarding the specification of M_T, the vast majority of the empirical research is focused on which factors should be included in the vector X_T.

In conditional asset pricing models, however, a nonparametric estimation is not as straightforward and parametric formulations are common. For instance, in the linear case, as noted by Engle (2016), assuming a C_X which is not time-varying implies an SDF M_T^* computed by Equation (3) which may differ from an M_T computed by the same equation assuming a dynamic C_X. As a consequence, working with M_T^* instead of M_T may introduce a pricing error if the constant covariance assumption represents a significant misspecification. An empirical study of the value added of asset pricing models such as Equation (2) assuming a dynamic C_X and, therefore dynamic beta exposures β_i, is provided by Bali, Engle, and Tang (2016) using a parametric setup.

In this article, we develop a method to test the incremental impact of several stylized facts of X_T. We work with the unconditional formulation provided in Equation (1) and the method allows us to exclude certain empirical features while holding the set of basic assets fixed when evaluating the right-hand side. An advantage of the unconditional formulation is that it allows us to consider nonlinear SDFs which, to the best of our knowledge, have not been adopted in conditional models yet.

The remainder of this section is organized in the following way. We begin with a description of the nonlinear SDF model and the method for sampling from the misspecified SDF. Then, we explain how the pricing error for a portfolio of basic assets is computed and we describe an algorithm that bootstraps its distribution in order to test for significance.

1.1 Evaluating the Impact of a Misspecified SDF

Instead of focusing on linear models only, we consider a more general framework. Almeida and Garcia (2016) construct an SDF by imposing moment restrictions on all admissible SDFs which are determined by a homogeneous discrepancy function selected to be the Cressie–Read family of discrepancies, Cressie and Read (1984). The model generalizes Hansen and Jagannathan (1991), Snow (1991), and Černý (2003) among others.

1 Toda and Walsh (2017) discuss this issue in the context of the generalized methods of moments.

Subject to some regularity conditions, Almeida and Garcia (2016) derive the following expression for the SDF,

$$M_T = \begin{cases} (a^\gamma + \gamma\lambda'(X_T - R_{f,T}))^{1/\gamma}, & \gamma \neq 0 \\ ae^{\lambda'(X_T - R_{f,T})}, & \gamma = 0 \end{cases} \tag{4}$$

where $a = \mathbb{E}(M_T) = 1/R_{f,T}$, $\gamma \in \mathbb{R}$ can be interpreted as a risk-aversion parameter and λ is the solution to the following problem,

$$\begin{aligned} &\max_{\nu} \mathbb{E}[u_\gamma(\nu, X_T)] \\ &s.t.\ \ a^\gamma + \gamma\nu'(X_T - R_{f,T}) > 0. \end{aligned} \tag{5}$$

where the objective function is given by

$$u_\gamma(\nu, X_T) = \begin{cases} \dfrac{a^{\gamma+1}}{\gamma+1} - \dfrac{1}{\gamma+1}\left(a^\gamma + \gamma\nu'(X_T - R_{f,T})\right)^{\frac{\gamma+1}{\gamma}}, & \gamma \neq 0 \\ a - ae^{\nu'(X_T - R_{f,T})}, & \gamma = 0 \end{cases}$$

Almeida and Garcia (2016) show that this model prices exactly the factors X_T and that the model is arbitrage-free, that is, $\mathbb{E}[M_T(X_{i,T} - R_{f,T})] = 0$, $i = 1, \ldots, K$, and $M_T > 0$ in all states of the world.[2]

The optimization problem in Equation (5) is interpreted as an optimal portfolio problem of a representative investor with a HARA type of utility function in which γ is a risk-aversion parameter. Almeida and Garcia (2016) show that specific values of γ correspond to common utility functions, such as logarithmic ($\gamma = -1$), exponential ($\gamma = 0$), and quadratic ($\gamma = 1$). Finally, γ changes the relative weight of the higher-order moments of the distribution of νX_T in the objective function: γ close to one implies negligible weights of all higher-order moments, $\gamma < 1$ increases the relative weight of skewness and kurtosis and corresponds to an increasing degree of risk aversion, see Almeida and Garcia (2016).

A linear asset pricing model arises with $\gamma = 1$ and is a special case. Because linear SDFs are equivalent to linear factor models, see for example (Cochrane, 2005, Chapter 6), we pay special attention to linear and near-linear SDFs that belong to the class of SDFs defined by Equation (4).

The SDF M_T as calculated in Equation (4) depends on the empirical properties of the excess returns of the primitive assets X_T. If some empirical properties are ignored, then the resulting SDF is misspecified and can fail to price the primitive assets correctly leading to pricing errors of some magnitude. It is, therefore, of practical and theoretical interest to quantify the magnitude of the pricing errors that can result from a misspecified SDF.

In particular, we are interested in finding an estimate of the pricing error arising from a misspecified SDF where the misspecification concerns failure to incorporate the following stylized facts: (i) autocorrelations, (ii) dynamics of the conditional covariances, (iii) dynamics of correlations, and (iv) the fat tails of the residual process. A discussion of these properties for commonly used equity factors is provided by Christoffersen and Langlois (2013).

We note that such misspecifications may not necessarily lead to large pricing errors. For instance, an empirical phenomenon may not be very pronounced in the data and may be of

2 Without loss of generality, we consider the normalized SDF $M_T = a\frac{(a^\gamma + \gamma\lambda'(X_T - R_{f,T}))^{1/\gamma}}{\mathbb{E}[(a^\gamma + \gamma\lambda'(X_T - R_{f,T}))^{1/\gamma}]}$ which guarantees $\mathbb{E}M_T = 1/R_{f,T}$.

little consequence for asset prices for some (or all) levels of risk aversion. Alternatively, a well-pronounced empirical property may turn out to be insignificant for asset prices.

Denote by M_T^* a misspecified SDF computed in Equation (4) by ignoring certain empirical phenomena. If the correctly specified M_T and the misspecified M_T^* provide identical prices for the risk-free asset,

$$\mathbb{E}[M_T] = \mathbb{E}[M_T^*] = \frac{1}{R_{f,T}},$$

then the pricing error resulting from the misspecification can be quantified in the following way

$$\mathbb{E}[R_{i,T} - R_{f,T}] - \mathbb{E}^*[R_{i,T} - R_{f,T}] = -\frac{\text{cov}(M_T - M_T^*, R_{i,T} - R_{f,T})}{\mathbb{E}[M_T]}, \tag{6}$$

where i denotes a financial asset and $\mathbb{E}^*[R_{i,T} - R_{f,T}]$ denotes the expected risk premium computed according to M_T^*. Equation (6) holds for any financial asset or a portfolio of financial assets.

Equation (6) can be used directly only if we can specify a set of financial assets for which we wish to assess the asset pricing error because the right-hand side depends on that choice. For example, in empirical asset pricing, it is common to work with certain portfolios of stocks or with a certain cross-section of stocks. The right-hand side of Equation (6) is then evaluated for each of the portfolios or the stocks.

We adopt the more general approach of Hansen and Jagannathan (1997) which is asset independent. Consider the following bound,

$$\left|\frac{\mathbb{E}[R_{i,T} - R_{f,T}] - \mathbb{E}^*[R_{i,T} - R_{f,T}]}{\sigma(R_{i,T} - R_{f,T})}\right| \leq \frac{\left(\mathbb{E}(M_T - M_T^*)^2\right)^{1/2}}{\mathbb{E}[M_T]}, \tag{7}$$

which follows from the Cauchy–Schwartz inequality applied to Equation (6). The upper bound is asset independent and has a direct economic interpretation as the maximal absolute Sharpe ratio difference across all risky assets, see Hansen and Jagannathan (1997) and Almeida and Garcia (2012) for a generalization. Our setting is simpler in that M_T is defined through Equation (4) and we do not consider the infimum of the upper bound in Equation (7) across all admissible SDFs as in Hansen and Jagannathan (1997). Our objective is to study the behavior of the upper bound as a function of γ.

We consider four types of misspecification for different values of the γ parameter in Equation (4) describing near-linear and also highly nonlinear SDFs:

A. The excess factor returns are not autocorrelated.
B. The covariance matrix of the excess factor returns is constant.
C. The correlations of the excess factor returns are constant but the volatilities are time-varying.
D. The excess factor returns are conditionally Gaussian with time-varying volatilities and correlations.

Using the upper bound in Equation (7), we can rank the misspecifications A through D. Furthermore, by comparing Case B to Case C we can draw conclusions about the relative significance of volatility dynamics and correlation dynamics.

To evaluate the upper bound in Equation (7), we need the bivariate distribution of the random vector (M_T, M_T^*). We suggest obtaining a sample of the joint distribution by using a model general enough to describe the corresponding stylized properties which we also use to obtain synthetic data by simplifying the model while keeping the states of the world fixed.

1.2 Sampling from Misspecified SDFs

In this section, first we describe our assumption for the multivariate process of the excess factor returns and then we explain how joint observations on (M_T, M_T^*) are obtained. In the remaining part of this article, the notation changes in the following way. Because we estimate the expectation in Equation (7) by a time series average, the index t is used to denote the time of observation and essentially identifies a state of the world. To simplify notation further, we drop the index T in the SDF notation.

We assume that the K-dimensional vector of factor returns X_t follows a multivariate AR-DCC-GARCH process[3] of the following type

$$\begin{aligned} X_t &= \mu + \phi' X_{t-1} + \eta_t,\ \eta_t | I_{t-1} \in N(0, \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t) \\ \mathbf{D}_t^2 &= \mathrm{diag}(\omega_i) + \mathrm{diag}(\kappa_i)^{\circ} \eta_{t-1} \eta_{t-1}' + \mathrm{diag}(\lambda_i)^{\circ} \mathbf{D}_{t-1}^2 \\ Z_t &= \mathbf{D}_t^{-1} \eta_t \\ \mathbf{Q}_t &= \overline{\mathbf{Q}}(1 - \alpha - \beta) + \alpha(Z_{t-1} Z_{t-1}') + \beta \mathbf{Q}_{t-1} \\ \mathbf{R}_t &= \mathrm{diag}(\mathbf{Q}_t)^{-1} \mathbf{Q}_t \mathrm{diag}(\mathbf{Q}_t)^{-1} \end{aligned} \tag{8}$$

in which ϕ is a vector of coefficients, $\mathbf{D}_t^2$ is a diagonal matrix with the factor return variances on the main diagonal, the notation $^{\circ}$ denotes the Hadamard product, $\mathbf{Q}_t$ is the covariance matrix of the vector of i.i.d. residuals Z_t, $\mathbf{R}_t$ is the dynamic correlation matrix, and α and β are positive constants with $\alpha + \beta < 1$ and have the general interpretation of the parameters of the one-dimensional GARCH(1,1).[4] The notation $\mathrm{diag}(\omega_i)$ is a diagonal matrix with $\omega = (\omega_1, \ldots, \omega_K)$ on the main diagonal, $\mathrm{diag}(\hat{\mathbf{Q}}_t)$ denotes a diagonal matrix with the diagonal of $\hat{\mathbf{Q}}_t$ on the main diagonal, and $\eta_t | I_{t-1}$ denotes conditioning on the information set at $t-1$.

The structure of the process in Equation (8) is standard in the empirical literature, see, for example, Engle (2002). The variances of the components of η_t follow a univariate GARCH(1,1) process with parameters ω_i, κ_i, and λ_i satisfying the condition $\kappa_i + \lambda_i < 1$ for the i-th component.

The model is estimated using the Gaussian quasi-maximum likelihood method which can lead to a consistent and asymptotically normal estimator even if the true distribution is non-Gaussian under the assumption that the innovation has a finite fourth moment, see Elie and Jeantheau (1995), Hall and Yao (2003), and Horv and Kokoszka (2003).[5] After estimating the model, we calculate the normalized residual

3 Christoffersen and Langlois (2013) note that very small differences are observed between the modified DCC model by Aielli (2013) and the original formulation by Engle (2002) in their empirical study. Our empirical setup is very similar and we use the original formulation by Engle (2002).

4 The choice of the lag in the AR component is justified in Section 4.1.

5 As a robustness check, we also estimate the model parameters using a multivariate Student t distribution and obtain very similar results.

$$\hat{\epsilon}_t = \hat{\mathbf{R}}_t^{-1/2} \hat{\mathbf{D}}_t^{-1} \hat{\eta}_t. \tag{9}$$

The intuition behind the method to obtain joint samples can be explained as follows. Suppose for simplicity that the true values of the parameters are known. Then, the actual observed excess returns are deterministic functions of ϵ_t provided by Equation (8). In this context, the four misspecifications can be viewed as special cases of the model in Equation (8); that is, each misspecification is represented by a simpler function of the same residual process.

Consider misspecification A. Joint observations on the two SDFs (M, M^*) are obtained in the following way:

A1. M is obtained from Equation (4) by solving Equation (5) with the historical data.
A2. We estimate the AR-DCC-GARCH model in Equation (8) using the full available history of the equity factors and compute $\hat{\epsilon}_t$ using Equation (9).
A3. Synthetic data X_t^* for the excess factor returns are computed by

$$X_t^* = \hat{\mu}_\phi + \hat{\mathbf{D}}_t \hat{\mathbf{R}}_t^{1/2} \hat{\epsilon}_t,$$

in which we ignore the autocorrelation term in Equation (8).[6] Each component of $\hat{\mu}_\phi$ equals $\hat{\mu}_{\phi,i} = \hat{\mu}_i / (1 - \hat{\phi}_i)$.
A4. M^* is computed from Equations (4) and (5) for a fixed γ using the synthetic data X_t^*. Because the states of the world as represented by $\hat{\epsilon}_t$ are fixed, we obtain joint observations on the pair (M, M^*) and the calculation of the upper bound in Equation (7) is straightforward.

Because the states of the world defined by the normalized residuals are kept the same, the only difference between X_t and X_t^*, and therefore between M and M^*, is the autocorrelation component of Equation (8).

Steps A1, A2, and A4 are the same for all misspecifications A through D. The differences are in the way the synthetic data X_t^* are generated. In the following, we describe the third step for the remaining three misspecifications.

Step B3 of misspecification B takes the following form:

B3a. Compute the unconditional covariance matrix $\hat{\mathbf{C}}$ of the estimated residual process $\hat{\eta}_t$.
B3b. Synthetic data X_t^* for the excess factor returns are computed by

$$X_t^* = \hat{\mu} + \hat{\phi} X_{t-1}^* + \hat{\mathbf{C}}^{1/2} \hat{\epsilon}_t.$$

In this case, the difference between X_t and X_t^* is that the matrix multiplying $\hat{\epsilon}_t$ is not time-varying.

Step C3 of misspecification C takes the following form:

C3a. Compute the unconditional correlation matrix $\hat{\mathbf{R}}$ of the estimated residual process $\hat{\eta}_t$.
C3b. Synthetic data X_t^* for the excess factor returns are computed by

6 Note that by virtue of the estimation process, if we set $X_t^* = \hat{\mu} + \hat{\phi} X_{t-1}^* + \hat{\mathbf{D}}_t \hat{\mathbf{R}}_t^{1/2} \hat{\epsilon}_t$, then we recover the historical data; that is, $X_t = X_t^*$.

$$X_t^* = \hat{\mu} + \hat{\phi} X_{t-1}^* + \hat{\mathbf{D}}_t \hat{\mathbf{R}}^{1/2} \hat{\epsilon}_t.$$

This misspecification is different from Case B in that the conditional correlation matrix is not time-varying while the volatilities are allowed to be time-varying.
Step D3 of misspecification D takes the following form:

D3a. For each equity factor, the empirical distribution function of the normalized residual is evaluated at the value of the observed normalized residuals. The resulting values have a uniform distribution.
D3b. For each equity factor, the results from the previous step are plugged into the quantile function of a Gaussian distribution estimated from the corresponding normalized residual. As a result, we have a new ϵ_t^* such that the vector component follows a Gaussian distribution while retaining the historical joint dependence of the components.
D3c. Synthetic historical data X_t^* are generated using Equation (9) with the fitted $\hat{\mathbf{D}}_t$ and $\hat{\mathbf{R}}_t$,

$$X_t^* = \hat{\mu} + \hat{\phi} X_{t-1}^* + \hat{\mathbf{D}}_t \hat{\mathbf{R}}_t^{1/2} \hat{\epsilon}_t^*.$$

Note that by construction $\hat{\epsilon}_t^*$ is a function of $\hat{\epsilon}_t$ and the assumption that the states of the world should be fixed has not been violated. Furthermore, the transformation affects not only the marginal distributions but also the severity of the joint extremes. Because the pair (X_t, X_t^*) has to be defined jointly in the same states of the world, disentangling the impact of the marginal distribution from that of the joint extremes is not straightforward.
In this case, the conditional covariances remain as initially estimated and the difference between M and M^* are driven by the assumption of normality. If the estimated normalized residuals are almost Gaussian, then X_t and X_t^* would be very similar which would result in similar M and M^* in each state of the world.

1.3 The Pricing Error for Portfolios of Basic Assets

The upper bound in Equation (7) may not be attained in a given set of risky assets. Across all assets, the parameter that determines the proximity to the upper bound is the correlation between the SDF difference $M_\gamma - M_\gamma^*$ and $R_{i,T} - R_{f,T}$ where $R_{i,T}$ denotes the gross return of a risky asset. This is demonstrated by rewriting Equation (6) in the following way

$$\frac{\mathbb{E}[R_i - R_f] - \mathbb{E}^*[R_i - R_f]}{\sigma(R_i - R_f)} = -\rho_i \frac{(\mathbb{E}[M_\gamma - M_\gamma^*]^2)^{1/2}}{\mathbb{E}[M_\gamma]}, \tag{10}$$

where ρ_i is the correlation between $M_\gamma - M_\gamma^*$ and $R_i - R_f$. The upper bound in Equation (7) is attained for risky assets with $|\rho_i| = 1$.

In order to find portfolios of the primitive assets most affected by the misspecification, we solve the following problem for every type of misspecification A through D.

$$\min_w \mathbb{E}(\Delta_M - w'X)^2 \tag{11}$$

where $X = (X_1, \ldots, X_K)$ denotes the excess returns of the basic assets. The solution to this problem replicates Δ_M in the best possible way in the mean-squared sense and represents a linear combination of systematic exposures which is most overpriced by M^*. The optimization problem in Equation (11) is a projection onto the linear span of the basic assets and

has the closed-form solution $\overline{w} = (X'X)^{-1}X'\Delta_M$. Flipping the sign of $\overline{w}$ provides the linear combination of systematic exposures which is most underpriced by M_T^*.

Because M and M^* are in general nonlinear functions of the basic assets, the upper bound of the pricing error may not be attained in the linear span of the basic assets. Therefore, a large upper bound in Equation (7) does not necessarily imply a significant pricing error on the linear span of the basic assets. The asset most affected by the misspecified SDF might be a highly exotic derivative of little practical significance. For this reason, apart from evaluating the upper bound in Equation (7), we also evaluate the impact of the misspecification on portfolios of basic assets by means of solving the problem in Equation (11).

1.4 Computing the Significance of the Pricing Error

Equation (6) computes the pricing error of a risky asset, which can be one of the basic assets or a portfolio of them. In practice, the right-hand side is estimated by a time series average in which the two SDFs are computed using Equations (4) and (5) together with the corresponding specifications of X_t. As a consequence, it is important to be able to test if the estimated pricing error is statistically significant given a pair (M, M^*).

The distribution of the pricing error for a particular asset is bootstrapped numerically using the following approach:

1. Take the real and the synthetic data (X_t, X_t^*) where $t \in \mathcal{T}$ denotes a period of time. The algorithms in Section 1.2 guarantee that the real and the synthetic data are defined on a common probability space.
2. Draw a random sequence of indices $\mathcal{T}_b$ from $\mathcal{T}$ with replacement.
3. Calculate the pair (M, M^*) by solving the optimization problem in Equation (5) using the data (X_t, X_t^*) where $t \in \mathcal{T}_b$.
4. Calculate the right-hand side of Equation (7) using $\mathcal{T}_b$ for the particular asset being considered.

Repeating this process many times results in a sample from the distribution of the pricing error. The calculation of the p-value of the pricing error obtained from the original data (X_t, X_t^*) where $t \in \mathcal{T}$ is straightforward.

2 Data

The dataset used in the risk assessment application consists of monthly returns of the five factors in the five-factor model by Fama and French (2005)—market (MKT), size (SMB), value (HML), investment (CMA), and profitability (RMW)—as well as momentum (MOM) for the period from July 1963 to September 2017. We use the one-month Treasury bill rate for the same period as a proxy for the risk-free rate. We set the risk-free rate to be the average one-month Treasury bill rate for the full period, which equals 0.39%. The returns were downloaded from the data library of Kenneth French.[7] In total, we work with 651 monthly observations.

As a robustness study, we repeat the analysis with weekly data for the same period and for the same factors. The empirical analysis is, however, based on the monthly data.

7 See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. We use excess returns for the market factor.

3 Empirical Analysis

The empirical analysis is divided into three parts. First, we discuss the relevant stylized facts of the empirical data and the degree to which they are pronounced in the monthly returns. Second, we quantify the relative importance of the misspecifications in A through D described in Section 1.1 for different levels of risk aversion by means of the upper bound in Equation (7). We look at both near-linear and highly nonlinear SDFs characterized by high levels of risk aversion.

Finally, we focus on equity portfolios whose systematic risk is a linear combination of the six factors. We calculate both the pricing error of the primitive assets and the factor exposures which result in the largest possible overpricing by the misspecified SDF.

Because we work with monthly returns of the five Fama–French and the momentum factors, our results in the section are reported as monthly statistics. Even though the upper bound in Equation (7) is related to the Sharpe ratio, we do not annualize because the econometric model in Equation (8) does not allow for a straightforward way to do that, see Lo (2002).

3.1 Stylized Facts of the Data

In this section, we confirm the presence of the following phenomena in the monthly returns data: autocorrelation, clustering of volatility, correlation dynamics, and a non-Gaussian residual process. Christoffersen and Langlois (2013) report similar phenomena in the weekly returns but because our analysis is based on monthly data, as a first step we verify if at that frequency the same empirical phenomena are sufficiently pronounced.

The autocorrelation plots of the factor returns in Figure 1 indicate that HML, RMW, and CMA have a significant autocorrelation of order one while the same order is borderline insignificant for MKT. The AR component of the model in Equation (8) is supposed to capture that statistically.

The estimated parameters of the AR-DCC-GARCH model for each factor are provided in Table 1. These parameter estimates include the intercept vector $\hat{\mu}$, the autoregression coefficients in the vector $\hat{\phi}$, and the individual factor GARCH(1,1) parameters $\hat{\omega}_i$, $\hat{\kappa}_i$, and $\hat{\lambda}_i$. The estimates of the joint parameters are $\hat{\alpha} = 0.104$ and $\hat{\beta} = 0.802$ and both are strongly significant. Table 1 suggests significant volatility and correlation dynamics.

An illustration of the volatility and correlation dynamics is provided in Figures 2 and 3. Figure 2 includes plots of the monthly data together with 95% confidence bands computed using the estimated AR-DCC-GARCH model. The dynamics of the confidence bands reflect the temporal behavior of the volatility as captured by the model.

Figure 3 provides plots of the estimated conditional correlations of all factor pairs which include the market factor (MKT). Correlation dynamics tend to be rather pronounced in some cases with correlations changing from -0.4 to 0.4 in the course of several months. All factor pairs exhibit a similar degree of correlation dynamics, Figure 3 includes only the pairs with MKT because of the significance of that factor for asset pricing.

Finally, Table 2 includes descriptive statistics of the normalized residual. The skewness and kurtosis indicate substantial departures from normality in some cases. The Shapiro–Wilk test (Shapiro and Wilk, 1965) shows that we can accept normality only for the normalized residual of CMA. The normalized residuals of MKT, SMB, and MOM are strongly non-Gaussian.

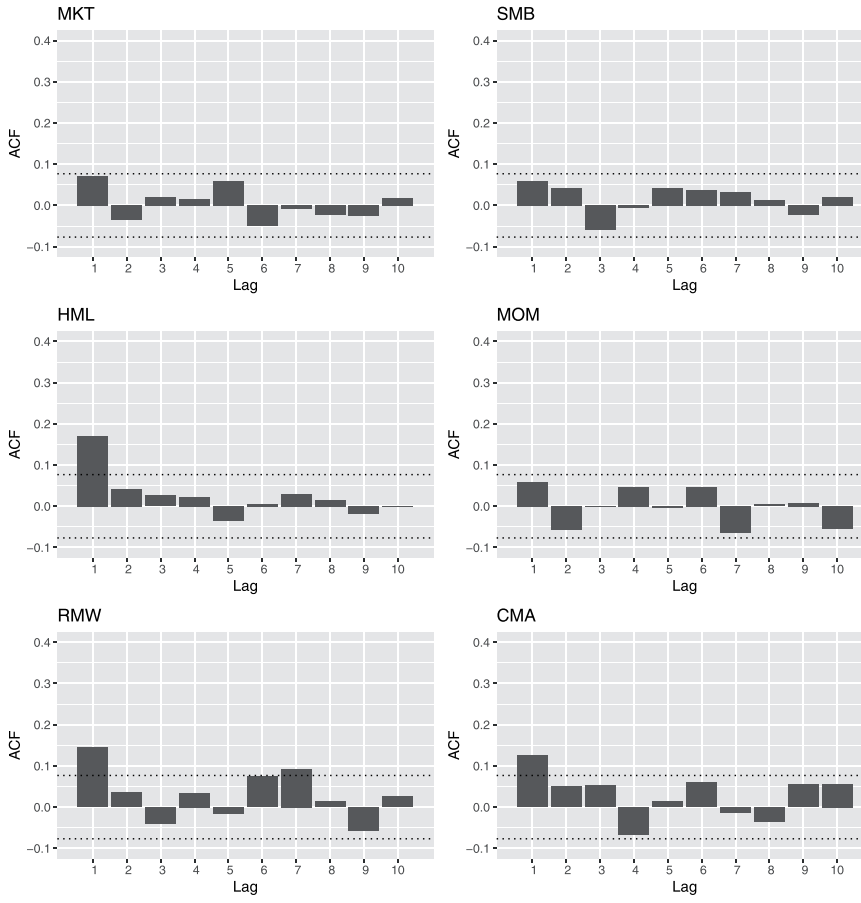

Figure 1. Estimated autocorrelation functions (ACF) of the monthly returns of the six factors up to lag ten with 95% confidence bounds. The HML, CMA, and RMW factors have significant autocorrelation of order one while the MKT, SMB, and MOM factors do not exhibit significant autocorrelations.

3.2 Impact on the Upper Bound

The empirical analysis in Section 3.1 implies that autocorrelations and fat tails are pronounced for some of the factors and the volatility and correlation dynamics are significant for all variables. In this section, we quantify the relative significance of these phenomena for asset pricing when the SDF is constructed according to Equation (4).

We introduce the following notation for the upper bound in Equation (7) for each of the four misspecifications,

$$\delta_{Missp,\gamma} \quad = \frac{1}{\mathbb{E}[M_\gamma]}(\mathbb{E}[M_\gamma - M^*_{Missp,\gamma}]^2)^{1/2},$$

in which the following abbreviations are used for the misspecifications in the subscript of $M^*_{Missp,\gamma}$ and $\delta_{Missp,\gamma}$: AR assumes no AR effect (misspecification A), Cov assumes a constant

Table 1. Estimated parameters of the model in Equation (8) and their *p*-values corresponding to each of the factors

		μ	ϕ	ω	κ	λ
MKT	Estimate	0.006441	0.044134	0.000073	0.128153	0.843764
	p-value	0.000121	0.306824	0.06154	0.000078	0
SMB	Estimate	0.001742	0.084794	0.000053	0.125718	0.821851
	p-value	0.153421	0.06585	0.033159	0.003065	0
HML	Estimate	0.002346	0.182289	0.000051	0.160977	0.771699
	p-value	0.037872	0.000045	0.019095	0.000045	0
MOM	Estimate	0.004571	0.022973	0.000137	0.419083	0.567123
	p-value	0.000476	0.620436	0.01574	0.001877	0
RMW	Estimate	0.002041	0.146328	0.000024	0.182671	0.753058
	p-value	0.002682	0.000304	0.008662	0.00323	0
CMA	Estimate	0.001861	0.148197	0.000017	0.159319	0.799946
	p-value	0.013726	0.000505	0.112547	0.000117	0

The estimated values for the joint parameters are $\hat{\alpha} = 0.10447$ and $\hat{\beta} = 0.802321$ both have *p*-values of zero. The factors are MKT, SMB, HML, MOM, CMA, and RMW.

conditional covariance matrix (misspecification B), Cor assumes a constant conditional correlation (misspecification C), and *N* assumes that the AR-DCC-GARCH residuals for all factors are Gaussian (misspecification D). Based on the average one-month Treasury bill rate, we estimate $\mathbb{E}[M_\gamma] = 1/R_f = 0.9961$.

The misspecified SDFs $M^*_{Missp,\gamma}$ and M_γ are computed according to the algorithms in Section 1.1 with γ fixed to one and the same value. We are interested in the behavior of the bounds for SDFs which are linear and near-linear functions of the factors and also in SDFs which are highly nonlinear. For this reason, we choose values for γ which are close to one and values that deviate away from one in both directions.

Plots of estimated SDFs are provided in Figure 4. The top plot shows the estimated M_γ, the plot in the middle corresponds to $M_{Cov,\gamma}$, and the bottom plot corresponds to $M_{N,\gamma}$. The plots are produced with the risk-aversion parameter set to $\gamma = 1$. High values of an SDF identify unfavorable states of the economy such as recessions or market crashes. This is clearly visible on the top plot in which the most extreme values correspond to the market declines in October 1978, October 1979, March 1980, the Black Monday crash, the aftermath of the tech bubble bursting, and the financial crisis of 2008. Because the covariance matrix is assumed constant, the SDF $M_{Cov,\gamma}$ would fail to recognize states with high conditional correlations and volatilities as very risky and states with low conditional correlations and volatilities as very favorable. As a consequence, M_1 is expected to have more pronounced extremes than $M_{Cov,1}$. This is confirmed by comparing the top plot to the middle plot in Figure 4.

Comparing the top plot to the bottom plot in Figure 4 suggests that the main differences between M_1 and $M_{N,1}$ are in the upper tails. Because the conditional distribution of the residual process is incorrectly assumed to be Gaussian, $M_{N,1}$ underestimates the risk of states associated with severe market crashes. The lower tails of M_1 and $M_{N,1}$ appear to be rather similar.

Estimated values of the upper bound in Equation (7) are provided in Table 3. Because we work with monthly data, the numbers in each row are interpreted as maximal differences

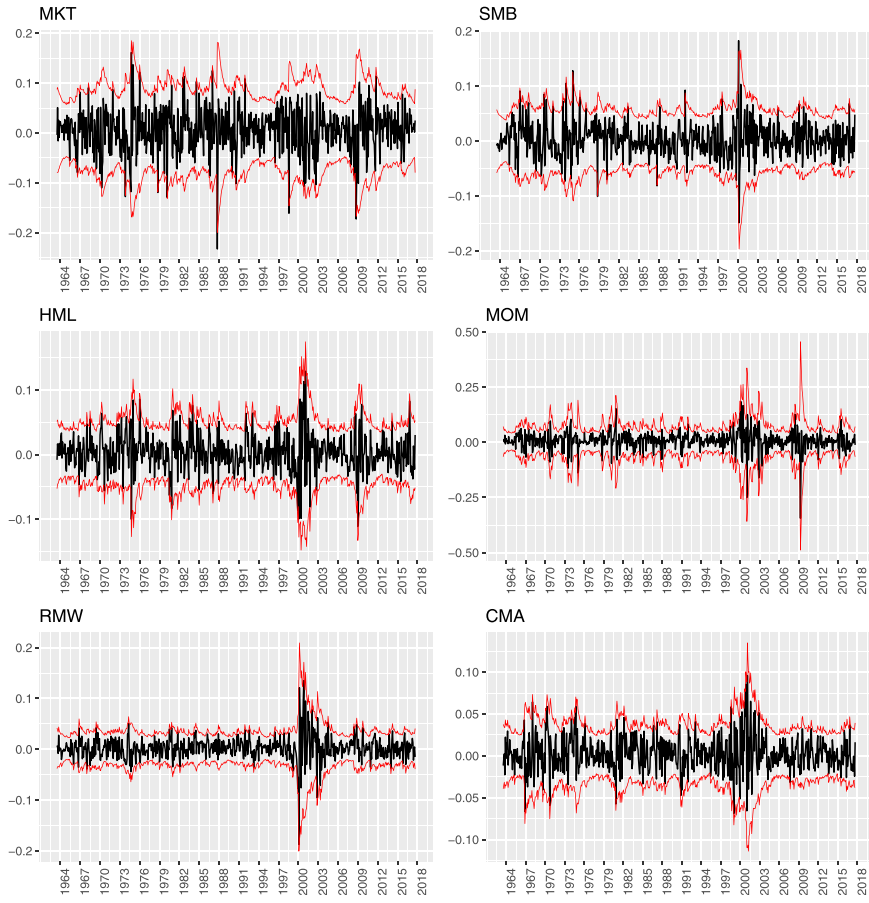

Figure 2. Monthly factor returns together with 95% confidence bands computed through the fitted model. The clustering of volatility effect is clearly visible. The factors considered are MKT, SMB, HML, MOM, CMA, and RMW.

between the monthly Sharpe ratios computed according to M_γ and the respective $M^*_{Missp,\gamma}$ across all risky assets with the risk-aversion parameter γ fixed to the corresponding value in the column.

Table 3 shows that in the case of linear and near-linear SDFs (γ close to one), ignoring the covariance dynamics and the correlation dynamics lead to the largest pricing errors. If we compare $\hat{\delta}_{Cov,\gamma}$ to $\hat{\delta}_{Cor,\gamma}$, we notice that correlation dynamics seem to be of relatively higher significance than volatility dynamics.[8] In contrast, if the misspecification consists of ignoring autocorrelations or the fat tails, then the pricing errors are expected to be much smaller.

When the level of risk aversion increases (γ declines), the estimated values of the upper bound increase suggesting higher sensitivity to all misspecifications. The largest increase,

8 Both correlations and volatilities are assumed constant in $\hat{\delta}_{Cov,\gamma}$, while only correlations are assumed constant in $\hat{\delta}_{Cor,\gamma}$.

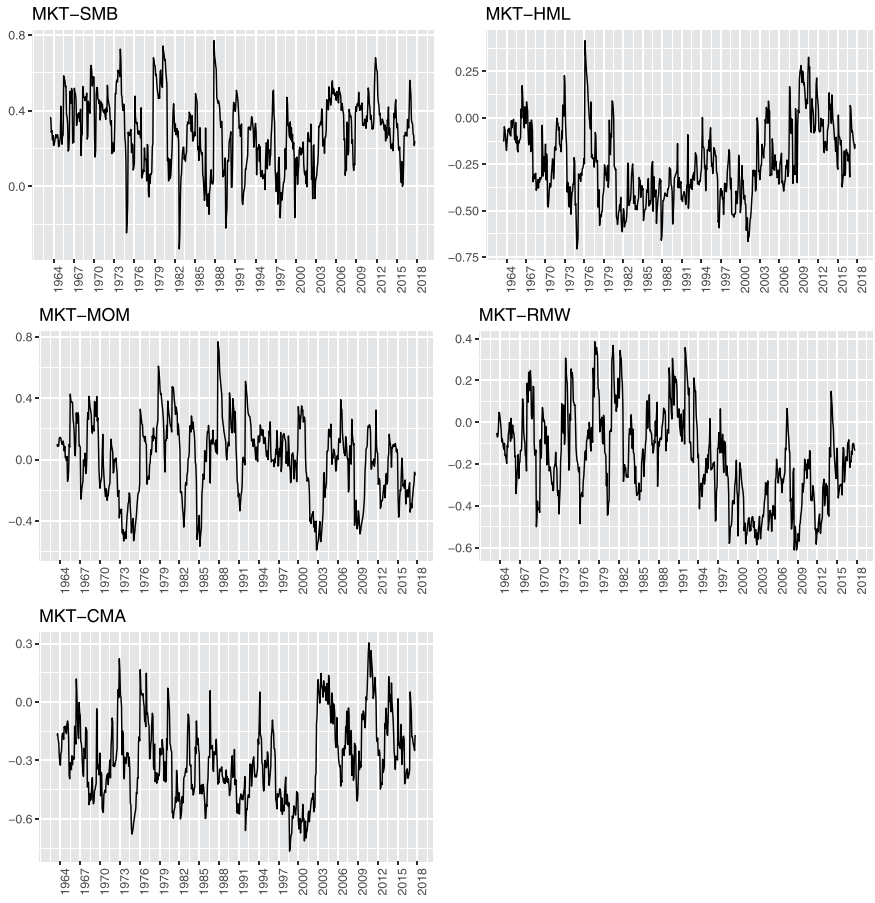

Figure 3. Estimated conditional correlations of the monthly returns of the market factor and the remaining five factors. The correlation dynamics present in the other pairs is of similar magnitude. The following abbreviations are used: MKT, SMB, HML, MOM, CMA, and RMW.

Table 2. Distributional characteristics of the normalized residual from the DCC AR-GARCH model together with *p*-values of the Shapiro–Wilk normality test

	Mean	Std	Skewness	Kurtosis	SW *p*-value
MKT	−0.041	0.995	−0.632	4.81	1.4e-08
SMB	0.019	1.025	0.239	4.13	4.3e-05
HML	−0.021	1.008	0.191	3.64	0.0158
MOM	0.064	0.996	−0.679	6.20	1.1e-12
RMW	−0.027	0.962	−0.080	3.66	0.0109
CMA	**0.002**	**1.006**	**−0.070**	**3.29**	**0.2282**

The row in bold is the only case in which the Gaussian distribution is not rejected by the Shapiro–Wilk test. The factors considered are the MKT, SMB, HML, MOM, CMA, and RMW factors.

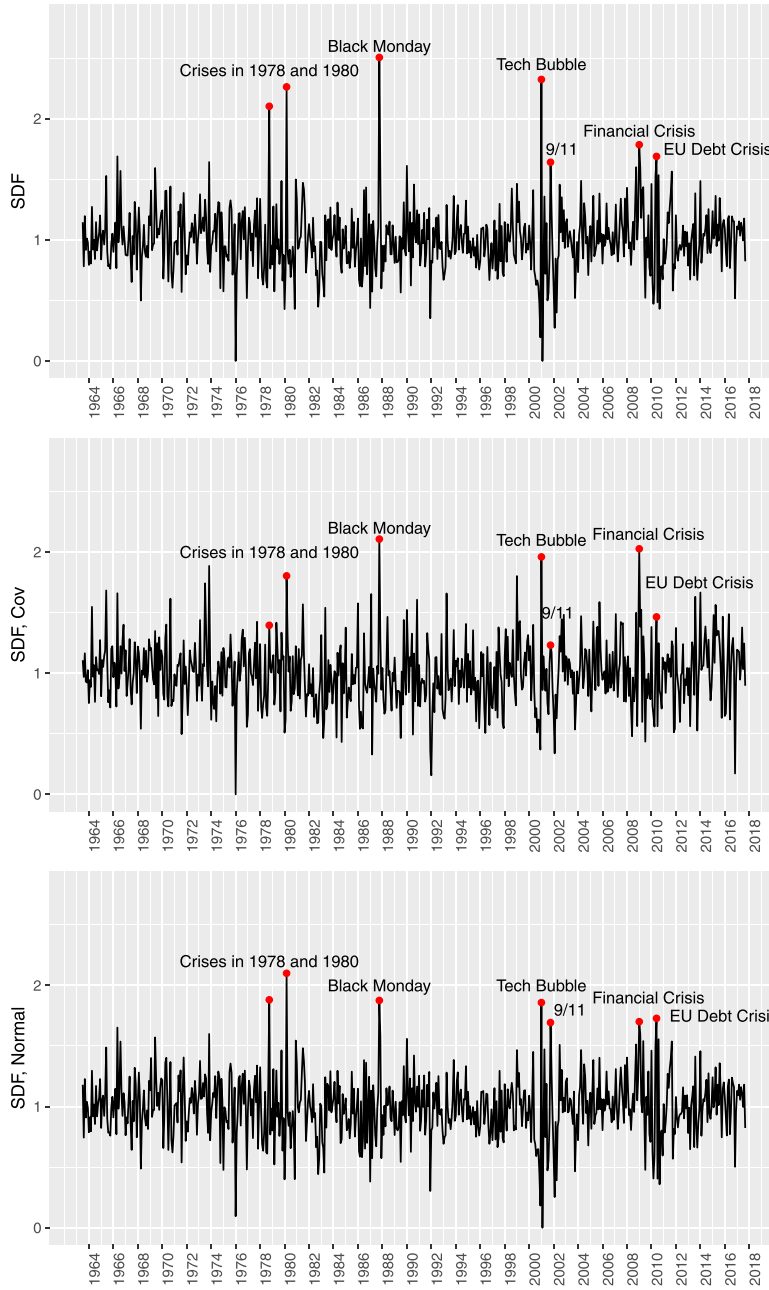

Figure 4. Monthly values of estimated SDFs using the historical data (top), assuming a constant covariance matrix (middle), and assuming a Gaussian residual (bottom). The risk-aversion parameter $\gamma = 1$ for all cases.

Table 3. Estimated values of the upper bound in Equation (7) for different values of the risk-aversion parameter γ with monthly data

γ	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
$\hat{\delta}_{AR,\gamma}$	0.081	0.054	0.049	0.047	0.045	0.032	**0.030**	0.027	0.026	0.028
$\hat{\delta}_{Cov,\gamma}$	0.782	0.530	0.333	0.238	0.209	0.188	**0.193**	0.189	0.184	0.166
$\hat{\delta}_{Cor,\gamma}$	0.341	0.251	0.198	0.169	0.153	0.118	**0.109**	0.099	0.093	0.086
$\hat{\delta}_{N,\gamma}$	0.391	0.215	0.163	0.113	0.084	0.055	**0.049**	0.048	0.040	0.038

The classical linear SDF arises with $\gamma = 1$ and is in bold. The numbers represent the maximal monthly Sharpe ratio deviations resulting from the corresponding misspecification. The following abbreviations are used: AR assumes no AR effect, Cov assumes a constant conditional covariance matrix, Cor assumes a constant conditional correlation, and N assumes the conditional distribution for all factors is Gaussian.

however, is in $\hat{\delta}_{N,\gamma}$ which shows that ignoring the heavy tail may become marginally more important when the level of risk aversion is higher.

Table 3 contains only point estimates and we need interval estimates to assess the significance of the differences between the numbers. To sample from the corresponding estimators, we repeat the procedure from Section 1.4 500 times where $\mathcal{T}$ is the collection of all months from July 1963 to September 2017.

Boxplots of the bootstrapped distributions for near-linear cases and nonlinear cases are provided in Figure 5. The top plot suggests that even though some of the corresponding values in Table 3 appear similar in magnitude, they are statistically different from each other. The bottom plot confirms that the increase in all estimates as γ declines is significant.

Based on the results in Table 3 and Figure 5, we draw the following conclusions. First, across all levels of risk aversion, the phenomenon with highest impact on pricing errors is the dynamics of the covariance matrix of the returns of the basic assets. Correlation dynamics is of first-order importance—assuming constant correlations while allowing for clustering of volatility can lead to monthly Sharpe ratio errors of up to 0.109 in the linear case. Given that most of the theory and practice of finance in equity asset pricing is based on linear models, this conclusion emphasizes the importance of research in dynamic beta models for asset pricing, see, for example, Engle (2016) and Bali, Engle, and Tang (2016).

Second, assuming the monthly returns are not autocorrelated leads to significantly smaller pricing errors compared to the remaining three types of misspecification. This result is surprising and the reason may be the relative simplicity of the AR model which is only statistical and has no state variable to provide a structural explanation. Candidates for such state variables include macroeconomic variables such as inflation and money supply, see Flannery and Protopapadakis (2002) for a comprehensive analysis albeit in a different context.

Finally, working with a Gaussian model when in fact the excess returns of the basic assets are fat-tailed leads to much more significant pricing errors when the degree of risk aversion is more pronounced. This is not surprising given that the skewness and kurtosis of the SDF have a higher weight in the corresponding Taylor series expansion (see Almeida and Garcia, 2016) when γ declines below one. Thus, the fat tails of the basic assets can lead to higher pricing errors on a relative basis as γ declines.

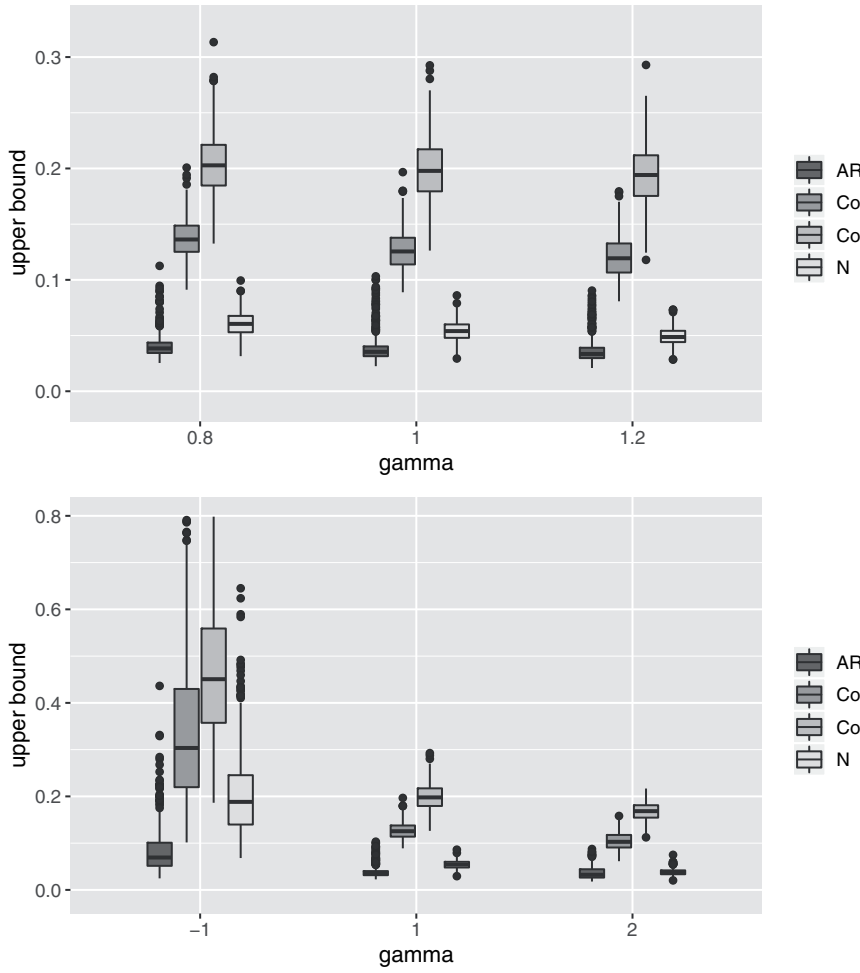

Figure 5. Bootstrapped distributions of the upper bounds estimated in Table 3 for near-linear SDFs (top) and nonlinear SDFs (bottom). The case of linear SDF ($\gamma = 1$) is included in both plots as a benchmark. The following abbreviations are used: AR assumes no AR effect, Cov assumes a constant conditional covariance matrix, Cor assumes a constant conditional correlation, and *N* assumes the conditional distribution for all factors is Gaussian.

3.3 Impact on Portfolios of Primitive Assets

Hansen and Jagannathan (1997) note that a misspecified SDF may introduce pricing errors on the set of primitive assets as well as on the collection of all derivative claims. Because we assess the pricing errors by the upper bound in Equation (7), it is unclear if the pricing errors on the primitive assets and some portfolios of them can approach the upper bound. In this section, we verify if the pricing errors of risky assets whose systematic risk is represented by a linear combination of the primitive assets are substantial.

The pricing error represented by the right-hand side of Equation (10) for each single primitive asset and type of misspecification is provided in Table 4. Using the algorithm in

Section 1.3, we calculate the p-values of the pricing errors using 500 bootstrapped samples. The pricing errors which are significant at the 5% level of significance are provided in bold.

In general, the pricing errors across all panels in Table 4 are insignificant and remain relatively small compared to the upper bounds reported in Table 3 with some notable exceptions such as MKT and CMA in Panel B and RMW in Panel C for some values of γ_1.

Risky assets with systematic exposures which are linear combinations of the primitive assets may have higher pricing errors. We check that hypothesis by finding a linear combination that best replicates the corresponding SDF difference $\Delta_M = M_\gamma - M^*_{Missp,\gamma}$ which is done by first solving problem (11) and then by computing the right-hand side of Equation (10) for the asset.

The results are presented in Table 5. Each panel corresponds to one of the four misspecifications and contains the optimal exposures, the pricing error for different values of the risk-aversion parameter γ, and its p-value computed using 500 bootstrapped samples generated according to the algorithm in Section 1.4. The pricing errors resulting from the assumption of constant covariances and constant correlations (Panels B and C) are significant at all levels of γ with one exception. Ignoring autocovariances and assuming the returns are Gaussian (Panels A and D) also leads to significant pricing errors for some levels of risk aversion, but the magnitude of the pricing errors is much smaller compared to Panels B and C indicating smaller economic significance.

From a more general perspective, the results in Table 4 indicate that the basic assets are not severely mispriced by any of the misspecified SDFs. Nevertheless, Table 5 shows that certain portfolios, viewed as linear functions of the basic assets, can be substantially mispriced. For example, ignoring the dynamics of covariances (Panel B) or the dynamics of correlations (Panel C) can lead to a monthly Sharpe ratio overvaluation of about 0.108 for the linear SDF ($\gamma = 1$) of Panel B and about 0.054 for the linear SDF of Panel C. Finally, the maximal Sharpe ratio deviations reported in Table 3 are larger in absolute value as they represent the maximal possible error which materializes for a very particular derivative claim.

The gap between the numbers in Table 5 and Table 3 indicates that larger errors are possible for certain derivative instruments, which can be viewed as certain nonlinear functions of the basic assets, as illustrated by Equation (10). As a consequence, the relatively larger gap between the pricing errors reported in Panel D of Table 5 and the upper bound in Table 3 implies that assuming a Gaussian conditional distribution can have a larger impact on the price of certain derivative instruments. The same observation can be made for the no-autocorrelation misspecification. However, its upper bound in Table 3 is much smaller than that of the no-fat-tails misspecification which limits the impact on the pricing error of all risky assets with either linear or nonlinear exposures to the equity factors.

3.4 Robustness Analysis

Even though it is common to work with monthly returns in empirical asset pricing, we check if the conclusions change in any material way if we use weekly data. As noted previously, the weekly returns exhibit similar empirical properties which are carefully documented by Christoffersen and Langlois (2013) for a similar time period.

The same analysis performed on the weekly returns indicates that largely the same conclusions hold. Like the case of monthly returns, the pricing errors of the basic assets are

Table 4. The pricing errors of the primitive assets computed by Equation (10) using monthly data for different levels of the risk-aversion parameter γ

Panel A: No autocorrelation

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−0.002	−0.002	−0.002	−0.002	−0.002	−0.002	−0.003	−0.002	−0.001	−0.008
SMB	0.000	0.000	−0.001	−0.002	−0.003	−0.004	−0.005	−0.006	−0.005	−0.005
HML	0.011	0.010	0.008	0.005	0.004	0.011	0.010	0.008	0.008	0.013
MOM	0.002	0.002	0.002	0.001	0.001	0.001	0.000	0.000	0.000	0.000
RMW	−0.001	0.000	0.001	0.002	0.002	0.010	0.011	0.011	0.010	0.008
CMA	0.015	0.014	0.011	0.008	0.007	0.014	0.013	0.013	0.012	0.012

Panel B: Constant covariances

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−0.013	−0.003	0.006	0.010	0.011	**0.075**	**0.094**	**0.090**	**0.083**	**0.072**
SMB	0.021	0.024	0.023	0.020	0.018	0.030	0.037	0.033	0.033	0.031
HML	−0.038	−0.025	−0.014	−0.010	−0.008	−0.032	−0.031	−0.039	−0.049	−0.047
MOM	0.057	0.044	0.035	0.033	0.034	0.023	0.026	0.027	0.029	0.029
RMW	0.057	0.048	0.043	0.043	0.043	−0.027	−0.044	−0.045	−0.045	−0.042
CMA	0.027	0.018	0.012	0.011	0.012	−0.041	−0.050	**−0.056**	**−0.061**	**−0.051**

Panel C: Constant correlations

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	0.000	0.000	0.000	0.000	0.000	0.038	0.048	0.037	0.033	0.022
SMB	0.006	0.005	0.004	0.004	0.004	0.003	0.007	0.001	−0.001	−0.002
HML	−0.011	−0.010	−0.009	−0.006	−0.004	−0.020	−0.006	−0.005	−0.004	0.004
MOM	0.045	0.041	0.039	0.037	0.035	0.006	0.005	0.004	0.003	0.002
RMW	**0.059**	**0.057**	**0.050**	0.041	0.035	−0.017	−0.022	−0.015	−0.010	−0.012
CMA	0.010	0.009	0.009	0.011	0.011	−0.014	−0.004	0.005	0.011	0.016

Panel D: The conditional excess returns are Gaussian

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−0.022	−0.018	−0.014	−0.010	−0.008	−0.001	0.000	−0.012	−0.006	−0.009
SMB	−0.008	−0.006	−0.003	−0.001	0.000	0.003	0.003	−0.003	0.000	0.001
HML	0.004	0.004	0.003	0.002	0.002	0.000	−0.001	−0.009	−0.011	−0.009
MOM	−0.059	−0.039	−0.015	−0.002	0.004	0.013	0.014	0.014	0.012	0.013
RMW	0.016	0.013	0.009	0.007	0.006	0.005	0.004	0.012	0.012	0.012
CMA	−0.004	−0.002	0.001	0.002	0.002	0.001	0.000	0.000	−0.004	−0.007

The panels correspond to the four types of misspecification. Negative errors imply that the misspecified SDF M_γ^* overprices the corresponding primitive asset. A linear SDF corresponds to $\gamma = 1$, a lower level of γ indicates a higher degree of risk aversion. The factors considered are the MKT, SMB, HML, MOM, CMA, and RMW factors. The numbers in bold indicate statistical significance at 5%.

Table 5. Linear combinations of factor exposures most exposed to the SDF difference $\Delta_M = M_\gamma - M_\gamma^*$, the corresponding pricing errors, and their p-values

Panel A: No autocorrelation

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−0.089	−0.082	−0.068	−0.052	−0.042	−0.139	−0.133	−0.133	−0.161	0.059
SMB	−0.036	−0.024	0.000	0.031	0.054	0.017	0.057	0.069	0.078	0.046
HML	−0.018	−0.002	0.028	0.048	0.037	−0.028	0.056	0.119	0.142	−0.248
MOM	−0.054	−0.045	−0.034	−0.021	−0.012	−0.017	0.023	0.037	0.039	−0.003
RMW	−0.028	−0.068	−0.095	−0.105	−0.100	−0.515	−0.533	−0.532	−0.508	−0.278
CMA	−0.804	−0.739	−0.635	−0.498	−0.381	−0.757	−0.809	−0.835	−0.837	−0.310
error	−0.016	−0.015	−0.012	−0.009	−0.007	−0.019	−0.018	−0.018	−0.017	−0.016
p-value	0.036	0.028	0.046	0.084	0.176	0.018	0.034	0.052	0.060	0.082

Panel B: Constant covariances

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−0.179	−0.290	−0.434	−0.546	−0.599	−1.525	−1.872	−1.727	−1.473	−1.288
SMB	−1.656	−1.548	−1.352	−1.178	−1.093	−0.146	−0.023	0.114	0.041	−0.059
HML	4.128	2.741	1.719	1.402	1.291	−0.157	−0.909	−0.658	−0.206	0.203
MOM	−0.665	−0.567	−0.505	−0.506	−0.536	−0.773	−1.021	−1.012	−0.973	−0.905
RMW	−3.705	−3.178	−2.822	−2.742	−2.711	0.677	1.495	1.650	1.661	1.492
CMA	−5.791	−4.037	−2.829	−2.569	−2.522	0.930	1.898	2.053	2.076	1.274
error	−0.124	−0.096	−0.079	−0.075	−0.074	−0.084	−0.108	−0.106	−0.102	−0.091
p-value	0.006	0.006	0.004	0.004	0.002	0.002	0.000	0.000	0.000	0.000

Panel C: Constant correlations

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−0.421	−0.401	−0.377	−0.352	−0.329	−0.891	−1.214	−1.068	−1.033	−0.759
SMB	−0.870	−0.812	−0.711	−0.594	−0.516	0.386	0.379	0.417	0.409	0.355
HML	1.315	1.264	1.131	0.948	0.788	0.511	0.043	0.477	0.750	0.456
MOM	−0.751	−0.671	−0.642	−0.643	−0.645	−0.224	−0.295	−0.174	−0.102	−0.104
RMW	−3.180	−3.055	−2.695	−2.221	−1.885	0.528	0.671	0.348	0.124	0.309
CMA	−2.268	−2.151	−2.009	−1.847	−1.672	−0.447	−0.750	−1.484	−2.033	−1.775
error	−0.084	−0.079	−0.071	−0.062	−0.056	−0.043	−0.054	−0.046	−0.046	−0.038
p-value	0.024	0.000	0.000	0.002	0.012	0.026	0.024	0.042	0.044	0.074

Panel D: The conditional excess returns are Gaussian

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	0.697	0.534	0.365	0.240	0.160	−0.019	−0.036	0.227	0.115	0.252
SMB	−0.175	−0.166	−0.141	−0.122	−0.109	−0.139	−0.132	−0.101	−0.191	−0.229
HML	0.410	0.234	0.070	−0.005	−0.032	−0.111	−0.054	0.570	0.593	0.236
MOM	1.552	1.036	0.413	0.083	−0.053	−0.307	−0.334	−0.206	−0.169	−0.228
RMW	−0.922	−0.707	−0.422	−0.283	−0.242	−0.190	−0.185	−0.512	−0.596	−0.489
CMA	0.267	0.225	0.137	0.081	0.051	0.046	0.024	−0.380	−0.349	0.270
error	−0.071	−0.049	−0.024	−0.012	−0.010	−0.015	−0.016	−0.023	−0.021	−0.022
p-value	0.086	0.128	0.190	0.152	0.110	0.048	0.022	0.014	0.008	0.004

Negative errors imply overpricing by the misspecified SDF M_γ^*. The portfolios are computed by Equation (11) and the pricing errors are calculated by Equation (10) using monthly returns. The factors considered are the MKT, SMB, HML, MOM, CMA, and RMW factors.

Table 6. Linear combinations of factor exposures most exposed to the SDF difference $\Delta_M = M_\gamma - M_\gamma^*$, the corresponding pricing errors, and their p-values

Panel A: No autocorrelation

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	0.015	0.004	−0.002	−0.006	−0.008	0.092	0.079	0.070	0.025	−0.285
SMB	−0.050	0.003	0.031	0.041	0.045	0.198	0.180	0.168	0.270	0.433
HML	0.031	0.077	0.100	0.096	0.084	0.328	0.272	0.235	0.179	0.327
MOM	0.154	0.065	0.016	−0.005	−0.015	0.071	0.010	−0.030	0.057	0.036
RMW	−0.289	−0.160	−0.080	−0.042	−0.019	0.614	0.503	0.426	0.171	0.229
CMA	−0.033	−0.122	−0.168	−0.170	−0.156	0.084	0.129	0.158	0.307	−0.188
error	−0.004	−0.002	−0.002	−0.001	−0.001	−0.007	−0.006	−0.005	−0.005	−0.009
p-value	0.210	0.240	0.286	0.292	0.300	0.022	0.044	0.066	0.066	0.048

Panel B: Constant covariances

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−1.371	−1.426	−1.422	−1.417	−1.413	−1.204	−0.786	−0.754	−1.425	−1.876
SMB	−1.804	−1.516	−1.392	−1.323	−1.283	−0.947	−0.320	0.065	0.567	0.805
HML	−0.886	−0.357	−0.166	−0.097	−0.074	0.471	1.550	2.169	2.637	2.840
MOM	−1.022	−0.954	−0.914	−0.908	−0.910	−0.701	−0.273	−0.087	0.007	0.158
RMW	−4.460	−3.727	−3.514	−3.421	−3.370	−2.072	0.490	2.029	3.275	3.211
CMA	−3.680	−3.800	−3.784	−3.760	−3.735	−3.295	−2.440	−2.607	−4.728	−5.326
error	−0.057	−0.051	−0.049	−0.048	−0.047	−0.036	−0.025	−0.033	−0.051	−0.057
p-value	0.002	0.000	0.000	0.000	0.000	0.000	0.016	0.006	0.000	0.000

Panel C: Constant correlations

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	−1.341	−1.051	−0.828	−0.727	−0.682	−0.892	−0.752	−0.660	−0.633	−1.242
SMB	−1.847	−1.378	−1.034	−0.879	−0.800	−1.073	−0.854	−0.706	−0.473	−0.447
HML	−1.052	−0.183	0.301	0.452	0.486	−0.793	−0.859	−0.887	−0.941	−0.659
MOM	−2.573	−1.947	−1.454	−1.154	−1.021	−1.160	−0.970	−0.837	−0.577	−0.338
RMW	−4.012	−2.789	−1.863	−1.451	−1.256	−2.670	−1.881	−1.370	−0.818	−0.419
CMA	−4.664	−4.359	−3.971	−3.755	−3.618	−3.444	−2.811	−2.394	−1.913	−2.656
error	−0.075	−0.058	−0.045	−0.039	−0.036	−0.046	−0.038	−0.032	−0.026	−0.031
p-value	0.000	0.000	0.000	0.004	0.006	0.000	0.000	0.000	0.000	0.000

Panel D: The conditional excess returns are Gaussian

	−1.5	−1	−0.5	0	0.4	0.8	1	1.2	1.6	2
MKT	0.124	0.097	0.069	0.040	0.016	0.111	0.114	0.145	0.200	0.176
SMB	0.108	0.072	0.056	0.043	0.032	0.196	0.194	0.038	−0.312	−0.277
HML	−0.243	−0.203	−0.186	−0.198	−0.213	0.075	0.057	0.055	0.060	0.005
MOM	−0.030	−0.052	−0.081	−0.111	−0.133	−0.023	0.004	−0.148	−0.485	−0.418
RMW	−0.335	−0.226	−0.148	−0.123	−0.116	0.596	0.568	0.735	1.022	0.949
CMA	−0.118	−0.123	−0.111	−0.093	−0.080	0.159	0.151	0.016	−0.222	−0.167
error	−0.007	−0.005	−0.004	−0.004	−0.004	−0.005	−0.005	−0.007	−0.014	−0.013
p-value	0.048	0.036	0.064	0.060	0.060	0.042	0.038	0.088	0.002	0.000

Negative errors imply overpricing by the misspecified SDF M_γ^*. The portfolios are computed by Equation (11) and the pricing errors are calculated by Equation (10) using weekly returns. The factors considered are the MKT, SMB, HML, MOM, CMA, and RMW factors.

insignificant with similar exceptions. We provide the optimal exposures from Equation (11), their pricing error, and the corresponding p-values in Table 6.

One difference from the case of monthly returns is the higher relative importance of the dynamic of correlations and the slight increase in the statistical significance in the pricing error resulting from ignoring the fat tails. We note that it is difficult to compare the pricing errors across frequencies, for example, Tables 5 and 6. The reason is that the pricing errors are interpreted in terms of Sharpe ratios which are not easily transformed across frequencies with the assumed dynamics in Equation (8).

The similarity across the weekly and monthly frequency results indicates that the conclusions are robust rather than strictly characteristic of a given return frequency.

4 Conclusions

In empirical equity asset pricing, the SDF is commonly assumed to be a linear function of a certain set of primitive assets, also known as equity factors. The empirical properties of the factors, therefore, have an influence on the properties of the SDF. Ignoring certain empirical characteristics leads to a misspecified SDF which may introduce a pricing error.

Even though many factors have been suggested in the empirical literature, we consider the five factors proposed by Fama and French (2005) and the momentum factor proposed by Carhart (1997). We investigate the impact of four different misspecifications related to well-known stylized facts for monthly returns: autocorrelation, dynamics of the covariances, correlation dynamics, and fat tails of the conditional factor return distribution. The asset pricing model is based on the nonlinear formulation by Almeida and Garcia (2016) and the pricing error is evaluated by means of the upper bound of Hansen and Jagannathan (1997).

We find that correlation dynamics is of first-order significance followed by volatility dynamics. In contrast, the misspecification resulting from assuming no autocorrelations leads to much smaller pricing errors. Finally, assuming the conditional distribution of the factor returns is Gaussian results in small pricing errors for near-linear SDFs but can lead to larger pricing errors for higher levels of risk aversion.

A misspecified SDF can misprice both the primitive assets and all derivative claims. We find that the pricing errors of the factors on a stand-alone basis are insignificant, with some notable exceptions. However, equity portfolios whose systematic risk is a linear combination of the five Fama–French and the momentum factors can be significantly overpriced by the misspecified SDF. For near-linear SDFs, this holds for all misspecifications but the pricing error is economically more significant for the constant correlation and the constant covariance misspecifications.

For SDFs with high risk aversion, the pricing errors of linear portfolios resulting from neglecting fat tails tend to increase but become insignificant. Our analysis, however, implies that this misspecification can have a more significant impact on certain derivative claims with nonlinear exposures to the equity factors.

Our results complement Christoffersen and Langlois (2013) and emphasize the importance of volatility and correlation dynamics in research in conditional asset pricing models, such as Engle (2016) and Bali, Engle, and Tang (2016).

References

Aielli, G. 2013. Dynamic Conditional Correlation: On Properties and Estimation. *Journal of Business & Economic Statistics* 31: 282–299.

Almeida, C., K. Ardison, and R. Garcia. 2020. Nonparametric Assessment of Hedge Fund Performance. *Journal of Econometrics* 214: 349–378.

Almeida, C., and R. Garcia. 2012. Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators. *Journal of Econometrics* 170: 519–537.

Almeida, C., and R. Garcia. 2016. Economic Implications of Nonlinear Pricing Kernels. *Management Science* 63: 3361–3380.

Bali, T., R. Engle, and Y. Tang. 2016. Dynamic Conditional Beta Is Alive and Well in the Cross Section of Daily Stock Returns. *Management Science* 63: 3760–3779.

Campbell, J. 2000. Asset Pricing at the Millennium. *The Journal of Finance* 55: 1515–1567.

Carhart, M. 1997. On Persistence in Mutual Fund Performance. *The Journal of Finance* 52: 57–82.

Černy, A. 2003. Generalized Sharpe Ratios and Asset Pricing in Incomplete Markets. *Review of Finance* 7: 191–233.

Christoffersen, P., and H. Langlois. 2013. The Joint Dynamics of Equity Market Factors. *Journal of Financial and Quantitative Analysis* 48: 1371–1404.

Cochrane, J. 2005. *Asset Pricing*. Princeton University Press, Princeton, NJ.

Cressie, N., and T. Read. 1984. Multinomial Goodness-of-Fit Tests. *Journal of the Royal Statistical Society: Series B (Methodological)* 46: 440–464.

Elie, L., and T. Jeantheau. 1995. Consistency in Heteroskedastic Models. *Comptes Rendus de L'Académie Des Sciences* 320: 1255–1258.

Engle, R. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 20: 339–350.

Engle, R. 2016. Dynamic Conditional Beta. *Journal of Financial Econometrics* 14: 643–667.

Fama, E., and F. French. 2005. A Five-Factor Asset Pricing Model. *Journal of Financial Economics* 116: 1–22.

Flannery, M., and A. Protopapadakis. 2002. Macroeconomic Factors Do Influence Aggregate Stock Returns. *Review of Financial Studies* 15: 751–782.

Hall, P., and Q. Yao. 2003. Inference in ARCH and GARCH Models with Heavy–Tailed Errors. *Econometrica* 71: 285–317.

Hansen, L., and R. Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99: 225–262.

Hansen, L., and R. Jagannathan. 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *The Journal of Finance* 52: 557–590.

Horv, L., and P. Kokoszka. 2003. GARCH Processes: Structure and Estimation. *Bernoulli* 9: 201–227.

Lo, A. 2002. The Statistics of Sharpe Ratios. *Financial Analysts Journal* 58: 36–52.

Shapiro, S., and M. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52: 591–611.

Snow, K. 1991. Diagnosing Asset Pricing Models Using the Distribution of Asset Returns. *The Journal of Finance* 46: 955–983.

Toda, A., and K. Walsh. 2017. Fat Tails and Spurious Estimation of Consumption-Based Asset Pricing Models. *Journal of Applied Econometrics* 32: 1156–1177.

Journal of Financial Econometrics, 2021, Vol. 19, No. 1, 202–234
doi: 10.1093/jjfinec/nbaa026
Advance Access Publication Date: 14 October 2020
Article

Local-Linear Estimation of Time-Varying-Parameter GARCH Models and Associated Risk Measures

Atsushi Inoue[1], Lu Jin[2,*], and Denis Pelletier[3]

[1]Vanderbilt University, [2]Bank of America and [3]North Carolina State University

Address correspondence to Denis Pelletier, Department of Economics, North Carolina State University, Box 8110 Raleigh, North Carolina 27695, or e-mail: denis_pelletier@ncsu.edu.

Received August 30, 2019; revised June 29, 2020; editorial decision July 1, 2020; accepted July 1, 2020

Abstract

In this article, we propose a nonparametric approach to estimating generalized autoregressive conditional heteroskedasticity (1,1) models with time-varying parameters. We model the time-varying parameters as a smooth function of time and estimate them using a local linear estimator. We show that our estimator is consistent and is asymptotically normal and that the proposed estimator outperforms a rolling window estimator in Monte Carlo simulation experiments. We present strong evidence of parameter instabilities using daily returns of stock indices and explore implications to risk management measures, such as value-at-risk and expected shortfall, through backtesting.

Key words: time-varying parameters, expected shortfall, value-at-risk, realized volatility

JEL classification: C14, C51, C58

Since the seminal work of Engle (1982) and Bollerslev (1986), the generalized autoregressive conditional heteroskedasticity (GARCH) model has become a workhorse in empirical finance. Yet, there is strong evidence of parameter instability in GARCH models. For example, Mikosch and Stărică (2004) and Rapach and Strauss (2008) find evidence of structural change in the GARCH parameters in S&P500 return volatility and in exchange rate return volatility, respectively. Parameter instability may not be captured by a one-time break model; Andreou and Ghysels (2002) find multiple breaks in stock return volatility series. The presence of regime switching in the GARCH parameters (Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996; Bauwens et al., 2010; Kim and Hwang, 2018) can also be thought of as evidence for multiple structural changes. Parameter instability may also exist in multivariate GARCH models (Jin and Maheu, 2016).

* The opinions expressed here are those of the authors do not reflect the views of Bank of America.

When it is not accounted for, it is known that the estimated GARCH process appears more persistent than it actually is, as pointed out by Diebold (1986), Lamoureux and Lastrapes (1990), and Hillebrand (2005) to name a few. Such upward biases in the GARCH parameter estimates have undesirable consequences. For example, West and Cho (1995) and Rapach and Strauss (2008) find that allowing for change in the unconditional variance improve the performance of GARCH models to forecast the exchange rate volatility. This has important implications for risk management, such as value-at-risk (VaR) and expected shortfall (ES). Under the normality assumption, these risk measures are proportional to the conditional variance. When a parameter break is not taken into account, this means that the volatility forecast is more tied to the current volatility (observed or latent) than it should be, which may lead to under- or over-estimation of financial risk. In a multivariate context, changes in the data-generating process (DGP) of returns not captured by the model will affect the accuracy of VaR and ES of a portfolio (Berens et al., 2015), the computation of hedging ratios (Ewing and Malik, 2013). Pettenuzzo and Timmermann (2011) also show that breaks in models used for return predictions can have a large impact on portfolio allocation.

An important part of the literature on models with parameters changing over time focuses on structural breaks. However, as argued for example in Chen and Hong (2016) and Hansen (2001) smoothly changing parameters are more likely than structural breaks. The impact of economic events (trade agreements, change in interest rate policies by a central bank) or financial market participants (algorithmic trading) could evolve over a period of time, not instantaneously. This increases the interest in models allowing parameters to be time-varying. For example, in macroeconomics time-varying parameter VAR models such as in Primiceri (2005) and Nakajima (2011), where parameters follow a random walk, are quite popular.

In this article, we propose to estimate time-varying-parameter GARCH(1,1) models. Although financial news may be captured by heavy-tailed distributions or by jumps, institutional changes may be better captured by time-varying parameters. We model the GARCH parameters as smooth functions of time and estimate them by a local linear estimator. Specifically, we compute local linear estimates by maximizing a local-quasi-log likelihood function. We show that the local linear estimator is consistent and is asymptotically normally distributed.

This article contributes to the literature on non-parametric estimation of time-varying parameter models. To the best of our knowledge, Robinson (1989) is the first to consider nonparametric estimation of linear regression models in which parameters are modeled as deterministic functions of time. Cai (2007) develops local linear estimators of such models. Viewing rolling-window estimators as a local constant estimator with the uniform kernel, Inoue et al. (2017) develop a method for selecting the window size for out-of-sample forecasting. Chen and Hong (2012, 2016) propose non-parametric tests of structural change in linear regression and GARCH models, respectively. In particular, Chen and Hong (2016) uses a local constant estimator to allow for time-varying parameter GARCH(p, q) models under the alternative hypothesis. Dahlhaus and Rao (2006) develop a local constant estimator for ARCH(∞) models with time-varying parameters. Fryzlewicz et al. (2008) develop a kernel normalized-least-squares locally constant estimator of ARCH(p) models by using the (linear) AR(p) representation for the squared return. Rohan and Ramanathan (2013) and Rohan (2013) further extend this approach to GARCH(p, q) models by using the (linear)

autoregressive moving average (ARMA) representation of the squared return. The estimation is performed in two steps: first, a local polynomial estimation of a long auto-regression for the squared return is performed, then latent variances are replaced by predicted values from the first stage and a second local polynomial estimation is performed. This is a non-parametric version of a method introduced by Hannan and Rissanen (1982). Hafner and Linton (2010) consider non-parametric estimation of multivariate GARCH models in which the unconditional variance–covariance matrix is modeled as a deterministic function of time. Kristensen and Lee (2019) develop local polynomial estimators for a general class of dynamic models, such as Markov-type and generalized autoregressive models.[1] Other approaches to allowing for changes in the unconditional variance of GARCH models include one based on spline functions (Engle and Rangel, 2008) and another on logistic functions (Amado and Teräsvirta, 2013; Amado and Terásvirta, 2014).

Unlike the estimators of Rohan and Ramanathan (2013) and Rohan (2013) that require choosing the order of an autoregressive process for the squared return and selecting a bandwidth in each of the two steps, our estimator is one-shot and requires selecting only one bandwidth because it is the local-quasi-log likelihood function of a GARCH(1,1) model. While Rohan and Ramanathan (2013), Rohan (2013), and Kristensen and Lee (2019) focus on the consistency of intercept estimators of local polynomial approximations, we provide sufficient conditions for the consistency of slope estimators in local linear approximations in addition to the consistency of intercept estimators. The slopes may be useful for long-horizon volatility forecasts based on time-varying parameter GARCH models.

GARCH models allowing time-varying parameters can be quite interesting for risk management applications. On one hand, it is tempting to use non-parametric estimators to compute risk measures, like computing VaR with Historical Simulations (HS). HS is the sample quantile of returns from a rolling window. However, as argued by Pritsker (2006), HS is under-responsive to risk. Christoffersen and Pelletier (2004) and Berkowitz et al. (2011) have shown that we can develop backtests to detect that VaR computed with HS lead to predictable violations. An improvement on HS is to use Filtered HS where we apply HS on returns standardized by a volatility model like GARCH. This approach is more flexible, however, could be further improved if applied in the time-varying parameter GARCH framework we consider. Our simulation and empirical results suggest that our local-linear (LL) estimator can give better estimates of parameters and forecasts at longer horizons. At shorter horizons, it remains difficult to out-perform the forecasts from a model with constant parameters estimated on a window.

The rest of the article is organized as follows: In Section 1, we present our model. In Section 2, we discuss our assumptions and theoretical results. Results of Monte Carlo simulations illustrating the performance of our LL estimator and our bandwidth selection procedure are in Section 3. An empirical application to daily returns on a set of indices is in Section 4. Section 5 presents some concluding remarks. Proofs of the theorems are in the Appendix A. The lemmas and their proofs, as well as a presentation of the backtesting methods employed in the empirical section are in the online Appendix.

1 For time-varying-parameter GARCH models, they consider local KLs. We focus on a local linear version of QMLE of GARCH(1, 1) models.

1 Model Specification

In this section, we consider a GARCH(1,1) process with time-varying parameters. As in Robinson (1989), Cai (2007), and Inoue et al. (2017), the time variation in the parameters is represented by a smooth function of time, which is defined on an equally spaced grid over [0,1]. The model setup for GARCH(1,1) can be rewritten as:

$$y_t = \sqrt{h_t}\epsilon_t, \tag{1}$$

where

$$h_t = \omega\left(\frac{t-1}{T}\right) + \alpha\left(\frac{t-1}{T}\right)y_{t-1}^2 + \beta\left(\frac{t-1}{T}\right)h_{t-1}, \tag{2}$$

$\omega(\cdot)$, $\alpha(\cdot)$, and $\beta(\cdot)$ in (2) are smooth functions of time satisfying $\omega(u), \alpha(u), \beta(u) > 0$, and $0 < \alpha(u) + \beta(u) < 1$ for all $u \in [0,1]$, and ϵ_t is independent and identically distributed (i.i.d.) with zero mean and unit variance. We will state assumptions in a more precise manner in the next section. We model $\omega(\cdot)$, $\alpha(\cdot)$ and $\beta(\cdot)$ as a function of $(t-1/T)$ rather than (t/T) to coincide with h_t being conditional on the information set at time $t-1$.

Let $u_0 \in [0,1]$ be the scaled time of interest. For example, if one is to forecast volatility at the end of the sample, one may be interested in $\omega(1)$, $\alpha(1)$, and $\beta(1)$ and thus $u_0 = 1$. To estimate $\omega(u_0)$, $\alpha(u_0)$, and $\beta(u_0)$, we use local linear approximations of the time-varying parameters to construct a quasi-likelihood function:

$$\bar{\omega}\left(\frac{t-1}{T}, u_0\right) = \omega_0 + \omega_1\left(\frac{t-1}{T} - u_0\right), \tag{3}$$

$$\bar{\alpha}\left(\frac{t-1}{T}, u_0\right) = \alpha_0 + \alpha_1\left(\frac{t-1}{T} - u_0\right), \tag{4}$$

$$\bar{\beta}\left(\frac{t-1}{T}, u_0\right) = \beta_0 + \beta_1\left(\frac{t-1}{T} - u_0\right), \tag{5}$$

where $\alpha_0 = \alpha(u_0)$, $\alpha_1 = \alpha^{(1)}(u_0)$, $\beta_0 = \beta(u_0)$, $\beta_1 = \beta^{(1)}(u_0)$, $\omega_0 = \omega(u_0)$, and $\omega_1 = \omega^{(1)}(u_0)$.

Our goal is to estimate α_0, α_1, β_0, β_1, ω_0, and ω_1 in Equations (3)–(5). Because α_0, β_0, and ω_0 are $\alpha(u_0)$, $\beta(u_0)$, and $\omega(u_0)$, respectively, in Equation (2), they are often the parameters of interest. However, α_1, β_1, and ω_1 in Equations (3)–(5) may be useful for long-horizon forecasting and testing. Let, $\omega_0^*, \omega_1^*, \alpha_0^*, \alpha_1^*, \beta_0^*, \beta_1^*$ denote the true parameter values of $\omega_0, \omega_1, \alpha_0, \alpha_1, \beta_0, \beta_1$, respectively, $\theta = (\omega_0, \omega_1, \alpha_0, \alpha_1, \beta_0, \beta_1)^\top$, and $\theta^* = (\omega_0^*, \omega_1^*, \alpha_0^*, \alpha_1^*, \beta_0^*, \beta_1^*)^\top$. The estimator of θ considered in this section is the maximizer of a quasi-log-likelihood function that is locally weighted around scaled time point u_0.

$$\tilde{L}_{R,T}(u_0, \theta) = \sum_{s=2}^{T} \frac{1}{R} \cdot k\left(\frac{s-t_0}{R}\right) \cdot \tilde{l}_{s,T}(u_0, \theta), \tag{6}$$

where $k(\cdot)$ is a kernel function, R is a bandwidth and

$$\tilde{l}_{s,T}(u_0, \theta) = -\frac{1}{2}\left(\frac{y_s^2}{\tilde{h}_s(u_0, \theta)} + \ln \tilde{h}_s(u_0, \theta)\right), \tag{7}$$

$$\tilde{h}_s(u_0,\theta)=\bar{\omega}\left(\frac{s-1}{T},u_0\right)+\bar{\alpha}\left(\frac{s-1}{T},u_0\right)y_{s-1}^2+\bar{\beta}\left(\frac{s-1}{T},u_0\right)\tilde{h}_{s-1}(u_0,\theta), \tag{8}$$

$\bar{\omega}(\frac{s-1}{T},u_0)$, $\bar{\alpha}(\frac{s-1}{T},u_0)$ and $\bar{\beta}(\frac{s-1}{T},u_0)$ are defined in Equations (3)–(5), respectively.

If, $u_0\in(0,1)$, then define interior time points $t_0=[u_0T]+1$. If, $u_0=0$, let $t_0-1=cR$ for some $c\in[0,1]$ represent the left boundary time points. Similarly, if $u_0=1$, let $t_0-1=T-cR$ represent the right boundary time points.[2]

To derive the consistency and the asymptotic properties of the proposed estimator, we consider the log-likelihood function of the constant-parameter Gaussian GARCH model whose true parameter values are given by α_0, β_0, and ω_0:

$$\ddot{L}_{R,T}(u_0,\vartheta)=\sum_{s=2}^{T}\frac{1}{R}\cdot k\left(\frac{s-t_0}{R}\right)\cdot\ddot{l}_{s,T}(u_0,\vartheta), \tag{9}$$

where

$$\ddot{l}_{s,T}(u_0,\vartheta)=-\frac{1}{2}\left(\frac{\ddot{y}_s^2(u_0)}{\ddot{h}_s(u_0,\vartheta)}+\ln\ddot{h}_s(u_0,\vartheta)\right), \tag{10}$$

$$\ddot{h}_s(u_0,\vartheta)=\omega_0+\alpha_0\ddot{y}_{s-1}^2(u_0)+\beta_0\ddot{h}_{s-1}(u_0,\vartheta), \tag{11}$$

and $\ddot{y}_s(u_0)$ follows

$$\ddot{y}_t(u_0)=\sqrt{\ddot{h}_s(u_0,\vartheta^*)}\epsilon_t, \tag{12}$$

where $\vartheta^*=(\theta_1^*,\theta_3^*,\theta_5^*)^\top=(\omega_0^*,\alpha_0^*,\beta_0^*)^\top$. The constant-parameter GARCH process $\{\ddot{y}_t(u_0)\}$ and its log-likelihood function are simply a device to help develop asymptotic theory for the proposed estimator. The process is not observed and is not used in the estimation procedure. We locally approximate the true conditional variance h_t in (2) by its locally linearized version $\tilde{h}_t(u_0,\theta)$ in (8), which in turn we locally approximate by $\ddot{h}_t(u_0,\vartheta)$ in (11).

2 Asymptotic Theory

To derive the consistency and asymptotic normality of our estimator, we impose the following conditions. Note that C denotes a generic finite positive constant.

Assumption 1 *Parameter functions $\omega(\cdot)$, $\alpha(\cdot)$ and $\beta(\cdot)$ map $[-\eta,1+\eta]$ to $\Re^+$ and are twice continuously differentiable on $(-\eta,1+\eta)$ for some $\eta>0$.*

Assumption 2 *Parameter functions $\omega(\cdot)$, $\alpha(\cdot)$, and $\beta(\cdot)$ are positive and are uniformly bounded above: There exist constants $0<\kappa\le K<\infty$ and $0<\delta\le\rho<1$ such that $\kappa\le\omega(u)\le K$, $\alpha(u)\ge\delta$ and $\beta(u)\ge 0$, and $\alpha(u)+\beta(u)\le\rho$ for all $u\in[0,1]$.*

Assumption 3 *The parameter space of θ, denoted by Θ, is a compact subset of $\mathbb{R}^6$. The population parameter value of θ, θ^*, belongs to the interior of the parameter space Θ.*

2 The choice of c is not unique. For example, c can be set to zero if one is concerned about bias while it can be set to one if one is concerned about the asymptotic variance.

Assumption 4 *(a)* $h_{t_0-R} > 0$ *and* $E(h_{t_0-R}) < \infty$*; (b)* $\tilde{h}_{t_0-R}(u_0,\theta)$ *is three times continuously differentiable in* θ*.* $\tilde{h}_{t_0-R}(u_0,\theta) > 0$ *and* $E[\sup_{\theta\in\Theta}|\tilde{h}_{t_0-R}(u_0,\theta)|] \le C$*; (c)* $\ddot{h}_0(u_0,\vartheta)$ *is drawn from the invariant distribution.*

Assumption 5 $E\left[\sup_{\theta\in\Theta}|\frac{\partial\tilde{h}_{t_0-R}(u_0,\theta)}{\partial\theta_i}|\right] \le C$ for $i=1, 3, 5$ and $E\left[\sup_{\theta\in\Theta}|\frac{T\partial\tilde{h}_{t_0-R}(u_0,\theta)}{R\partial\theta_i}|\right] \le C$ for i = 2, 4, 6, where $\theta_1=\omega_0$, $\theta_2=\omega_1$, $\theta_3=\alpha_0$, $\theta_4=\alpha_1$, $\theta_5=\beta_0$, and $\theta_6=\beta_1$.

Assumption 6 *For* $i=1, 3, 5$ *and* $j=1, 3, 5$ $E\left[\sup_{\theta\in\Theta}|\frac{\partial^2\tilde{h}_{t_0-R}(u_0,\theta)}{\partial\theta_i\partial\theta_j}|\right] \le C$.

Assumption 7 *For* $i=1, 3, 5$, $j=1, 3, 5$ *and* $k=1, 3, 5$, $E\left[\sup_{\theta\in\Theta}|\frac{\partial^3\tilde{h}_{t_0-R}(u_0,\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}|\right] \le C$.

Assumption 8 *Weight function,* $k:[-1,1]\to\Re^+$*, is bounded and satisfies* $\int_{-1}^{1}k(u)du = 1$, $\int_{-1}^{1}uk(u)du = 0$, $\int_{-1}^{1}k(u)^2du < \infty$ *and* $\int_{-1}^{1}u^2k(u)du < \infty$.

Assumption 9 *As T goes to infinity, R diverges to infinity satisfying* $R = O(T^{\frac{4}{5}})$.

Assumption 10 ϵ_t *is independent and identically distributed with zero mean and unity variance, and* $E[\epsilon_t^{4+\delta}] < \infty$ *for some* $\delta > 0$.

Remarks:

1. The assumption that the parameters are smooth function of time in Assumption 1 is standard for nonparametric estimators of time varying parameters.
2. Once we assume that $\omega(\cdot)$, $\alpha(\cdot)$, and $\beta(\cdot)$ are continuously differentiable over the compact interval $[0,1]$ as in Assumption 1, the compactness of Θ in Assumption 3 follows.
3. We derive the asymptotic distribution of our estimator using the first order conditions. To work with the first-order conditions, we assume that θ^* belongs to the interior of Θ. Together with Assumption 2, this assumption implies that $\omega(u_0) > 0$, $\alpha(u_0) > 0$ and $\beta(u_0) > 0$.
4. Assumption 4(a) and 4(b) are high-level conditions and guarantee that $h_t > 0$ and $\tilde{h}_t > 0$ hold for $t = t_0 - R, t_0 - R + 1, \ldots, t_0 + R$.
5. The derivatives in the second inequality in Assumption 5 are the derivatives with respect to the slope parameters ω_1, α_1, and β_1 and thus converge to zero at rate R/T over the estimation window. These derivatives are multiplied by T/R because they converge to zero otherwise.
6. Assumption 4(a) can be shown to be satisfied under Assumption 2 and the assumption that $h_0 > 0$ and $E(h_0) < \infty$, for example. Assumptions 4(b), 5, and 6 are satisfied if $\tilde{h}_{t_0-R}(u_0,\theta)$ is set to the unconditional variance given θ, for example.
7. As an example, consider the case in which the parameters are linear functions of time, that is, $\omega(t/T) = \bar{\omega}_0 + \omega_1(t/T)$, $\alpha(t/T) = \bar{\alpha}_0 + \alpha_1(t/T)$, $\beta(t/T) = \bar{\beta}_0 + \beta_1(t/T)$, and $\varepsilon_t \overset{\text{iid}}{\sim} N(0,1)$, where $\bar{\omega}_0, \bar{\alpha}_0, \bar{\beta}_0 > 0, \bar{\omega}_0 + \omega_1, \bar{\alpha}_0 + \alpha_1, \bar{\beta}_0 + \beta_1$ are all positive, and $\bar{\alpha}_0 + \bar{\beta}_0$ and $\bar{\alpha}_0 + \alpha_1 + \bar{\beta}_0 + \beta_1$ are strictly smaller than one. In addition, assume that $h_0 > 0$, $E(\sup_{\theta\in\Theta}h_0(\theta)) < \infty$, $E(y_0^2) < \infty$, $E(\sup_{\theta\in\Theta}|\partial h_0(\theta)/\partial\theta_i|) < \infty$ for $i = 1, 2, 3, 4, 5, 6$, $E(\sup_{\theta\in\Theta}|\partial^2 h_0(\theta)/\partial\theta_i\partial\theta_j|) < \infty$ for $i, j = 1, 2, 3, 4, 5, 6$, and $E(\sup_{\theta\in\Theta}|\partial^3 h_0(\theta)/\partial\theta_i\partial\theta_j\partial k|) < \infty$ for $i, j, k = 1, 2, 3, 4, 5, 6$.

Under the constraints on $\bar{\alpha}_0, \alpha_1, \bar{\beta}_0, \beta_1, \bar{\omega}_0, \omega_1$, Assumption 2 is satisfied. It follows from recursive substitutions that $h_t > 0$ and $E(y_t^2) < \infty$, and $E(h_t) < \infty$ for all $t = 1, 2, \ldots, T$ (which also implies $\tilde{h}_t(u_0, \theta^*) > 0$ and $E(\tilde{h}_t(u_0, \theta^*) < \infty$ because of the linear specification). By mathematical induction, $E(\sup_{\theta \in \Theta} \tilde{h}_t(u_0, \theta)) < \infty$ for all $t = 1, 2, \ldots, T$. Thus, Assumption 4 is satisfied. Assumptions 5, 6, and 7 follow by using the arguments similar to the ones used in Lemmas 3, 4, and 5 in the online Appendix.

8. To simplify the proof, we assume that the kernel function has a bounded support as in Assumption 8.
9. As we increase R, the bias of our estimator increases while its variance decreases. The condition $R = O(T^{4/5})$ guarantees that the bias is not too large.

By abuse of notation, define $\nabla \ddot{h}_s(u_0, \vartheta) = (\partial \ddot{h}_s(u_0, \vartheta)/\partial \omega_1, \partial \ddot{h}_s(u_0, \vartheta)/\partial \alpha_1, \partial \ddot{h}_s(u_0, \vartheta)/\partial \beta_1)^\top$ where

$$\frac{\partial \ddot{h}_s(u_0, \vartheta)}{\partial \omega_1} = \sum_{j=1}^{s-t_0+R} \left(\frac{s - t_0 - (j-1)}{R} \right) \beta_0^{j-1}, \tag{13}$$

$$\frac{\partial \ddot{h}_s(u_0, \vartheta)}{\partial \alpha_1} = \sum_{j=1}^{s-t_0+R} \left(\frac{s - t_0 - (j-1)}{R} \right) \beta_0^{j-1} \ddot{y}_{s-j}(u_0)^2, \tag{14}$$

$$\frac{\partial \ddot{h}_s(u_0, \vartheta)}{\partial \beta_1} = \sum_{j=2}^{s-t_0+R} \sum_{k=1}^{j-1} \left(\frac{s - t_0 - (k-1)}{R} \right) \beta_0^{j-2} \ddot{c}_{s-j}(u_0) + \sum_{k=1}^{s-t_0+R} \left(\frac{s - t_0 - (k-1)}{R} \right) \beta_0^{s-t_0+R-1} \ddot{h}_{t_0-R}(u_0, \vartheta). \tag{15}$$

and $\ddot{c}_t(u_0) = \omega_0 + \alpha_0 \ddot{y}_t^2(u_0)$.

These are not partial derivatives in the conventional sense and are introduced to approximate $\partial \tilde{h}_s(u_0, \theta)/\partial \omega_1$, $\partial \tilde{h}_s(u_0, \theta)/\partial \alpha_1$, and $\partial \tilde{h}_s(u_0, \theta)/\partial \beta_1$ by functions of the stationary process $\ddot{y}_s(u_0)$.

In the theorem below, we first show that the estimator $\widehat{\theta}_{R,T} = \text{argmax}_{\theta \in \Theta} \tilde{L}_{R,T}(u_0, \theta)$ is a consistent estimator of θ.

Theorem 1 *Let* $\widehat{\theta}_{R,T} = (\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3, \widehat{\theta}_4, \widehat{\theta}_5, \widehat{\theta}_6)^\top = \text{argmax}_{\theta \in \Theta} \tilde{L}_{R,T}(u_0, \theta)$, $\widehat{\vartheta} = (\widehat{\omega}_0, \widehat{\alpha}_0, \widehat{\beta}_0)^\top$, $\widehat{\boldsymbol{\theta}} = (\widehat{\omega}_1, \widehat{\alpha}_1, \widehat{\beta}_1)^\top$, $\theta^* = (\omega_0^*, \omega_1^*, \alpha_0^*, \alpha_1^*, \beta_0^*, \beta_1^*)^\top$, $\vartheta^* = (\omega_0^*, \alpha_0^*, \beta_0^*)^\top$, and $\boldsymbol{\theta}^* = (\omega_1^*, \alpha_1^*, \beta_1^*)^\top$. *In other words,* $\widehat{\vartheta}$ *consists of the constant term estimates in* $\widehat{\theta}_{R,T}$ *while* $\widehat{\boldsymbol{\theta}}$ *consists of the slope coefficient estimates in* $\widehat{\theta}_{R,T}$.

Suppose that Assumptions 1–10 hold. Then:

a. $\widehat{\vartheta}_{R,T} \xrightarrow{p} \vartheta^*$.
b. *In addition, if* $R/T^{\frac{2}{3}} \to \infty$ *and if*

$$E\left[\frac{\nabla \ddot{h}_s(u_0, \vartheta^*) \nabla \ddot{h}_s(u_0, \vartheta^*)^\top}{\ddot{h}_s(u_0, \vartheta^*)^2} \right]$$

is positive definite, then $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$.

Part (a) of the theorem shows that we can estimate $\omega(u_0)$, $\alpha(u_0)$ and $\beta(u_0)$ in Equation (2) consistently. Part (b) provides sufficient conditions for consistent estimation of the slope coefficients, ω_1, α_1, and β_1 in Equations (3)–(5). Because the dependence of the quasi-maximum likelihood estimation (QMLE) criterion function on $\boldsymbol{\theta}$ vanishes asymptotically, we

show that $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ is the only solution to the probability limit of the scaled and linearized first order conditions for $\widehat{\boldsymbol{\theta}}$. The two additional assumptions make that feasible.

Next, we show the asymptotic distribution of our estimator under Assumption 9. In the derivation of the asymptotic properties of this estimator, we leverage a local stationary approximation $\ddot{y}_s^2(u_0)$ and focus on the asymptotic distribution of a subvector of the parameters.

Theorem 2 *As in Theorem 1, let ϑ, ϑ^* and $\widehat{\vartheta}_{R,T}$ denote the subvectors that consist of the first, third and fifth elements of θ, θ^* and $\widehat{\theta}_{R,T}$, respectively. Also let $\nabla = \frac{\partial}{\partial\theta} = \left(\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_3}, \frac{\partial}{\partial\theta_5}\right)^\top$. Under Assumptions 1–10,*

$$\sqrt{R}\left(\widehat{\vartheta}_{R,T} - \vartheta^* + \left(\frac{R}{T}\right)^2 \frac{\kappa_2}{2}\boldsymbol{\theta}^*\right) \underline{d} N\left(0, \frac{k_2}{k_1^2}(E\left[\epsilon_s^4\right] - 1)[\ddot{B}(\vartheta^*)]^{-1}\right), \tag{16}$$

where $k_i = \int_{-1}^{1} k^i(u)\mathrm{du}$ if $u_0 \in (0,1)$, $k_i = \int_{-c}^{1} k^i(u)\mathrm{du}$ *if* $u_0 = 0$, $k_i = \int_{-1}^{c} k^i(u)\mathrm{du}$ if $u_0 = 1$, $\kappa_2 = \int_{-1}^{1} u^2 k(u)\mathrm{du}$ if $u_0 \in (0,1)$, $\kappa_2 = \int_{-c}^{1} u^2 k(u)\mathrm{du}$ if $u_0 = 0$, $\kappa_2 = \int_{-1}^{c} u^2 k(u)\mathrm{du}$ if $u_0 = 1$, *c is the constant defined just after Equation (8)*, and $\ddot{B}(\vartheta^*) = E[\nabla\ddot{h}_s(u_0,\vartheta^*)\nabla\ddot{h}_s(u_0,\vartheta^*)^\top/\ddot{h}_s^2(u_0,\vartheta^*)]$.

Since $\ddot{B}\left(\vartheta^*(u_0)\right)$ is defined based on a stationary GARCH(1,1) model structure, it depends on three parameters, that include θ_1, θ_3, θ_5. The derivation of the asymptotic properties of parameters θ_2, θ_4, θ_6 would need an analog of the term $\ddot{B}\left(\vartheta^*(u_0)\right)$ that depends on all six parameters in θ. In other words, we need to approximate a nonstationary GARCH(1,1) model with smooth time-varying parameters by a non-stationary GARCH(1,1) model with the local linear representations in parameters. This would further require more sophisticated properties on the dependence structure in time-varying GARCH(1,1) models.

Lastly, we provide a consistent estimator of the asymptotic covariance matrix in Theorem 2:

Theorem 3 *Given the estimator $\widehat{\theta}_{R,T}$, we can construct the fitted conditional variance as*

$$\widehat{\ddot{h}}_s\left(u_0, \widehat{\theta}_{R,T}\right) = \widehat{\ddot{\omega}}\left(\frac{s-1}{T}, u_0\right) + \widehat{\ddot{\alpha}}\left(\frac{s-1}{T}, u_0\right)y_{s-1}^2 + \widehat{\ddot{\beta}}\left(\frac{s-1}{T}, u_0\right)\widehat{\ddot{h}}_{s-1}\left(u_0, \widehat{\theta}_{R,T}\right), \tag{17}$$

where $\widehat{\ddot{\omega}}(\frac{s-1}{T}, u_0) = \widehat{\omega_0} + \widehat{\omega_1}(\frac{s-1}{T} - u_0)$, $\widehat{\ddot{\alpha}}(\frac{s-1}{T}, u_0) = \widehat{\alpha_0} + \widehat{\alpha_1}(\frac{s-1}{T} - u_0)$, $\widehat{\ddot{\beta}}(\frac{s-1}{T}, u_0) = \widehat{\beta_0} + \widehat{\beta_1}(\frac{s-1}{T} - u_0)$. Denote $\nabla = \frac{\partial}{\partial\theta} = (\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_3}, \frac{\partial}{\partial\theta_5})^\top$ and let $V = \frac{k_2}{k_1^2}(E[\epsilon_s^4] - 1)[\ddot{B}(\vartheta^*)]^{-1}$, where

$$\ddot{B}(\vartheta^*) = E\left[\frac{\nabla\ddot{h}_s(u_0,\vartheta^*)\nabla\ddot{h}_s(u_0,\vartheta^*)^\top}{\ddot{h}_s^2(u_0,\vartheta^*)}\right]. \tag{18}$$

Then, the following estimator

$$\widehat{V} = \frac{k_2}{k_1^2} \cdot \frac{1}{R} \sum_{s=2}^{T} k\left(\frac{s-t_0}{R}\right) \left[\left(\frac{y_s^2}{\widehat{\widehat{h}}_s\left(u_0, \widehat{\theta}_{R,T}\right)} \right)^2 - 1 \right] \tag{19}$$
$$\cdot \left[\frac{1}{R} \sum_{s=2}^{T} k\left(\frac{s-t_0}{R}\right) \frac{\nabla \widehat{\widehat{h}}_s\left(u_0, \widehat{\theta}_{R,T}\right) \nabla \widehat{\widehat{h}}_s\left(u_0, \widehat{\theta}_{R,T}\right)^\top}{\widehat{\widehat{h}}_s^2\left(u_0, \widehat{\theta}_{R,T}\right)} \right]^{-1}$$

is a consistent estimator of V.

Remarks: Although we show our estimator is consistent for α_0, α_1, β_0, β_1, ω_0, ω_1 in Theorem 1, we show it is asymptotically normally distributed only for α_0, β_0 and ω_0 in Theorem 2. This is due to the approach we take given the complexity involved in the GARCH(1,1) model with time-varying parameters. The parameters α_0, β_0, and ω_0 are useful for understanding the time-varying nature of the GARCH parameters and for out-of-sample forecasting at short horizons (horizons fixed relative to the sample size). The derivatives, α_1, β_1, and ω_1, may be useful for out-of-sample forecasting at long horizons (those that are proportional to the sample size), however. We leave it for future research to investigate asymptotic properties of our estimator of the derivatives.

In recent articles, Chen and Hong (2016) use local constant estimators of time-varying parameter GARCH models, and Kristensen and Lee (2019) develop local polynomial estimators for a class of dynamic models with time-varying parameters and use local constant estimator when considering a GARCH(1,1) model. The advantage of local linear estimators over local constant estimators is that the former estimator has smaller bias and suffers less from the boundary problem. This is important because out-of-sample forecasting is based on the parameter value at the end of the sample ($u_0 = 1$).

The local polynomial estimator in Rohan and Ramanathan (2013) and Rohan (2013) for time-varying GARCH(p, q) models appears more general than the results we present for a LL estimator, however, one has to recognize that it is built on the weak representation of a strong GARCH. To see this, taking a GARCH(1,1) with constant parameters, the equation for the variance can be rearranged as an ARMA(1,1) for the squared return:

$$y_t^2 = \omega + (\alpha + \beta) y_{t-1}^2 - \beta \nu_{t-1} + \nu_t$$

where $\nu_t \equiv y_t^2 - \sigma_t^2$ is a m.d.s. From Francq and Zakoïan (1998) and Francq and Zakoïan (2000), we know that nonlinear least squares would be a consistent estimator of (ω, α, β). A more simple, however, still consistent estimator can be obtained using the method proposed in Hannan and Rissanen (1982). In a first step, we approximate the ARMA(1,1) by a long auto-regression. In a second step, we replace ν_{t-1} by residuals from the long auto-regression and perform another regression. This approach is extended to time-varying parameters and local polynomial estimators by replacing the two regressions by local polynomial regressions. Two important drawbacks of this approach are loss of efficiency coming from using the weak representation (the conditional variance of ν_t is σ_t^4, however, we do not do GLS) and weak identification of the ARMA representation. In practice for GARCH models, α is small and $\alpha + \beta$ is close to one. It follows that for the ARMA representation the AR and MA operators are very similar so the ARMA(1,1) is close to being an

ARMA(0,0). Our method is based on the likelihood; it does not suffer from these issues, however, theoretical results are more difficult to derive.

It would be interesting to extend our results to other forms of the GARCH model, like the GARCH-GJR of Glosten et al. (1993) and the EGARCH of Nelson (1991) or likelihoods build on distributions other than Gaussian. It would require redoing most of the lemmas since they use the specific form of the basic GARCH with Gaussian innovations. We leave this for future work.

3 Simulations

In this section, we present results from Monte Carlo simulations to illustrate the behavior of our methods. First, we investigate the performance of two bandwidth selection procedures. The first consists in selecting R according to a pseudo out-of-sample criterion commonly used in the forecasting literature. At time t, for a given bandwidth R, we estimate the parameters of the time-varying GARCH model using observations $(y_1, \ldots, y_t)$. Since the focus is on out-of-sample performance we set $u_0 = 1$. We denote this estimate by $\widehat{\theta}_{R,t}$. With it, at time t, we compute the j-step ahead forecast of the variance,[3] $\widehat{h}_{t+j,R}$, for different values of j. In this simulation environment, it is possible to compare the forecast $\widehat{h}_{t+j,R}$ with the true value h_{t+j} and we evaluate the mean-squared forecast error:

$$\mathrm{FE}^{(0)}(R) = \frac{1}{T - \max\mathcal{R}} \sum_{t=\max\mathcal{R}}^{T-j} \left(h_{t+j} - \widehat{h}_{t+j,R}\right)^2, \tag{20}$$

where T is the sample size. This criterion is evaluated for different values of R in a set $\mathcal{R}$. For a given horizon j, the optimal bandwidth $\widehat{R}^{(0)}$ would be the one that minimizes $FE^{(0)}(R)$. A feasible alternative of this criterion when using real data and h_{t+j} is latent is to use the squared return y_{t+j}^2 as a proxy for h_{t+j}. See for example, Andersen and Bollerselv (1998) and Patton (2011) among many articles. It leads to the criterion

$$\mathrm{FE}^{(1)}(R) = \frac{1}{T - \max\mathcal{R}} \sum_{t=\max\mathcal{R}}^{T-j} \left(y_{t+j}^2 - \widehat{h}_{t+j,R}\right)^2 \tag{21}$$

and associated bandwidth $\widehat{R}^{(1)}$. The second bandwidth selection procedure is to use cross-validation (CV), as proposed in Kristensen and Lee (2019) for example.

In the simulations below, we take $\mathcal{R} = \{300, 400, \ldots, 2000\}$, sample size $T = 3000$ and we use either the Epanechnikov kernel or a uniform kernel. The parameters are re-estimated as we move forward through time. The number of replications is limited to 100 because they require a large number of numerical optimizations. We consider six different DGPs. The values taken by the parameters are plotted in Figure 1. In DGP 1, the persistence parameters α and β are constant and only ω changes over time following a sinus function. In DGP 2, all the parameters are increasing over time leading to more persistence and higher level of volatility. DGP 3 has constant parameters. In DGP 4, there is a discrete change in the value of the parameters. This DGP violates our assumption of smoothly

3 Because the LL approximation of the parameters is not random, the computation of multi-step ahead forecasts is the same as when parameters are constant. We proceed recursively using the law of iterated expectations since the variance at time $t+1$ is know at time t.

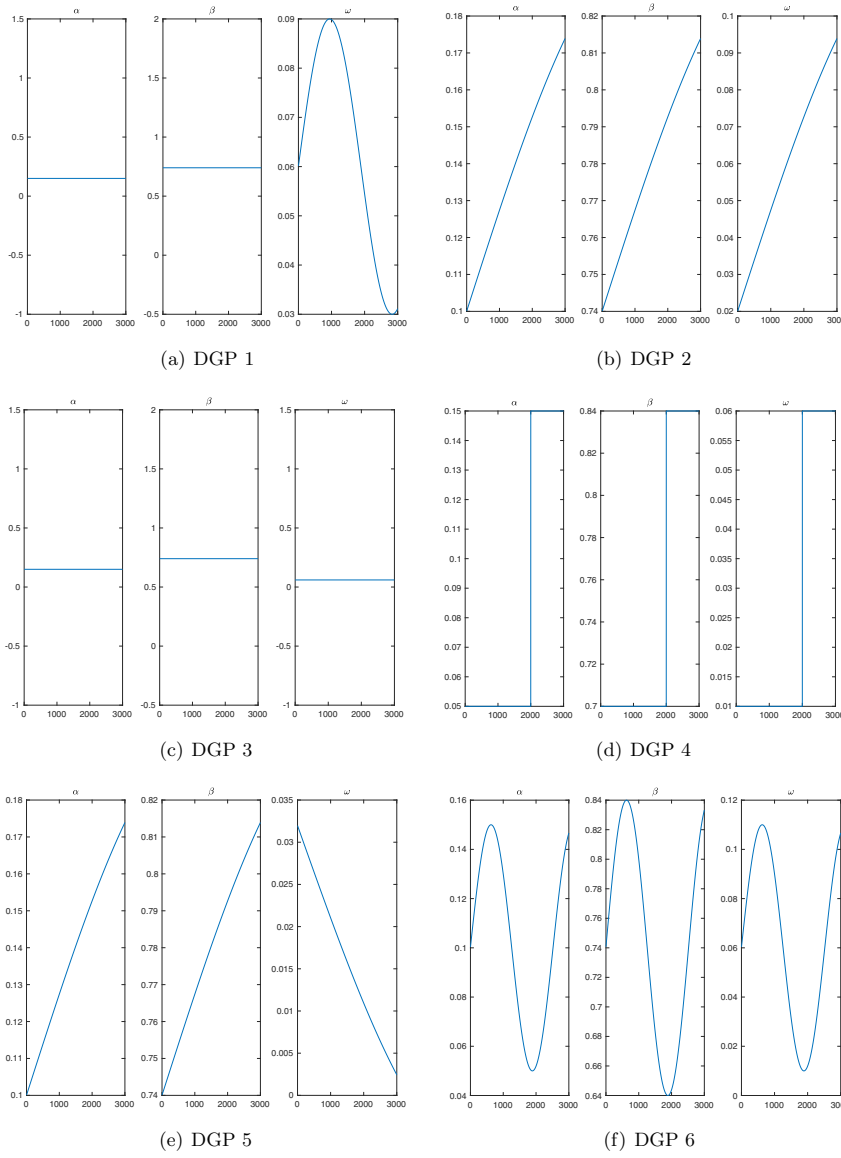

Figure 1 Values of the parameters for the different DGPs.

changing parameters over time. DGP 5 has increasing persistence of the volatility, however, ω is such that the overall level of volatility $\omega/(1-\alpha-\beta)$ is constant. In DGP 6, all three parameters follow a sinus function over time and they go through at least one full cycle.

Note that if $\bar{\alpha}(t-1/T,u_0)+\bar{\beta}(t-1/T,u_0)$ is increasing with t, then it is possible to go far enough in the future where the persistence will exceed one and as a result the multi-step ahead forecasts will start diverging with the forecasting horizon. Our practical experience

is that this is not really an issue, in part because the slopes α_1 and β_1 are typically not too large and in part because relevant forecasting horizons do not extend so far out as a fraction of the sample size. Ensuring $\bar{\alpha}(t-1/T,u_0)+\bar{\beta}(t-1/T,u_0) < 1$ for a specific range of values of t exceeding T can be imposed as a restriction when estimating the parameters if one so desired.

3.1 Bandwidth Selection

Results from these simulations for the bandwidth selection can be found in Table 1. We report the average value of the optimal bandwidth, using the true variance (columns 1 through 6) or the squared return as a proxy (columns 7 through 12). The last two columns are for the bandwidth selected by CV. For each criterion, we report the average of the optimal bandwidth for the LL estimator with either the Epanechnikov kernel or a uniform kernel, with or without bias-correction (BC). We also report the average of the optimal bandwidth for the locally constant estimator (KL) of Kristensen and Lee (2019) with the Epanechnikov kernel. We also include a model with constant parameters estimated using the most recent R observations (CST_R), with an implied uniform kernel.

Discussing first the results for the forecasting criteria, we can see that on average for the LL estimator the selected bandwidth is fairly large with values generally between 1200 and 1900. Across DGPs and forecast horizons, the average of the optimal bandwidth for the constant parameter GARCH is always smaller than for the LL estimator. This result is not surprising since we can see the LL estimator as reducing the amount of misspecification, thus allowing the model to "correctly" fit a longer time span. It is also perhaps not so surprising to have such large bandwidths considering that the DGPs, we considered have a high level of persistence for the variance and estimation of a GARCH model is data intensive since we are trying to capture the persistence of a latent process. The optimal bandwidths for the locally KL are slightly larger than for the constant parameters estimator when using the forecasting criteria. Overall, the optimal bandwidths increase with the forecasting horizons.

When using CV, the optimal bandwidths for the locally KL are much smaller than for the forecasting criteria across all DGPs. For the LL estimator, the optimal bandwidth varies a lot across the six DGPs. The more the parameters change over time, the smaller the optimal bandwidth. We go from the max of 2000 for DGP 1 through 3 to less than 500 for DGPs 5 and 6.

We next discuss the value of the forecasting criteria for the optimal bandwidths. In Tables 2 and 3, we present the value of the criteria when using the true variance or the squared return, respectively. Discussing the results in Table 2 first, a few patterns emerge. Bias correction does not help forecasting. To do bias correction we need to use the estimate of the slopes. As we will see later in this section, bias correction does improve the local estimation of the parameters, however, it hurts forecasting. For the LL estimator, we get better forecasts with the uniform kernel than with the Epanechnikov kernel whether or not we do bias correction (which depends on the kernel). Not surprisingly since it is not designed to minimize the forecasting criterion, cross validation leads to larger values of $\widehat{\widehat{R}}^{(0)}$. At shorter horizons, the locally KL (with the Epanechnikov kernel) gives smaller values of $\widehat{\widehat{R}}^{(0)}$ than the local linear estimator; however, the gap is pretty much closed when the forecasting horizon is increased to 30 periods ahead. For some of the DGPs if we use the uniform kernel the *LL* estimator can slightly out-perform the locally KL. Except for DGP 1 where the persistence parameters α and β are constant, in terms of forecasting nothing beats the constant

Table 1 Bandwidth selection results for simulations from the different DGP in Figure 1

	$\widehat{\bar{R}}^{(0)}$						$\widehat{\bar{R}}^{(1)}$							
DGP	LL_E	LL_{E+BC}	LL_U	LL_{U+BC}	CST_R	KL	LL_E	LL_{E+BC}	LL_U	LL_{U+BC}	CST_R	KL	LL_{CV}	KL_{CV}
	Forecasts 1 period ahead													
1	1703	1791	1459	1504	546	690	1269	1223	1200	1458	642	721	2000	432
2	1758	1884	1671	1713	1020	1385	1311	1373	1275	1791	986	1167	1970	416
3	1918	1969	1758	1865	1704	1858	1465	1337	1369	1763	1001	1265	1974	366
4	1318	1191	1481	1147	566	747	1202	954	946	1471	1136	1253	1000	523
5	1772	1933	1677	1840	968	1302	1320	1234	1177	1738	1131	1337	316	408
6	792	826	638	698	721	605	1118	982	1017	863	976	1118	465	334
	Forecasts 5 periods ahead													
1	1638	1671	1380	1410	485	579	1470	1596	1305	1344	518	617	2000	432
2	1787	1948	1604	1841	1028	1463	1439	1885	1359	1793	958	1350	1970	416
3	1918	1982	1799	1890	1645	1853	1697	1932	1535	1791	1279	1451	1974	366
4	1316	1409	1350	1252	714	1043	1202	1271	1036	1511	993	1108	1000	523
5	1771	1974	1652	1866	1059	1481	1406	1798	1277	1737	1102	1397	316	408
6	1321	1044	1140	855	1223	1271	1166	1193	1206	1014	926	1040	465	334
	Forecasts 20 periods ahead													
1	1619	1342	1381	1357	433	503	1481	1366	1280	1331	452	542	2000	432
2	1714	1212	1628	1835	699	1346	1585	1130	1504	1823	870	1286	1970	416
3	1939	1797	1785	1875	1605	1837	1844	1788	1683	1802	1336	1617	1974	366
4	1282	1070	1132	1461	685	1050	1247	1096	1075	1564	798	1146	1000	523
5	1744	1388	1599	1803	1161	1598	1603	1375	1466	1751	1163	1503	316	408
6	1498	1552	1413	941	1060	1267	1328	1500	1334	948	818	1032	465	334
	Forecasts 30 periods ahead													
1	1608	1638	1355	1374	427	484	1514	1581	1297	1308	445	507	2000	432
2	1755	1928	1611	1862	695	1253	1676	1922	1562	1860	790	1326	1970	416
3	1937	1965	1819	1882	1599	1814	1794	1935	1643	1809	1245	1488	1974	366
4	1320	1419	1088	1542	648	1050	1347	1438	1107	1597	749	1093	1000	523
5	1753	1904	1633	1811	1292	1659	1708	1865	1565	1786	1218	1581	316	408
6	1486	1127	1471	923	1025	1236	1419	1167	1419	964	755	1125	465	334

LL refers to the local-linear estimator and KL refers to the locally constant estimator. CST_R refers to the constant parameters estimator using the most recent R observations. The other subscripts represent the choice of kernel (E for Epanechnikov, U for uniform), the bias-correction of the LL estimator (BC), and the bandwidth selection done by CV.

parameters model estimated on a window (CST_R) with the largest performance gaps at shorter horizons.

To summarize the above results, from the point of view of forecasting, it is best to keep it simple and deal with parameters instability by using a uniform kernel, constant parameters and less than the full sample to estimate the model. At short horizons, constant parameters do better than locally constant, which does better than local linear. It is an example of the Keep it Sophisticatedly Simple principle of Diebold (2006). Further examples include forecasting exchange rates by a random walk being very hard to beat (see Rossi, 2013) or sample averages doing better than individual forecasts (see Aruoba et al. (2013) and Diebold and Shin (2019)). On the other hand, our simulations indicate that at longer

Table 2 Value of the pseudo out-of-sample criteria $FE^{(0)}$ for the different DGPs, estimators, bandwidth selections and forecast horizons

DGP	LL_E	LL_{E+BC}	LL_U	LL_{U+BC}	LL_{CV}	CST_R	CST_{inc}	KL_{CV}	KL
	Forecasts 1 period ahead								
1	0.0245	0.0552	**0.0228**	0.0518	0.0452	0.0243	0.0452	0.0313	0.0239
2	0.2932	0.5430	0.2695	0.4304	0.4001	**0.1500**	0.2815	0.3442	0.2092
3	0.0402	0.0867	0.0347	0.0726	0.0722	0.0171	**0.0158**	0.0513	0.0199
4	0.4552	1.1770	0.3981	0.9539	0.8527	**0.3680**	0.9478	0.8317	0.4359
5	0.0191	0.0383	0.0173	0.0316	0.0284	**0.0092**	0.0201	0.0223	0.0101
6	0.0989	0.2440	0.0986	0.2361	0.1258	**0.0842**	0.1101	0.1229	0.0961
	Forecasts 5 periods ahead								
1	0.1410	0.1527	**0.1404**	0.1509	0.1506	0.1411	0.1537	0.1433	0.1410
2	2.5414	2.6390	2.5340	2.5791	2.5714	**2.4922**	2.5097	2.5539	2.5148
3	0.2252	0.2421	0.2231	0.2350	0.2398	**0.2192**	0.2193	0.2278	0.2197
4	6.2281	6.5046	6.2197	6.4054	6.3173	**6.1843**	6.2635	6.2985	6.2129
5	0.1495	0.1577	0.1490	0.1537	0.1525	**0.1461**	0.1482	0.1502	0.1468
6	0.4761	0.5445	0.4754	0.5342	0.4930	**0.4720**	0.4773	0.4898	0.4760
	Forecasts 20 periods ahead								
1	0.1760	0.2047	**0.1753**	0.1871	0.1927	0.1769	0.2050	0.1798	0.1761
2	4.5899	5.5014	4.5750	4.6374	4.6679	**4.4315**	4.4732	4.6013	4.5126
3	0.2863	0.3400	0.2817	0.2925	0.3157	**0.2771**	0.2772	0.2877	0.2775
4	12.099	12.682	12.236	12.774	12.466	**11.995**	12.106	12.375	12.165
5	0.2606	0.3141	0.2607	0.2691	0.2698	**0.2518**	0.2531	0.2626	0.2534
6	0.8078	0.9323	0.8062	0.8820	0.8621	**0.7987**	0.8113	0.8365	0.8076
	Forecasts 30 periods ahead								
1	0.1761	0.1897	**0.1755**	0.1877	0.2045	0.1776	0.2117	0.1803	0.1764
2	4.9841	5.2564	4.9612	5.0159	5.1663	**4.7583**	4.8151	4.9783	4.8634
3	0.2887	0.3024	0.2832	0.2929	0.3335	**0.2788**	0.2790	0.2892	0.2792
4	13.454	13.483	13.716	14.512	14.198	**13.371**	13.512	13.895	13.493
5	0.2815	0.2957	0.2814	0.2910	0.3011	**0.2689**	0.2705	0.2834	0.2710
6	0.9004	1.0093	0.8987	0.9930	0.9924	**0.8891**	0.9063	0.9308	0.9013

LL represents the local-linear estimator, KL represents the locally constant estimator, and CST represents the constant parameters estimator. The subscripts represent the Epanechnikov kernel (E), the uniform kernel (U), the bias-correction of the LL estimator (BC). The bandwidths chosen are the optimal values from Table 1. CST_R and CST_{inc} are the constant parameters estimators using, respectively, the most recent R observations or an increasing window. Entries in bold correspond to the model that gives the lowest value of the criterion for a given DGP and forecast horizon.

horizons the more complicated *LL* estimator can dominate because it can track future evolution of the parameters.

The same patterns can be observed in Table 3. We know from Andersen and Bollerselv (1998) that the squared return is a very noisy proxy for the variance. As a result, the values of $FE^{(1)}$ are much larger than the corresponding values of $FE^{(0)}$. Overall, we see the same ranking.[4] Although the bandwidths obtained with $FE^{(1)}$ are similar, if not slightly smaller, to the ones

4 The magnitude of $FE^{(0)}$ and $FE^{(1)}$ varies greatly across DGPs because the variance of the simulated returns varies a lot across DGPs.

Table 3 Value of the pseudo out-of-sample criteria $FE^{(1)}$ for the different DGPs, estimators, bandwidth selections, and forecast horizons

DGP	LL_E	LL_{E+BC}	LL_U	LL_{U+BC}	LL_{CV}	CST_R	CST_{inc}	KL_{CV}	KL
	Forecasts 1 period ahead								
1	0.5430	0.5462	**0.5427**	0.5456	0.5450	0.5428	0.5450	0.5438	0.5429
2	7.5001	7.5291	7.4967	7.5145	7.5112	**7.4851**	7.5039	7.5109	7.4896
3	0.8615	0.8659	0.8609	0.8644	0.8644	**0.8603**	0.8614	0.8629	0.8606
4	**20.081**	20.130	20.153	20.197	20.195	20.134	20.211	20.198	20.110
5	0.4613	0.4633	0.4612	0.4627	0.4621	**0.4602**	0.4610	0.4621	0.4606
6	1.7356	1.7545	1.7357	1.7538	1.7442	1.7354	1.7409	1.7442	**1.7350**
	Forecasts 5 periods ahead								
1	0.5576	0.5607	**0.5573**	0.5600	0.5608	0.5575	0.5614	0.5587	0.5576
2	7.9935	8.0352	7.9884	8.0030	8.0156	**7.9637**	7.9763	8.0040	7.9831
3	0.8878	0.8916	0.8869	0.8895	0.8924	**0.8858**	0.8862	0.8888	0.8861
4	**21.328**	21.337	21.402	21.494	21.487	21.356	21.378	21.461	21.355
5	0.4844	0.4870	0.4842	0.4857	0.4864	**0.4823**	0.4830	0.4852	0.4831
6	1.8454	1.8632	1.8450	1.8610	1.8551	**1.8428**	1.8477	1.8530	1.8450
	Forecasts 20 periods ahead								
1	0.5632	0.5672	**0.5628**	0.5662	0.5702	0.5632	0.5729	0.5646	0.5632
2	8.7038	8.8371	8.6940	8.7363	8.7593	**8.6122**	8.6409	8.7136	8.6625
3	0.9056	0.9106	0.9039	0.9073	0.9163	**0.9021**	0.9025	0.9061	0.9025
4	23.193	**23.169**	23.344	23.694	23.524	23.199	23.277	23.448	23.175
5	0.5206	0.5260	0.5208	0.5262	0.5270	**0.5164**	0.5178	0.5224	0.5176
6	2.0324	2.0675	2.0319	2.0626	2.0594	**2.0275**	2.0362	2.0466	2.0331
	Forecasts 30 periods ahead								
1	0.5596	0.5638	**0.5592**	0.5629	0.5697	0.5601	0.5723	0.5613	0.5598
2	8.9653	9.1134	8.9531	8.9677	9.0739	**8.8223**	8.8625	8.9649	8.8931
3	0.9063	0.9108	0.9044	0.9074	0.9224	**0.9029**	0.9034	0.9068	0.9033
4	23.709	**23.580**	24.014	24.460	24.302	23.757	23.857	24.123	23.765
5	0.5316	0.5379	0.5314	0.5360	0.5427	**0.5236**	0.5247	0.5327	0.5252
6	2.1969	2.2417	2.1958	2.2351	2.2424	**2.1903**	2.2021	2.2122	2.1983

LL represents the local-linear estimator, KL represents the locally constant estimator, CST represents the constant parameters estimator. The subscripts represent the Epanechnikov kernel (E), the uniform kernel (U), and the bias-correction of the LL estimator (BC). The bandwidths chosen are the optimal values from Table 1. CST_R and CST_{inc} are the constant parameters estimators using, respectively, the most recent R observations or an increasing window. Entries in bold correspond to the model that gives the lowest value of the criterion for a given DGP and forecast horizon.

from $FE^{(0)}$, our conclusion from these simulations is that we would rather not use the squared return as a proxy. Following Liu et al. (2015), in our empirical application we will use five-minutes realized variance as a proxy to form a loss function.

3.2 Parameter Estimation

We now focus our attention on the estimation of the time-varying parameters. For the optimal bandwidths in Table 1 with forecasting horizon equal to one day or for CV,[5] we

5 We do not re-optimize the bandwidth selection for the values of $u_0 < 1$ because it would require too much computation.

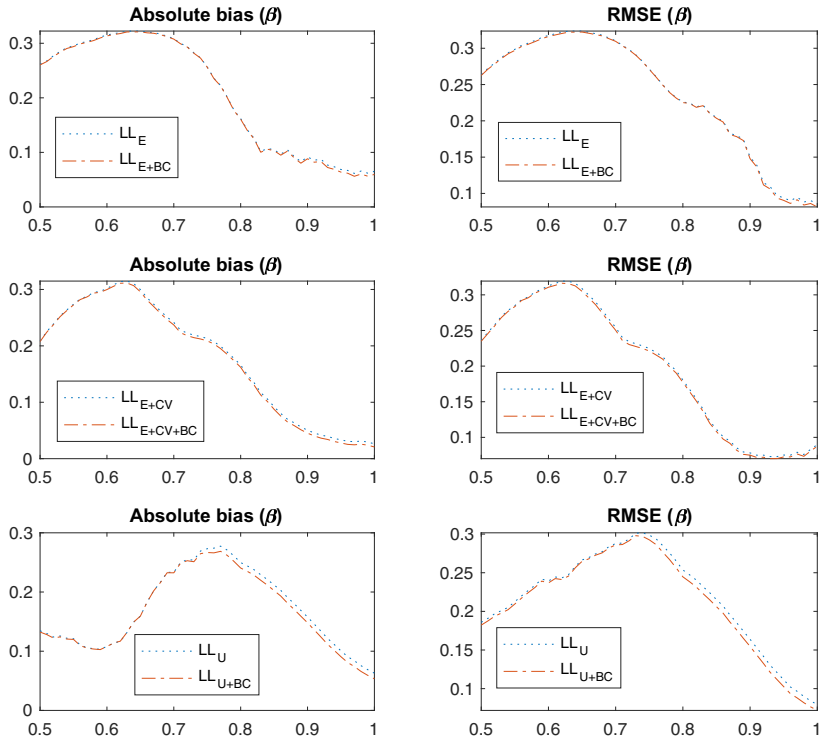

Figure 2 Impact of doing bias correction: absolute bias and RMSE for the estimation of $\beta(u_0)$ in DGP 6 with different versions of the LL estimators. The subscripts *E* and *U* mean that the estimator is using the Epanechnikov or uniform kernel, respectively. The bandwidths used are the ones that minimize the one-step ahead forecast criterion in Table 1 unless there is a subscript CV where in this case we use the CV bandwidth. The subscript BC represents the bias correction version of the LL estimator.

compare the estimation of the parameters $\theta(u_0)$ for a grid $u_0 \in [0.5, 0.51, 0.52, \ldots, 1]$. We present results for DGP 6, the DGP with the most parameter instability. Results for the other DGPs are available upon request. We look at bias and root mean square error (RMSE) across 1000 Monte Carlo replications, using the same sample size $T = 3000$.

We start by assessing the relative performance of the bias correction for the LL estimator. In Figures 2 and 3, we plot the absolute bias and RMSE across the value of u_0 for $\beta(u_0)$ and $\omega(u_0)$, respectively.[6] We compare the impact of doing bias correction when using the Epanechnikov or the uniform kernel, bandwidth optimized for the forecasting criterion $\mathrm{FE}^{(0)}$ or from CV. We can see that doing the bias correction always reduces the bias and RMSE, albeit the improvement can be very small. Moving forward, we will only consider the bias corrected LL estimator.

After confirming that BC improves the LL estimator in Figure 4, we assess the method for choosing the bandwidth and the choice of the kernel. To do so we look at the absolute

6 We get similar results for $\alpha(u_0)$.

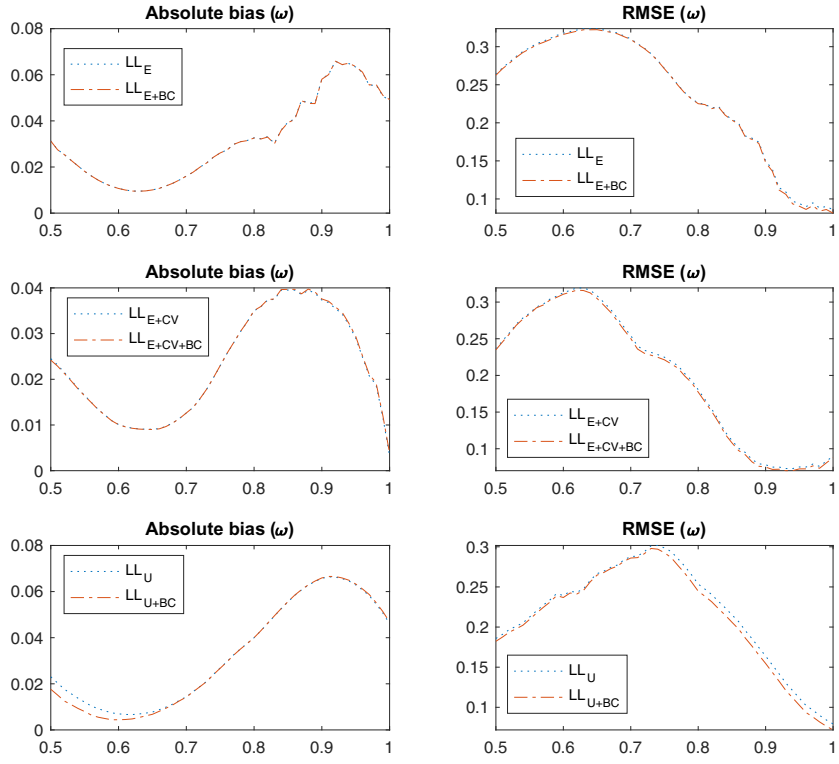

Figure 3 Impact of doing bias correction: absolute bias and RMSE for the estimation of $\omega(u_0)$ in DGP 6 with different versions of the LL estimator. The subscripts *E* and *U* mean that the estimator is using the Epanechnikov or uniform kernel, respectively. The bandwidths used are the ones that minimize the one-step ahead forecast criterion in Table 1 unless there is a subscript CV where in this case we use the CV bandwidth. The subscript BC represents the bias correction version of the LL estimator.

bias and RMSE for the estimation of $\beta(u_0)$ and $\omega(u_0)$ over the same range for u_0. Around the middle part of the sample (lower values of u_0), the uniform kernel does the best both in terms of bias and RMSE. However, as we move toward the end of the sample, the part that is likely to be of most interest to practitioners, the LL estimator using the Epanechnikov kernel performs the best when selecting the bandwidth through CV.

Finally, we then compare the performance of the best implementation of the LL estimator with the locally KL and the constant parameters model. Looking at Figure 5, while the model with constant parameters (CST_R) gave the best out-of-sample forecasts according to the criteria $FE^{(0)}$ and $FE^{(1)}$, it performs the worst with regards to bias and RMSE. Notably, toward the end of the interval there is a large bias, especially for the estimation of $\omega(u_0)$. The locally KL using $FE^{(0)}$ to select a bandwidth (KL) is almost as bad as CST_R. Using CV to select the bandwidth greatly reduces the bias of the locally KL by selecting a smaller bandwidth (334 instead of 605). However, all these estimators are generally dominated by the LL estimator. For most of the interval for u_0, it has a smaller bias and RMSE.

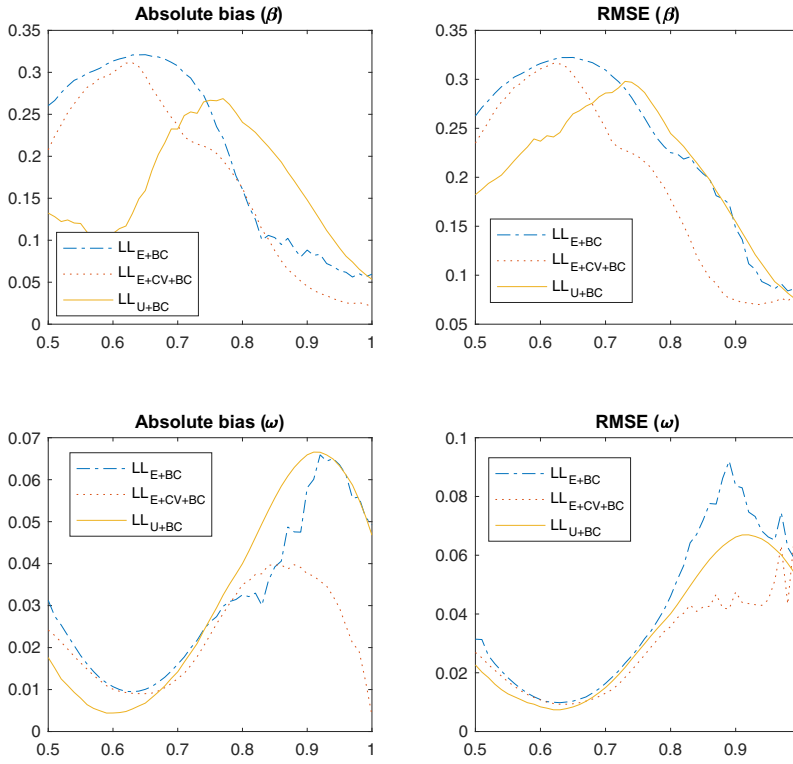

Figure 4 Absolute bias and RMSE for the estimation of $\beta(u_0)$ and $\omega(u_0)$ in DGP 6 with different versions of the LL estimator: impact of the bandwidth and kernel. The subscripts *E* and *U* mean that the estimator is using the Epanechnikov or uniform kernel, respectively. The bandwidth used is the optimal value from Table 1 for the one-step ahead criterion $FE^{(0)}$ unless the subscript includes CV in which case we use the CV bandwidth.

With the LL estimator we can estimate the slopes $(\alpha^{(1)}, \beta^{(1)}, \omega^{(1)})$ something that none of the other estimators can do. In Figure 6, we report the mean, 10 and 90% confidence bands for our LL estimation of the slopes for DGP 2. Results for the other DGPs are available upon request. We again see that the LL estimator is accurately tracking the true value of the slopes and as we go towards the end of the sample ($u_0 = 1$) the confidence bands are widening as one would expect.

4 Empirical Application

We apply our LL estimator to daily returns on different stocks indices. The dataset we use is the realized library of the Oxford-Man Institute of Quantitative Finance. Of the series that have at least 4000 daily observations, we use the following 10 series: Amsterdam Exchange Index (AEX), Dow Jones Industrial Average (DJI), CAC40 (FCHI), FTSE 100 (FTSE), S&P/TSX composite index (GSPTSE), Nasdaq 100 (IXIC), Nikkei 225 (N225), S&P 500 index (SPX), Swiss stock market index (SSMI), and Euro Stoxx 50 (STOXX).

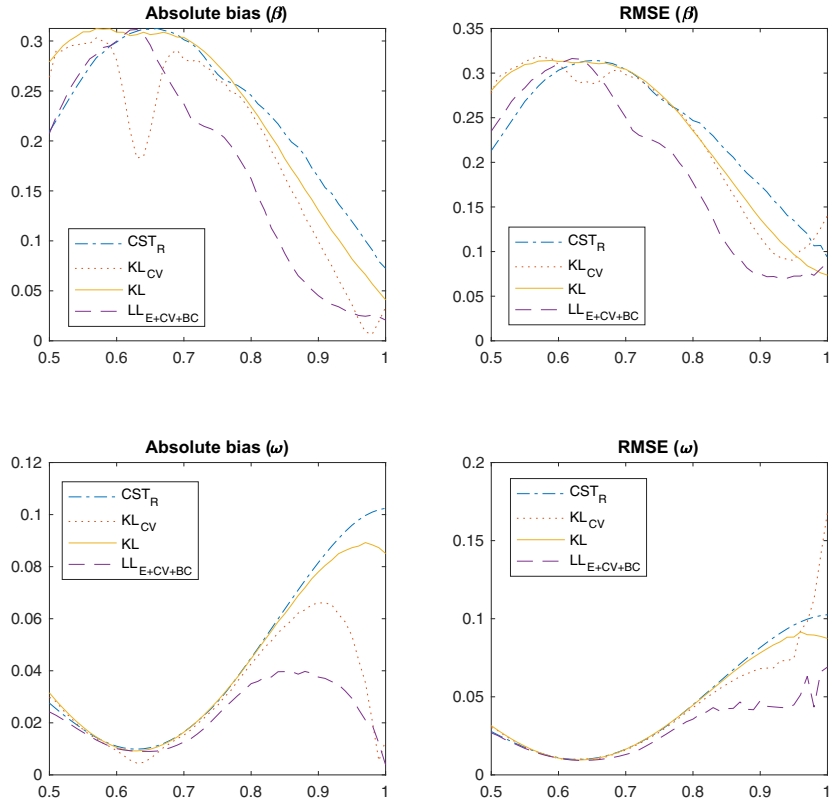

Figure 5 Absolute bias and RMSE for the estimation of $\beta(u_0)$ and $\omega(u_0)$ in DGP 6 for different estimators. CST_R is the constant parameters model using bandwidth R. KL_{CV} and KL are the locally KLs with the Epanechnikov kernel and bandwidth selected, respectively, by CV or the one-step ahead criterion $\text{FE}^{(0)}$. $\text{LL}_{E+\text{CV}+\text{BC}}$ is the bias-corrected LL estimator using the Epanechnikov kernel and the CV bandwidth.

Descriptive statistics for the series can be found in Table 4. All series start in January 2000 except FTSE (January 2001), GSPTSE (May 2002), and N225 (February 2000). All the series end on May 30, 2019. We compute the daily returns as the first difference of the log-price, times 100. We observe typical values for these statistics. The average daily return is very close to zero. The minimum and maximum daily returns for each series are fairly large. All the series are left skewed and leptokurtic. A plot of the daily return on AEX in Figure 7 illustrates the time-varying variance of these series.

This library also includes daily measures of volatility computed with high-frequency returns for each index. As is commonly done when forecasting the daily variance and high-frequency returns are available, we will compute forecast errors using realized variance computed with five-minutes returns (RV5) as a proxy for the unobserved variance. Descriptive statistics for the RV5 series of each index are in Table 4. Other high-frequency measures are included in the library; however, following Liu et al. (2015), we use RV5

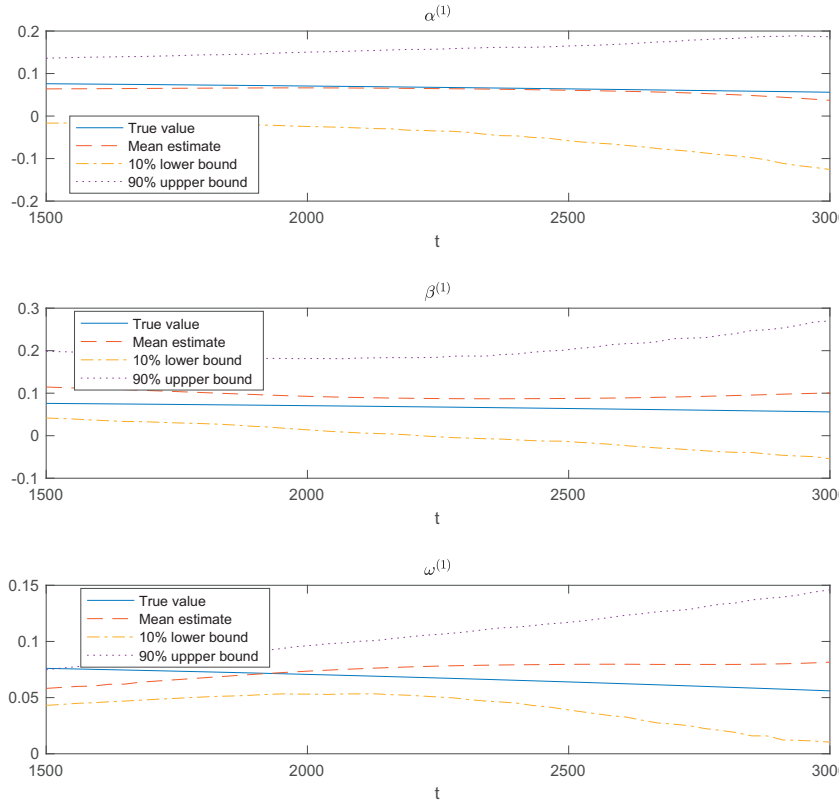

Figure 6 LL estimates of the slopes for DGP 2 using the Epanechnikov kernel and selecting the bandwidth with CV.

(scaled by 100^2 and not annualized so as to match our return series). It leads to the following criterion, focusing on one-step ahead forecasts:

$$\mathrm{FE}^{(2)}(R) = \frac{1}{T - \max\mathcal{R}} \sum_{t=\max\mathcal{R}}^{T} \left(\mathrm{RV5}_{t+1} - \widehat{h}_{t+1,R}\right)^2. \tag{22}$$

The bandwidth can be selected by minimizing $\mathrm{FE}^{(1)}$ or $\mathrm{FE}^{(2)}$. Results for $\mathrm{FE}^{(2)}$ and for CV will be presented below. We consider the following set $\mathcal{R} = \{300, \ldots, 3000\}$ setting $u_0 = 1$ since we are interested in out of sample forecasting. For the LL and locally KLs, we will focus on the Epanechnikov kernel and the models are re-estimated as we move forward through time. As a point of comparison, we also compute the optimal bandwidth for a GARCH model with constant parameters.

The results for the selection of the bandwidth are summarized in Table 5. Overall, the optimal bandwidths for the LL estimator across the 10 indices are larger than the values we found in the simulations for the different DGPs we considered in Section 3. We also find that for all series CV selects the largest bandwidths allowed. For the locally KL, as in the simulations the optimal bandwidths are smaller than for the LL estimator, sometimes by a

Table 4 Descriptive statistics for daily returns and RV5.

		Daily returns						RV5	
Index	T	Mean	Min	Max	Variance	Skewness	Kurtosis	Mean	Variance
AEX	4948	−0.005	−9.12	9.57	1.90	−0.18	9.48	1.15	3.93
DJI	4866	0.016	−8.61	10.53	1.26	−0.12	10.97	1.09	6.88
FCHI	4947	−0.003	−8.52	10.44	1.97	−0.10	7.96	1.33	5.06
FTSE	4893	0.001	−8.93	9.48	1.30	−0.17	9.47	1.17	7.45
GSPTSE	4273	0.017	−9.47	7.58	1.01	−0.69	12.19	0.87	25.84
IXIC	4865	0.012	−10.22	13.28	2.46	−0.05	9.01	1.22	4.91
N225	4723	0.001	−12.11	13.24	2.24	−0.43	9.33	1.04	2.79
SPX	4871	0.013	−9.69	10.64	1.42	−0.21	11.17	1.07	5.92
SSMI	4863	0.006	−9.07	10.79	1.35	−0.21	10.54	0.83	2.46
STOXX	4947	−0.008	−8.77	10.55	2.03	−0.09	7.90	1.57	9.83

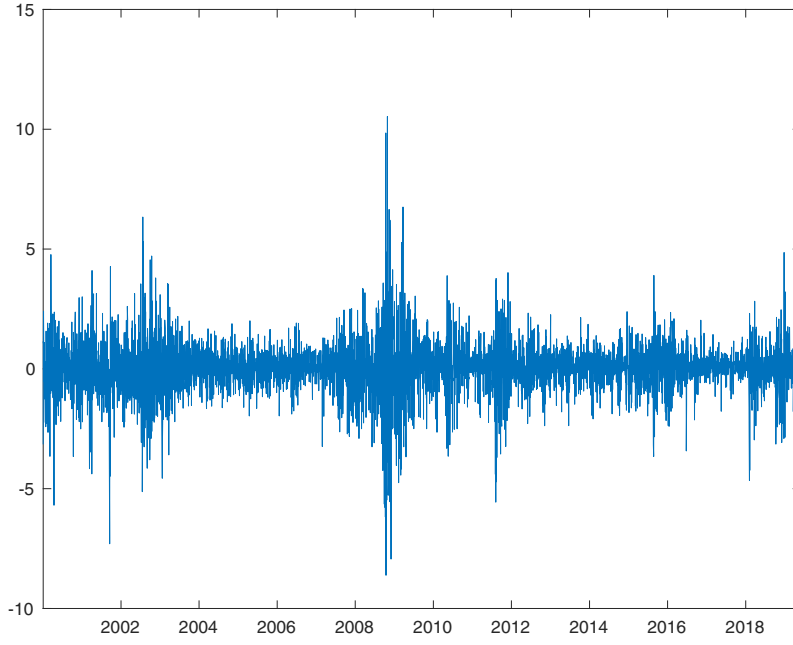

Figure 7 Daily return for AEX.

lot. For example, for STOXX and $FE^{(2)}$ the optimal bandwidths are 2900 and 1100, respectively, for the LL and locally KLs. It can be seen again as a sign that the local linear estimator reduces the level of misspecification of the model and allows it to fit a longer time span. For the model with constant parameters, the bandwidths are even smaller than for the locally KL. For the values of the criterion $FE^{(2)}$, which is based on one-step ahead

Table 5 Bandwidth selection and value of the criteria for the different indices.

	$\widehat{R}^{(2)}$			$\widehat{R}_{\mathrm{CV}}$		$\mathrm{FE}^{(2)}(\widehat{R}^{(2)})$			$\mathrm{FE}^{(2)}(\widehat{R}_{\mathrm{CV}})$	
Index	LL	KL	CST_R	LL	KL	LL	KL	CST_R	LL	KL
AEX	2500	1600	300	3000	700	0.963	0.955	0.896	0.973	1.007
DJI	2000	700	300	3000	1800	2.211	2.197	2.171	2.231	2.238
FCHI	2800	1100	500	3000	700	1.356	1.353	1.273	1.360	1.415
FTSE	1800	1000	1200	3000	400	2.180	2.181	2.179	2.204	2.209
GSPTSE	1900	800	700	3000	400	0.908	0.911	0.903	0.924	0.928
IXIC	1300	700	500	3000	1900	0.682	0.659	0.638	0.761	0.741
N225	2200	1200	1000	3000	700	3.606	3.582	3.373	3.713	3.779
SPX	1200	700	500	3000	1800	1.028	1.009	0.994	1.052	1.053
SSMI	2300	1600	800	3000	700	1.733	1.650	1.601	1.804	1.969
STOXX	2900	1100	300	3000	700	2.711	2.713	2.606	2.711	2.753

The bandwidth $\widehat{R}^{(2)}$ is obtained by minimizing the criterion $\mathrm{FE}^{(2)}$. The bandwidth $\widehat{R}_{\mathrm{CV}}$ is obtained by CV.

forecasts, the results are as in the simulation exercise. This criterion is favoring simpler models, with the constant parameters estimator ahead of the locally KL, which itself is ahead of the LL estimator.

In Table 6, we extend the results for one-step ahead forecasts of Table 5 to 5 and 30 periods ahead, as well as a broader set of models and estimators. We now also include a constant parameters GARCH estimated with an increasing window ($\mathrm{CST}_{\mathrm{inc}}$), the regime switching GARCH(1,1) model of Haas et al. (2004) with two regimes (RS) and the IGARCH(1,1) model (INT). Overall, the results are similar to the simulation results in Table 2. At short horizons (one and five days), the simpler model CST_R has the smallest value for the criterion, except for DJI and FTSE for five days ahead. At longer horizons (30 days), the results are mixed with $\mathrm{LL}_{\mathrm{BC}}$ and KL having the smallest value of $\mathrm{FE}^{(2)}$ for, respectively, three and four of the indices.

Next, for the selected bandwidth we look at estimates for different values of u_0 for the bias-corrected LL and locally KLs. As a point of comparison, we also include estimates of a GARCH model with constant parameters estimated on the whole sample. As an illustration, Table 7 reports estimates and standard errors for AEX. The estimates are clearly changing as we move through the sample. Focusing on persistence $(\alpha(u_0)+\beta(u_0))$ of the GARCH process, the estimates obtained with the LL estimator do not decrease as much as with the locally KL (0.9883 at $u_0 = 0.5$ to 0.9656 at $u_0 = 1$ versus 0.9921 at $u_0 = 0.5$ to 0.9184 at $u_0 = 1$). Figure 8 plot the LL estimates of $\bar{\alpha}(u_0)$, $\bar{\beta}(u_0)$, $\bar{\omega}(u_0)$ and $\bar{\alpha}(u_0)+\bar{\beta}(u_0)$ for AEX for a finer grid of values than in the table. The straight line in these figures is the estimate obtained from a GARCH with constant parameters estimated on the whole sample. For each sub-figure, the marks correspond to the estimates at the different values of u_0 at which we do the estimation. From each mark, we extend the linear approximations $\bar{\alpha}(t-1/T,u_0)$, $\bar{\beta}(t-1/T,u_0)$, and $\bar{\omega}(t-1/T,u_0)$ over the 100 days preceding the time point corresponding to the given value of u_0 so as to convey information about the estimates of $\alpha^{(1)}(u_0)$, $\beta^{(1)}(u_0)$ and $\omega^{(1)}(u_0)$. The sequence of linear approximations is pretty well aligned, pointing toward the location of the previous or next mark (the exception being the estimation of $\omega(u_0)$ in the middle part). Focusing on the lower right panel of Figure 8, we

Table 6 Value of the $FE^{(2)}$ criterion for out-of-sample forecasts with the ten indices and different models.

Index	CST_{inc}	CST_R	LL_{BC}	KL	RS	INT
	Forecasts 1 period ahead					
AEX	1.009	**0.896**	0.963	0.955	1.102	0.956
DJI	2.327	**2.171**	2.211	2.197	2.631	2.370
FCHI	1.391	**1.273**	1.356	1.353	1.624	1.317
FTSE	2.208	**2.179**	2.180	2.181	2.260	2.205
GSPTSE	0.934	**0.903**	0.908	0.911	0.985	0.939
IXIC	0.892	**0.638**	0.682	0.659	1.058	0.821
N225	3.575	**3.373**	3.606	3.582	4.878	3.546
SPX	1.117	**0.994**	1.028	1.009	1.422	1.139
SSMI	1.716	**1.601**	1.733	1.650	3.157	1.567
STOXX	2.776	**2.606**	2.711	2.713	3.786	2.722
	Forecasts 5 periods ahead					
AEX	1.177	**0.966**	1.088	1.083	1.225	1.101
DJI	2.592	2.565	2.570	**2.541**	2.839	2.579
FCHI	1.572	**1.356**	1.502	1.502	1.779	1.481
FTSE	2.347	2.304	2.292	**2.288**	2.372	2.329
GSPTSE	1.000	**0.971**	0.981	0.985	1.033	0.995
IXIC	1.126	**0.831**	0.886	0.850	1.215	0.990
N225	3.966	**3.694**	3.934	3.936	5.197	4.004
SPX	1.335	**1.244**	1.323	1.271	1.577	1.306
SSMI	1.838	**1.665**	1.826	1.706	3.236	1.706
STOXX	3.028	**2.720**	2.915	2.926	3.976	2.948
	Forecasts 30 periods ahead					
AEX	1.483	**1.018**	1.224	1.268	1.429	1.359
DJI	2.720	2.639	2.658	**2.610**	2.971	2.760
FCHI	1.988	**1.535**	1.840	1.844	2.118	1.902
FTSE	2.572	2.449	**2.392**	2.400	2.565	2.557
GSPTSE	1.054	1.002	**0.988**	1.012	1.087	1.057
IXIC	1.496	0.981	0.975	**0.974**	1.365	1.203
N225	4.431	4.095	**4.087**	4.227	5.779	4.900
SPX	1.494	1.358	1.428	**1.354**	1.692	1.472
SSMI	1.978	1.637	1.764	**1.631**	3.476	1.965
STOXX	3.451	**2.902**	3.268	3.255	4.468	3.413

The numbers in bold correspond to the model that gives the lowest value of the criterion for a given index and forecast horizon. CST_{inc} and CST_R are the constant parameters GARCH model estimated with an increasing window or a fixed window, respectively. LL_{BC} is the bias-corrected LL estimator using the Epanechnikov kernel. KL is the locally constant estimator with Epanechnikov kernel. RS is the regime switching GARCH(1,1) model of Haas et al. (2004) with two regimes and the IGARCH(1,1) model is INT. The value of the different bandwidths are from Table 5.

see that the LL estimate of $\alpha(u_0) + \beta(u_0)$ is always below the constant GARCH estimate, while this is not the case for the individual parameters. As previously noted in Hillebrand (2005), for GARCH models, we know that if the parameters are changing over time, estimators of $\alpha + \beta$ that do not take changes into account will converge to one. It is also interesting to note that the

Table 7 Parameter estimates for AEX for different values of u_0.

u_0	1.0	0.9	0.8	0.7	0.6	0.5
	Constant parameters					
A	0.1063	0.1063	0.1063	0.1063	0.1063	0.1063
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
B	0.8833	0.8833	0.8833	0.8833	0.8833	0.8833
	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)	(0.0026)
Ω	0.0192	0.0192	0.0192	0.0192	0.0192	0.0192
	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
	LL parameters					
$\alpha(u_0)$	0.1390	0.1184	0.1121	0.1084	0.1053	0.1042
	(0.0157)	(0.0252)	(0.0186)	(0.0162)	(0.0167)	(0.0157)
$\beta(u_0)$	0.8266	0.8552	0.8687	0.8756	0.8812	0.8841
	(0.0185)	(0.0234)	(0.0178)	(0.0163)	(0.0156)	(0.0147)
$\omega(u_0)$	0.0298	0.0272	0.0244	0.0225	0.0212	0.0204
	(0.0035)	(0.0051)	(0.0052)	(0.0049)	(0.0048)	(0.0046)
$\alpha^{(1)}(u_0)$	0.1214	0.0437	0.0347	0.0421	0.0109	0.0015
$\beta^{(1)}(u_0)$	−0.1571	−0.0710	−0.0546	−0.0586	−0.0249	−0.0136
$\omega^{(1)}(u_0)$	−0.0083	−0.0116	−0.0054	0.0114	0.0047	0.0035
	Locally constant parameters					
$\alpha(u_0)$	0.1518	0.1239	0.1253	0.0917	0.0759	0.1079
	(0.0568)	(0.0317)	(0.0284)	(0.0211)	(0.0158)	(0.0219)
$\beta(u_0)$	0.7666	0.8521	0.8547	0.8822	0.9057	0.8842
	(0.0718)	(0.0275)	(0.0230)	(0.0172)	(0.0107)	(0.0153)
$\omega(u_0)$	0.0537	0.0246	0.0282	0.0289	0.0240	0.0291
	(0.0254)	(0.0112)	(0.0131)	(0.0114)	(0.0098)	(0.0145)

The standard errors are between parentheses. The LL and locally constant estimators are using the Epanechnikov kernel and the bandwidth is selected by CV at $u_0 = 1$.

estimates of $\alpha(u_0) + \beta(u_0)$ steadily decrease as we move away from the financial crisis of 2007–2008, meaning that volatility has become less persistent.

Finally, another way to investigate the impact of allowing parameters to change over time is by computing risk measures. Namely, we compute one-day ahead VaR and ES. For a given coverage rate p, VaR, and ES are formally defined as

$$\Pr(y_t < -\mathrm{VaR}_{t+1}|\mathcal{F}_t) = p$$
$$\mathrm{ES}_{t+1} = -E[y_t|y_t < -\mathrm{VaR}_t, \mathcal{F}_t],$$

where $\mathcal{F}_t$ represents the information set at time t. The convention is to introduce minus signs so losses are presented as positive numbers. We compute VaR and ES in the following way: using returns from time 1 to t,

- Estimate the set of models and estimators[7] to get $\widehat{\theta}_t$ for each.

7 We implement the GARCH model with constant parameters with an increasing window ($\mathrm{CST}_{\mathrm{inc}}$) or a bandwidth ($\mathrm{CST}_R$), the LL ($\mathrm{LL}_{\mathrm{BC+CV}}$) and locally constant ($\mathrm{KL}_{\mathrm{CV}}$) estimators with the Epanechnikov kernel, the regime switching GARCH(1, 1) model of Haas et al. (2004) with two regimes (RS) and the IGARCH(1, 1) model (INT).

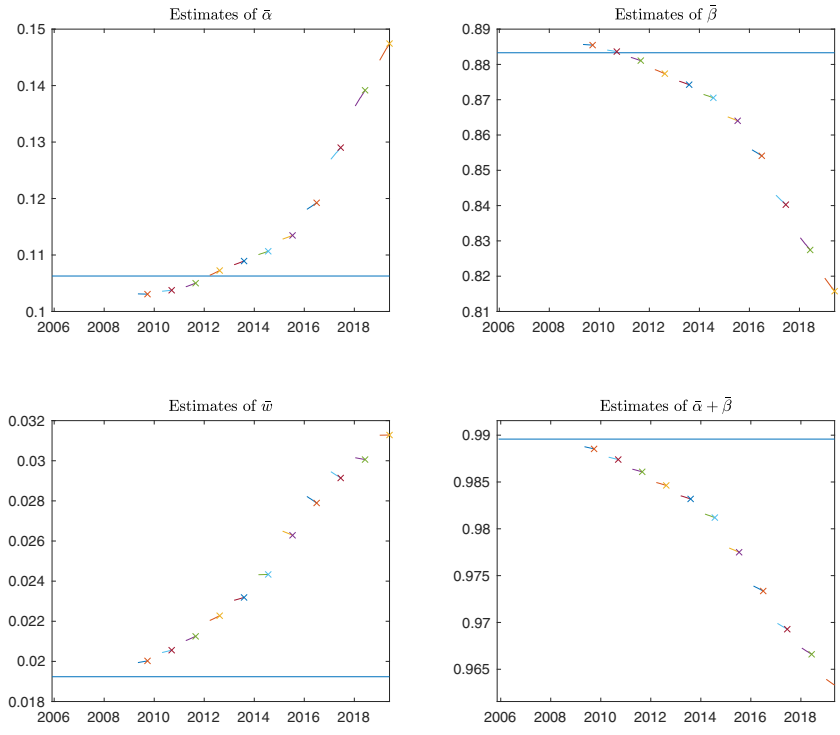

Figure 8 Estimates of $\bar{\alpha}(u_0)$, $\bar{\beta}(u_0)$, $\bar{\omega}(u_0)$, and $\bar{\alpha}(u_0)+\bar{\beta}(u_0)$ for different values of u_0. The asset is AEX. The straight lines correspond to the estimates of a GARCH with constant parameters applied to the whole sample.

- With $\widehat{\theta}_t$, compute the predicted variance $\widehat{h}_{t+1}$ and the standardized residuals $\widehat{\epsilon}_s = y_s/\sqrt{\widehat{h}_s}$ for $s = 1, \ldots, t$.
- Compute VaR and ES as

$$\mathrm{VaR}^p_{t+1} = -\sqrt{\widehat{h}_{t+1}}\,\mathrm{percentile}\big(\{\widehat{\epsilon}_s\}_{s=1,\ldots,t}, 100p\big), \tag{23}$$

$$\mathrm{ES}^p_{t+1} = -\sqrt{\widehat{h}_{t+1}}\,\frac{\sum_{s=1}^{t}\widehat{\epsilon}_s \mathbf{I}\big(\widehat{\epsilon}_s < -\mathrm{VaR}^p_s\big)}{\sum_{s=1}^{t}\mathbf{I}\big(\widehat{\epsilon}_s < -\mathrm{VaR}^p_s\big)}, \tag{24}$$

where $\mathbf{I}$ denotes the indicator function equal to one if the condition is true, zero if false. We iterate these steps from time t until time $T-1$.

VaR and ES are computed ex-ante. Ex-post we observe the actual return y_t and we compute the hit sequence, $I_{t+1} = 1$ if $y_{t+1} < -\mathrm{VaR}^p_{t+1}$, otherwise $I_{t+1} = 0$. We describe $I_t = 1$ as a violation. From the sequence of returns, VaR and ES we can backtest the hypothesis that the risk measures are correctly computed. We consider the following popular backtests:

- the Markov test of Christoffersen (1998) based on the hit sequence;

- the Weibull test of Christoffersen and Pelletier (2004) based on the durations between hits;
- the CAViaR test of Engle and Manganelli (2004) implemented with lagged hit and VaR as regressors; and
- the ES backtest of Christoffersen (2003) implemented with either the VaR (C_{VaR}) or the sample variance of the daily returns over the last 20 days (C_{vol}) as a regressor.

For all these tests, we test the "conditional coverage" hypothesis. Namely, we jointly test that hits happen at the rate prescribed by the coverage rate of the risk measure and that hits or magnitude of the violations cannot be predicted. A detailed presentation of the implementation of all these tests and the computation of the p-value by Monte Carlo simulations can be found in the online Appendix. We implement the backtests over the last 1500 observations for each series and set the coverage rate at $p = 5\%$.

The results of the backtesting exercise for indices AEX, DJI, and SPX are presented in Table 8. We report the value of the different backtests and the associated (Monte Carlo) p-value.[8] With a 5% coverage rate, the expected number of violations is 75. Some combinations of index and estimator are very close to that number (e.g., AEX with $\text{LL}_{\text{BC+CV}}$ or KL_{CV}), some are above (CST_R with SPX or AEX), and some are below (SPX with RS).

As for the outcome of the different backtests, the results are mixed. For AEX, none of the models and estimators are rejected by the ES backtests, however, all are rejected by some of the VaR backtests except the simplest model (CST). For DJI and SPX our proposed LL estimator is not rejected by any of the VaR backtests, something that cannot be said for any of the other methods. There is no method that is not rejected by some of the ES or VaR backtests for some of the indices. A more flexible method like our LL estimator might improve the accuracy of the computation of risk measures, however, the underlying GARCH(1,1) model is probably inadequate for these series.

5 Concluding Remarks

Given empirical evidence that the GARCH parameters are not stable over time, we consider GARCH models with time-varying parameters in this paper. In our approach, we model the parameters as smooth functions of time and estimate them with a LL estimator. We show that our estimator is consistent and is asymptotically normally distributed given a point of time. To select the bandwidth of the kernel used by our estimator, we propose minimizing a pseudo out-of-sample squared forecast error criterion. We illustrate the usefulness of our approach in Monte Carlo simulation experiments and empirical application to risk management.

It would be interesting to extend our pointwise results to joint uniform results, which would allow us to test a shape restriction on the GARCH parameters (e.g., whether or not the parameters are constant over time). We leave it for future research.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics online*.

8 For the seven other indices, we get no p-value $<10\%$ for FTSE, GSPTSE, N225, and SSMI. For FCHI and STOXX, a few of the p-values for the VaR backtests are $<10\%$ while most of the ES backtest of IXIC are rejected at the 10% significance level.

Table 8 Backtesting results for VaR and ES

Index	Model	Nb. hits	$C_{V\,aR}$	C_{vol}	Markov	Weibull	CAViaR
AEX	CST_{inc}	61	1.288	1.375	4.278	3.637	5.271
			(0.595)	(0.547)	(0.126)	(0.245)	(0.242)
	CST_R	84	1.399	1.380	**8.202**	3.316	**14.390**
			(0.576)	(0.546)	**(0.024)**	(0.280)	**(0.045)**
	LL_{BC+CV}	73	1.610	1.833	**12.021**	4.579	**11.339**
			(0.546)	(0.475)	**(0.004)**	(0.163)	**(0.063)**
	KL_{CV}	74	0.789	0.698	**8.912**	2.474	**11.558**
			(0.720)	(0.741)	**(0.019)**	(0.395)	**(0.060)**
	RS	63	0.759	1.517	**6.117**	3.727	8.282
			(0.597)	(0.527)	**(0.073)**	(0.235)	(0.115)
	INT	74	0.928	1.298	4.600	0.991	**10.576**
			(0.684)	(0.566)	(0.112)	(0.695)	**(0.073)**
DJI	CST_{inc}	63	3.387	7.895	**6.239**	5.321	6.268
			(0.313)	(0.113)	**(0.072)**	(0.117)	(0.198)
	CST_R	74	8.008	9.430	0.594	0.057	2.288
			(0.122)	(0.088)	(0.727)	(0.980)	(0.536)
	LL_{BC+CV}	68	7.769	**13.708**	2.164	1.916	2.186
			(0.127)	**(0.047)**	(0.347)	(0.494)	(0.556)
	KL_{CV}	79	**10.232**	8.352	**7.288**	**7.093**	**10.594**
			(0.089)	(0.104)	**(0.044)**	**(0.054)**	**(0.072)**
	RS	63	17.291	**16.529**	4.509	4.935	4.853
			(0.122)	**(0.029)**	(0.115)	(0.137)	(0.264)
	INT	68	6.818	**9.115**	3.541	2.445	4.771
			(0.152)	**(0.092)**	(0.177)	(0.399)	(0.268)
SPX	CST_{inc}	55	8.496	**9.171**	**10.380**	**8.135**	**15.476**
			(0.113)	**(0.092)**	**(0.010)**	**(0.034)**	**(0.042)**
	CST_R	81	**11.819**	**18.189**	2.016	1.940	1.509
			(0.070)	**(0.023)**	(0.373)	(0.490)	(0.701)
	LL_{BC+CV}	65	7.983	**12.523**	1.796	2.014	2.250
			(0.123)	**(0.057)**	(0.409)	(0.475)	(0.542)
	KL_{CV}	81	**14.179**	**11.723**	3.321	3.591	**11.672**
			(0.052)	**(0.063)**	(0.200)	(0.249)	**(0.060)**
	RS	59	11.175	**10.107**	**7.039**	**5.961**	**11.752**
			(0.116)	**(0.080)**	**(0.058)**	**(0.089)**	**(0.059)**
	INT	64	**10.891**	**9.724**	3.877	3.480	**13.509**
			(0.083)	**(0.084)**	(0.154)	(0.260)	**(0.048)**

The numbers between parentheses are the *p*-values for the hypothesis that the risk measure is correctly computed. Results with a *p*-value below 10% are in bold.

Appendix A: Proofs of Theorems 1, 2, and 3

The lemmas and their proofs are in the online Appendix.

Proof of Theorem 1. Part (a): By Lemma 2, we know $\tilde{L}_{R,T}(u_0,\theta) \xrightarrow{p} E\left[\ddot{L}_{R,T}(u_0,\vartheta)\right]$ uniformly in θ. Next, we will show that $E\left[\ddot{L}_{R,T}(u_0,\vartheta)\right]$ is well defined and $E\left[\ddot{L}_{R,T}(u_0,\vartheta)\right]$ has an

identifiably unique maximizer at ϑ^*. Since $\lim_{T\to\infty}\left\{E\left[\ddot{L}_{R,T}(u_0,\vartheta)\right]-E\left[\ddot{l}_{t_0,T}(u_0,\vartheta)\right]\right\}=0$, where $\ddot{l}_{t_0,T}(u_0,\vartheta)=-\frac{1}{2}\left(\frac{\ddot{y}^2_{t_0}(u_0)}{\ddot{h}_{t_0}(u_0,\vartheta)}+\ln\ddot{h}_{t_0}(u_0,\vartheta)\right)$, then on the one hand we have

$$E\left[\left(-\ddot{l}_{t_0,T}(u_0,\vartheta)\right)^-\right]\le E\left[\left(\ln\ddot{h}_{t_0}(u_0,\vartheta)\right)^-\right]\le E\left[\left(\ln\omega(u_0)\right)^-\right]<\infty, \tag{A1}$$

where f^- represents the negative part of f, that is, $f^-=\max(-f,0)$. On the other hand,

$$E\left[-\ddot{l}_{t_0,T}(u_0,\vartheta)\right]=E\left[\epsilon^2_{t_0}\right]+E\left[\ln\ddot{h}_{t_0}(u_0,\vartheta)\right]\le 1+\ln E\left[\ddot{h}_{t_0}(u_0,\vartheta)\right]<\infty, \tag{A2}$$

by Jensen's inequality. Hence, $E\left[\ddot{L}_{R,T}(u_0,\vartheta)\right]$ is well defined on $\bar{\mathbb{R}}$. For $\theta\in\Theta$, we have

$$E\left[-\ddot{l}_{t_0,T}(u_0,\vartheta)\right]-E\left[-\ddot{l}_{t_0,T}(u_0,\vartheta^*)\right]=\frac{1}{2}\left\{E\left[\frac{\ddot{h}_{t_0}(u_0,\vartheta^*)}{\ddot{h}_{t_0}(u_0,\vartheta)}\right]E\left[\epsilon^2_{t_0}\right]+E\left[\ln\left(\frac{\ddot{h}_{t_0}(u_0,\vartheta)}{\ddot{h}_{t_0}(u_0,\vartheta^*)}\right)\right]-1\right\}. \tag{A3}$$

Because function $f(x)=x-1-\ln x$ for $x>0$ is minimized when $x=1$ and $f(1)=0$, thus $\ln x\le x-1$ for $x>0$. Hence, $E\left[\ddot{l}_{t_0,T}(u_0,\vartheta)\right]-E\left[\ddot{l}_{t_0,T}(u_0,\vartheta^*)\right]\le 0$ with equality holds when $\frac{\ddot{h}_{t_0}(u_0,\vartheta^*)}{\ddot{h}_{t_0}(u_0,\vartheta)}=1$ almost surely, and if and only if $\vartheta=\vartheta^*$. By Assumption 3, we have $\widehat{\vartheta}_{R,T}\xrightarrow{p}\vartheta^*$.

Part (b): Let $\check{\theta}_{R,T}=\left(\widehat{\vartheta}_{R,T},\boldsymbol{\theta}\right)^\top$. Plugging the result in Lemma 12 into the gradient vector of $\tilde{L}_{R,T}(u_0,\theta)$ with respect to $\boldsymbol{\theta}$, evaluating the expression at $\theta=\theta_{R,T}$, and multiplying it by $(T/R)^2$, we obtain

$$\begin{aligned}
&\frac{1}{2R}\sum_{s=2}^{T}k\left(\frac{s-t_0}{R}\right)\left(\epsilon^2_s-1\right)\frac{T^2\nabla\tilde{h}_s\left(u_0,\theta_{R,T}\right)}{R^2\tilde{h}_s\left(u_0,\theta_{R,T}\right)}\\
&+\frac{1}{2R}\sum_{s=2}^{T}k\left(\frac{s-t_0}{R}\right)\frac{\epsilon^2_sT^2\left(h\left(\frac{s}{T}\right)-\tilde{h}_s\left(u_0,\theta_{R,T}\right)\right)\nabla\tilde{h}_s\left(u_0,\theta_{R,T}\right)}{R^2\tilde{h}_s\left(u_0,\theta_{R,T}\right)^2}\\
&=\frac{1}{2R}\sum_{s=2}^{T}k\left(\frac{s-t_0}{R}\right)\left(\epsilon^2_s-1\right)\frac{T^2\nabla\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)}{R^2\tilde{h}_s\left(u_0,\theta_{R,T}\right)}\\
&+\frac{1}{2R}\sum_{s=2}^{T}k\left(\frac{s-t_0}{R}\right)\frac{\epsilon^2_sT^2\nabla\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)\nabla\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)^\top}{R^2\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)^2}\left(\widehat{\vartheta}_{R,T}-\vartheta\right)\\
&+\frac{1}{2R}\sum_{s=2}^{T}k\left(\frac{s-t_0}{R}\right)\frac{\epsilon^2_sT^2\nabla\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)\nabla\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)^\top}{R^2\tilde{h}_s\left(u_0,\check{\theta}_{R,T}\right)^2}(\boldsymbol{\theta}^*-\boldsymbol{\theta})+o_p(1),
\end{aligned} \tag{A4}$$

where the rates in the last terms follows from Lemma 3, $(T/R)r_T=o(1)$ and $(T/R)|(s-t_0)/T|^2=o(1)$. It follows from Lemma 14(ii) and the assumption that $R/T^{\frac{2}{3}}\to\infty$ that the first term is $O_p\left(T/R^{\frac{3}{2}}\right)=o_p(1)$. Because of $\widehat{\vartheta}_{R,T}-\vartheta=O(r_T)$, $Tr_T/R=o(1)$ and Lemma 3, the second term is also $o_p(1)$. It follows from Lemma 9(iv) and the arguments used in the Proof of Lemma 18 that the last terms converges in probability to

$$\frac{k_1}{2}E\left[\frac{\nabla\ddot{h}_s(u_0,\vartheta^*)\nabla\ddot{h}_s(u_0,\vartheta^*)^\top}{\ddot{h}_s(u_0,\vartheta^*)^2}\right](\boldsymbol{\theta}^*-\boldsymbol{\theta}). \tag{A5}$$

Combining these results, the left hand side converges in probability to (A5). Because the matrix is positive definite, the first order condition is zero if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}^*$. ■

Proof of Theorem 2. The gradient vector of the log of the quasi-density function is

$$\nabla\tilde{l}_{t,T}(u_0,\theta)=-\frac{1}{2}\frac{\nabla\tilde{h}_t(u_0,\theta)}{\tilde{h}_t(u_0,\theta)}+\frac{y_t^2\nabla\tilde{h}_t(u_0,\theta)}{2\tilde{h}_t^2(u_0,\theta)}, \tag{A6}$$

and the Hessian matrix is

$$\nabla^2\tilde{l}_{t,T}(u_0,\theta)=\left(\frac{y_t^2}{\tilde{h}_t(u_0,\theta)}-1\right)\frac{\nabla^2\tilde{h}_t(u_0,\theta)}{2\tilde{h}_t(u_0,\theta)}+\left(1-\frac{2y_t^2}{\tilde{h}_t(u_0,\theta)}\right)\frac{\nabla\tilde{h}_t(u_0,\theta)\nabla\tilde{h}_t(u_0,\theta)^\top}{2\tilde{h}_t^2(u_0,\theta)}. \tag{A7}$$

Applying the mean value theorem to the first order conditions with respect to ϑ,

$$\nabla\tilde{L}_{R,T}\left(u_0,\widehat{\vartheta}_{R,T},\widehat{\boldsymbol{\theta}}_{R,T}\right)=0, \tag{A8}$$

we have

$$\nabla\tilde{L}_{R,T}\left(u_0,\vartheta^*,\widehat{\boldsymbol{\theta}}_{R,T}\right)+\nabla^2\tilde{L}_{R,T}\left(u_0,\tilde{\vartheta}_{R,T},\widehat{\boldsymbol{\theta}}_{R,T}\right)\left(\widehat{\vartheta}_{R,T}-\vartheta^*\right)=0, \tag{A9}$$

where $\tilde{\vartheta}_{R,T}$ lies between $\widehat{\vartheta}_{R,T}$ and ϑ^*. Thus, we can write

$$\widehat{\vartheta}_{R,T}-\vartheta^*=-\left(\nabla^2\tilde{L}_{R,T}\left(u_0,\tilde{\vartheta}_{R,T},\widehat{\boldsymbol{\theta}}_{R,T}\right)\right)^{-1}\nabla\tilde{L}_{R,T}\left(u_0,\vartheta^*,\widehat{\boldsymbol{\theta}}_{R,T}\right). \tag{A10}$$

It follows from Lemma 16 and the consistency of $\widehat{\vartheta}_{R,T}$ that

$$\nabla^2\tilde{L}_{R,T}\left(u_0,\tilde{\vartheta}_{R,T},\widehat{\boldsymbol{\theta}}_{R,T}\right)\xrightarrow{p}-\frac{k_1}{2}\ddot{B}(\vartheta^*). \tag{A11}$$

Two expansions (C.1) and (C.2) show that the difference between h_s and $\tilde{h}_s\left((s-1)/T,u_0\right)$ only comes from the difference between $\theta(\cdot)=\left(\omega(\cdot),\alpha(\cdot),\beta(\cdot)\right)^\top$ and $\bar{\theta}(\cdot,\cdot)=\left(\bar{\omega}(\cdot,\cdot),\bar{\alpha}(\cdot,\cdot),\bar{\beta}(\cdot,\cdot)\right)^\top$ and the difference between h_{t_0-R} and $\tilde{h}_{t_0-R}$. This means that the difference between $\nabla l_{s,T}$ and $\nabla\tilde{l}_{s,T}$ from the difference between $\theta(\cdot)$ and $\bar{\theta}_{(\cdot,\cdot)}$ and the difference between h_{t_0-R} and $\tilde{h}_{t_0-R}$.

Given the initial condition $h_{t_0-R}\left(\theta(\cdot)\right)$, define the Gateaux derivative of $l_s\left(\theta(\cdot)\right)$ at $\left\{\theta(\cdot)\right\}$ in the direction of $\delta(u)=1_{3\times1}$ for $(t_0-R)/T\le u\le(t_0+R)/T$ by:

$$\begin{aligned}\nabla_\delta l_s\left(\theta(\cdot)\right)&=\lim_{c\to0}\frac{l_s\left(\theta(\cdot)+c\delta(\cdot)\right)-l_s\left(\theta(\cdot)\right)}{c}\\&=-\frac{1}{2}\left(1-\frac{y_s^2}{h_s\left(\theta(\cdot)\right)}\right)\lim_{c\to0}\frac{h_s\left(\theta(\cdot)+c\delta(\cdot)\right)-h_s\left(\theta(\cdot)\right)}{c}\\&=-\frac{1}{2}(1-\epsilon_s^2)\frac{\nabla_\delta h_s\left(\theta(\cdot)\right)}{h_s(\theta)}.\end{aligned} \tag{A12}$$

Because ϵ_s is iid with zero mean and unit variance, and because $E\left[\left|\frac{\nabla_\delta h_s\left(\theta(\cdot)\right)}{h_s\left(\theta(\cdot)\right)}\right|\right]<\infty$ by Lemma 3,

$$E\left[\nabla_\delta l_s\left(\theta(\cdot)\right)\right] = 0. \tag{A13}$$

Thus, the bias term can be written as

$$\frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{R}\right) E\left(\nabla \tilde{l}_{s,T}(u_0, \bar{\theta}^*)\right) = \frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{R}\right)\left[E\left(\nabla \tilde{l}_{s,T}(u_0, \bar{\theta}^*)\right) - E\left(\nabla_\delta l_s\left(\theta\left(\frac{s-1}{T}\right)\right)\right)\right].$$

Note that the Gateaux derivatives of $l_s(\theta(\cdot))$ and $h_s(\theta(\cdot))$ are identical to the derivatives of $\tilde{l}_s(u_0, \bar{\theta}((s-1)/T))$ and $\tilde{h}_s(u_0, \bar{\theta}((s-1)/T))$, respectively, with $\bar{\theta}(\cdot)$ and $\tilde{h}_s(u_0, \bar{\theta}((s-1)/T))$ replaced by $\theta(\cdot)$ and $h_s(\theta(\cdot))$, respectively.

Thus, it follows from the above equation that

$$\begin{aligned}
&\frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{R}\right) E\left(\nabla \tilde{l}_{s,T}(u_0, \bar{\theta}^*)\right) \\
&= \frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{R}\right) E\left[\nabla^2 \tilde{l}_{s,T}\left(u_0, \check{\theta}^*\left(\frac{s-1}{T}\right)\right)\right]\left(\bar{\theta}^* - \theta^*\left(\frac{s-1}{T}\right)\right) + o(1) \\
&= -\frac{1}{4R}\sum_{s=2}^{T}\left(\frac{s}{T} - u_0\right)^2 k\left(\frac{s-t_0}{R}\right)\ddot{B}(\vartheta^*)\boldsymbol{\theta}^* + O_p\left(\left(\frac{R}{T}\right)^3\right)
\end{aligned} \tag{A14}$$

where $\check{\theta}((s-1)/T)$ is a point between $\bar{\theta}^*((s-1)/T)$ and $\theta^*((s-1)/T)$. (A14) holds because the derivatives of $E\left[\nabla^2 \tilde{l}_{s,T}(u_0, \theta)\right]$ are bounded uniformly on Θ by Lemma 6.

Thus, the desired result follows from Equations (A10)–(A14) and Lemmas 17. ■

Proof of Theorem 3. We can write

$$\frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{T}\right)\left(\frac{y_s^2}{\hat{\tilde{h}}_s\left(u_0, \hat{\theta}_{R,T}\right)}\right)^2 = \frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{T}\right)\left(\frac{\ddot{y}_s(u_0)^2}{\hat{\tilde{h}}_s\left(u_0, \hat{\vartheta}_{R,T}\right)}\right)^2 + R_{1,R,T}. \tag{A15}$$

Using arguments similar to the one in the Proof of Lemma 18, we can show that the first term converges to $k_1 E\left[\epsilon_s^4\right]$ in probability. Repeating arguments analogous to the Proof of Lemma 7, the second term can be shown to converge in probability to zero.

Next, we write

$$\begin{aligned}
&\frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{T}\right)\frac{\nabla\hat{\tilde{h}}_s\left(u_0, \hat{\theta}_{R,T}\right)\nabla\hat{\tilde{h}}_s\left(u_0, \hat{\theta}_{R,T}\right)^\top}{\hat{\tilde{h}}_s^2\left(u_0, \hat{\theta}_{R,T}\right)} \\
&= \frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{T}\right)\frac{\nabla\hat{\tilde{h}}_s\left(u_0, \hat{\vartheta}_{R,T}\right)\nabla\hat{\tilde{h}}_s\left(u_0, \hat{\vartheta}_{R,T}\right)^\top}{\hat{\tilde{h}}_s^2\left(u_0, \hat{\vartheta}_{R,T}\right)} + R_{2,R,T},
\end{aligned} \tag{A16}$$

where the first term converges in probability to $\ddot{B}(\vartheta^*)$ by using the similar argument in Lemma 18. The remainder term $R_{2,R,T}$ converges to zero in probability by using similar arguments as Lemmas 9(ii) and 16. Since $\ddot{B}(\vartheta^*)$ is positive definite, therefore we have

$$\left[\frac{1}{R}\sum_{s=2}^{T} k\left(\frac{s-t_0}{T}\right)\frac{\nabla\widehat{\widehat{h}}_s\left(u_0,\widehat{\theta}_{R,T}\right)\nabla\widehat{\widehat{h}}_s\left(u_0,\widehat{\theta}_{R,T}\right)^{\top}}{\widehat{\widehat{h}}_s^2\left(u_0,\widehat{\theta}_{R,T}\right)}\right]^{-1} \xrightarrow{p} \frac{1}{k_1}[\ddot{B}(\vartheta^*)]^{-1}. \tag{A17}$$

This completes the proof. ■

References

Amado, C., and T. Teräsvirta. 2013. Modelling Volatility by Variance Decomposition. *Journal of Econometrics* 175: 142–153.

Amado, C., and T. Terásvirta. 2014. Modelling Changes in the Unconditional Variance of Long Stock Return Series. *Journal of Empirical Finance* 25: 15–35.

Andersen, T. G., and T. Bollerselv. 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39: 885–905.

Andreou, E., and E. Ghysels. 2002. Detecting Multiple Breaks in Financial Market Volatility Dynamics. *Journal of Applied Econometrics* 17: 579–600.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song. 2013. *Improving U.S. GDP Measurement: A Forecast Combination Perspective*. New York: Springer, 1–25.

Bauwens, L., A. Preminger, and J. V. K. Rombouts. 2010. Theory and Inference for a Markov Switching GARCH Model. *Econometrics Journal* 13: 218–244.

Berens, T., G. N. Weiß, and D. Wied. 2015. Testing for Structural Breaks in Correlations: Does It Improve Value-at-Risk Forecasting? *Journal of Empirical Finance* 32: 135–152.

Berkowitz, J., P. Chirstoffersen, and D. Pelletier. 2011. Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science* 57: 2213–2227.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Cai, J. 1994. A Markov Model of Unconditional Variance in ARCH. *Journal of Business & Economic Statistics* 12: 309–316.

Cai, Z. 2007. Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors. *Journal of Econometrics* 136: 163–188.

Chen, B., and Y. Hong. 2012. Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression. *Econometrica* 80: 1157–1183.

Chen, B., and Y. Hong. 2016. Detecting for Smooth Structural Changes in GARCH Models. *Econometric Theory* 32: 740–791.

Christoffersen, P. F. 1998. Evaluating Interval Forecasts. *International Economic Review* 39: 841–862.

Christoffersen, P. F. 2003. *Elements of Financial Risk Management*. London: Academic Press.

Christoffersen, P. F., and D. Pelletier. 2004. Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics* 2: 84–108.

Dahlhaus, R., and S. S. Rao. 2006. Statistical Inference for Time-Varying ARCH Processes. *The Annals of Statistics* 1075–1114.

Diebold, F. 1986. Modeling the Persistence of Conditional Variances: A Comment. *Econometric Reviews* 5: 51–56.

Diebold, F. X. 2006. *Elements of Forecasting*, 4th edn. Mason, OH: South-Western College Pub.

Diebold, F. X., and M. Shin. 2019. Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian LASSO and Its Derivatives. *International Journal of Forecasting* 35: 1679–1691.

Engle, R. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of U.K. inflation. *Econometrica* 50: 987–1008.

Engle, R., and J. Rangel. 2008. The spline-GARCH Model for Low Frequency Volatility and Its Global Macroeconomic Causes. *Review of Financial Studies* 21: 1187–1222.

Engle, R. F., and S. Manganelli. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics* 22: 367–381.

Ewing, B., and F. Malik. 2013. Volatility Transmission between Gold and Oil Futures under Structural Breaks. *International Review of Economics & Finance* 25: 113–121.

Francq, C., and J.-M. Zakoïan. 1998. Estimating Linear Representations of Nonlinear Processes. *Journal of Statistical Planning and Inference* 68: 145–165.

Francq, C., and J.-M. Zakoïan. 2000. Estimating Weak GARCH Representations. *Econometric Theory* 16: 692–728.

Fryzlewicz, P., T. Sapatinas, and S. Subba Rao. 2008. Normalized Least-Squares Estimation in Time-Varying ARCH Models. *The Annals of Statistics* 36: 742–786.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance* 48: 1779–1801.

Gray, S. F. 1996. Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics* 42: 27–62.

Haas, M., S. Mittnik, and M. Paolella. 2004. A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics* 2: 493–530.

Hafner, C., and O. Linton. 2010. Efficient Estimation of a Multivariate Multiplicative Volatility Model. *Journal of Econometrics* 159: 55–73.

Hamilton, J., and R. Susmel. 1994. Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics* 64: 307–333.

Hannan, E. J., and J. Rissanen. 1982. Recursive Estimation of Mixed Autoregressive-Moving-Average Order. *Biometrika* 69: 81–94.

Hansen, B. E. 2001. The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity. *Journal of Economic Perspectives* 15: 117–128.

Hillebrand, E. 2005. Neglecting Parameter Changes in GARCH Models. *Journal of Econometrics* 129: 121–138.

Inoue, A., L. Jin, and B. Rossi. 2017. Rolling Window Selection for out-of-Sample Forecasting with Time-Varying Parameters. *Journal of Econometrics* 196: 55–67.

Jin, X., and J. Maheu. 2016. Modeling Covariance Breakdowns in Multivariate GARCH. *Journal of Econometrics* 194: 1–23.

Kim, Y., and E. Hwang. 2018. A Dynamic Markov Regime-Switching GARCH Model and Its Cumulative Impulse Response Function. *Statistics & Probability Letters* 139: 20–30.

Kristensen, D., and Y. Lee. 2019. *Local Polynomial Estimation of Time-Varying Parameters in Nonlinear Models*. London: University College.

Lamoureux, C., and W. Lastrapes. 1990. Persistence in Variance Structural Change and the GARCH Model. *Journal of Business & Economic Statistics* 8: 225–234.

Liu, L. Y., A. J. Patton, and K. Sheppard. 2015. Does Anything Beat 5-Minute RV? A Comparison of Realized Measures across Multiple Asset Classes. *Journal of Econometrics* 187: 293–311.

Mikosch, T., and C. Stărică. 2004. Non-Stationarities in Financial Time Series, the Long Range Dependence and the IGARCH Effects. *Review of Economics and Statistics* 86: 378–390.

Nakajima, J. 2011. Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications. *Monetary and Economic Studies* 29: 107–142.

Nelson, D. B. 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59: 347–370.

Patton, A. J. 2011. Volatility Forecast Comparison Using Imperfect Volatility Proxies. *Journal of Econometrics* 160: 246–256.

Pettenuzzo, D., and A. Timmermann. 2011. Predictability of Stock Returns and Asset Allocation under Structural Breaks. *Journal of Econometrics* 164: 60–78.

Primiceri, G. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies* 72: 821–852.

Pritsker, M. 2006. The Hidden Dangers of Historical Simulation. *Journal of Banking & Finance* 30: 561–582.

Rapach, D., and J. Strauss. 2008. Structural Breaks and GARCH Models of Exchange Rate Volatility. *Journal of Applied Econometrics* 23: 65–90.

Robinson, P. 1989. "Nonparametric Estimation of Time-Varying Parameters." In P. Hackl (ed.), *Statistical Analysis and Forecasting of Economic Structural Change*. Berlin, Germany: Springer. pp. 253–264.

Rohan, N. 2013. A Time Varying GARCH(p, q) Model and Related Statistical Inference. *Statistics & Probability Letters* 83: 1983–1990.

Rohan, N., and T. Ramanathan. 2013. Nonparametric Estimation of a Time-Varying GARCH Model. *Journal of Nonparametric Statistics* 25: 33–52.

Rossi, B. 2013. Exchange Rate Predictability. *Journal of Economic Literature* 51: 1063–1119.

West, K., and D. Cho. 1995. Predictive Ability of Several Models of Exchange Rate Volatility. *Journal of Econometrics* 69: 367–391.

The goal of *Journal of Financial Econometrics* is to reflect and advance the relationship between econometrics and finance, both at the methodological and at the empirical levels. Estimation, testing, learning, prediction and calibration in the framework of asset pricing or risk management represent the core focus. The scope includes topics relating to volatility processes, return modelling, dynamic conditional moments, machine learning, big data, fintech, extreme values, long memory, dynamic mixture models, endogenous sampling transaction data, and microstructure of financial markets.