



**OXFORD JOURNALS**  
OXFORD UNIVERSITY PRESS

---

On the Correlation Structure of Microstructure Noise: A Financial Economic Approach

Author(s): FRANCIS X. DIEBOLD and GEORG STRASSER

Source: *The Review of Economic Studies*, October 2013, Vol. 80, No. 4 (285) (October 2013), pp. 1304-1337

Published by: Oxford University Press

Stable URL: <https://www.jstor.org/stable/43551560>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Oxford University Press is collaborating with JSTOR to digitize, preserve and extend access to *The Review of Economic Studies*

JSTOR

# On the Correlation Structure of Microstructure Noise: A Financial Economic Approach

FRANCIS X. DIEBOLD

*Department of Economics, University of Pennsylvania*

and

GEORG STRASSER

*Department of Economics, Boston College*

*First version received October 2008; final version accepted January 2013 (Eds.)*

We introduce the financial economics of market microstructure to the financial econometrics of asset return volatility estimation. In particular, we derive the cross-correlation function between latent returns and market microstructure noise in several leading microstructure environments. We propose and illustrate several corresponding theory-inspired volatility estimators, which we apply to stock and oil prices. Our analysis and results are useful for assessing the validity of the frequently assumed independence of latent price and microstructure noise, for explaining observed cross-correlation patterns, for predicting as-yet undiscovered patterns, and most importantly, for promoting improved microstructure-based volatility empirics and improved empirical microstructure studies. Simultaneously and conversely, our analysis is far from the last word on the subject, as it is based on stylized benchmark models; it comes with a “call to action” for development and use of richer microstructure models in volatility estimation and beyond.

**Key words:** Realized volatility, Integrated volatility estimation, Market microstructure, High-frequency data, Microstructure noise, Noise correction, Structural volatility estimation

**JEL Codes:** G14, G20, D82, D83, C51

## 1. INTRODUCTION

Recent years have seen substantial progress in asset return volatility measurement, with important applications to asset pricing, portfolio allocation, and risk management. In particular, so-called realized variances and covariances (“realized volatilities”), based on increasingly available high-frequency data, have emerged as central for several reasons.<sup>1</sup> They are, *e.g.*, largely model-free (in contrast to traditional model-based approaches such as GARCH or stochastic volatility), they are computationally trivial, and they are in principle highly accurate.

1. Several surveys are now available, ranging from the comparatively theoretical treatments of Barndorff-Nielsen and Shephard (2007) and Andersen *et al.* (2010), to the applied perspective of Andersen *et al.* (2006, 2013).

A tension arises, however, linked to the last of the above desiderata. Econometric theory suggests the desirability of sampling as often as possible to obtain highly accurate volatility estimates, but financial market reality suggests otherwise. In particular, market microstructure noise (MSN), such as bid–ask bounce associated with ultra-high-frequency sampling, may contaminate the observed price, potentially rendering naively calculated realized volatilities unreliable.

Early work (*e.g.*, Andersen *et al.*, 2001a,b, 2003; Barndorff-Nielsen and Shephard, 2002a,b) addressed the sampling issue by attempting to sample often, but not “too often,” typically resulting in use of 5–30-min returns. Much higher frequency data are usually available, however, so reducing the sampling frequency to insure against MSN discards potentially valuable information.

To use all information, more recent work has emphasized MSN-robust realized volatilities that use returns sampled at very high frequencies. Examples include Zhang *et al.* (2005), Hansen and Lunde (2006), Bandi and Russell (2008), Aït-Sahalia *et al.* (2011), and Barndorff-Nielsen *et al.* (2008, 2011b). That literature is almost entirely *statistical*, however, which is unfortunate because statistics offers little guidance regarding the nature of latent price, MSN, and their interaction. Hence some authors such as Bandi and Russell assume no correlation (perhaps erroneously), whereas in contrast Barndorff-Nielsen *et al.* (2008, 2011a) allow for correlation (perhaps unnecessarily).

To improve this situation, we explicitly recognize that MSN results from the behaviour of economic agents, and we push towards integrating the financial economics of market microstructure with the financial econometrics of volatility estimation. In particular, we explore the implications of microstructure theory for the empirical relationship between latent price and MSN, characterizing the cross-correlation structure between latent price and MSN, contemporaneously and dynamically. We do so in a variety of leading benchmark environments, including Roll (1984), Glosten and Milgrom (1985), Kyle (1985), Easley and O’Hara (1992), and Hasbrouck (2002).<sup>2</sup> Simultaneously and conversely, we also emphasize that our analysis is far from the last word on the subject, as it is based on stylized benchmark models. In part it serves as a “call to action” for development of richer microstructure models that would facilitate more sophisticated analyses.

We proceed as follows. In Section 2 we introduce our general framework, which nests a variety of microstructure models. In Sections 3 and 4 we provide detailed analyses of private-information models, distinguishing two types of latent prices based on the implied level of market efficiency. In Section 5 we discuss the relationship between price change frequency and sampling frequency. Based on this, we suggest several microstructure-based estimators and apply them to stock and oil market data in Section 6. We conclude in Section 7, in which we highlight both the strengths and limitations of our analysis and briefly sketch aspects of extensions beyond the scope of the present article. We provide details of technical results in a web appendix.

## 2. THE FRAMEWORK

We begin in Section 2.1 by introducing a general framework relating latent prices, observed prices, and MSN in a wide range of market-making environments. We then introduce, in Section 2.2, market makers, or—more generally—learning market participants, who are central in the subsequent analyses.

2. See O’Hara (1995) and Hasbrouck (2007) for insightful surveys of the key models, and see Engle and Sun (2007) for a related but ultimately very different perspective based on conditional duration modelling.

### 2.1. Latent prices, observed prices, and microstructure noise

Let  $p_t^*$  denote the (logarithm of the) *strong form efficient price* of some asset in the calendar (or business) time period  $t$ . This price, strictly exogenously changing every  $T$ -th-period, could stem from sampling increments of standard Brownian motion every  $T$  periods, in which case the standard deviation  $\sigma$  would be proportional to  $T$ . At time  $t$ ,  $p_t^*$  is known only to the informed traders, and follows the process:

$$p_t^* = \begin{cases} p_{t-1}^* + \sigma \varepsilon_t, & \forall t = \kappa T, \kappa \in \mathbb{Z} \\ p_{t-1}^*, & \text{otherwise} \end{cases} \quad (2.1)$$

$$\text{with } \varepsilon_t \stackrel{\text{iid}}{\sim} (0, 1). \quad (2.2)$$

This price process is very restrictive. For simplicity of exposition we do not model jumps, time-varying volatility ( $\sigma_t$ ), or time-varying sampling intervals ( $T_t$ ), which are the subject of sophisticated models of market microstructure theory. In all its simplicity, however, this process is the discrete time analogue of the latent price process that estimators of integrated volatility (IV) are based on. As we show later in this article, different assumptions about the nature of the latent price process will lead to different estimates of IV. In particular, the properties of the latent price relevant in many applications depend on the information set. In this article we aim to bridge the gap between market microstructure theory and IV estimation by introducing for the first time a simple price determination framework founded on market microstructure theory to IV estimation.

Microstructure *noise* (MSN) is the difference between the observed market return and the latent return. Instead of *ad-hoc* assumptions about the properties of the *strong form noise*

$$\Delta u_t \equiv \Delta p_t - \Delta p_t^*, \quad (2.3)$$

which are common in the IV estimation literature, we add additional market microstructure that helps explain key properties of MSN.

Let  $q_t$  denote the direction of the trade in period  $t$ , where  $q_t = +1$  denotes a buy,  $q_t = -1$  a sell, and  $q_t = 0$  a no-trade period. Define  $p_t^e$  as the expected efficient price directly before the trade occurs. The *semi-strong form efficient price*, which summarizes the knowledge of the market maker after the trade, is in logarithmic terms

$$\tilde{p}_t^e = p_t^e + \lambda_t q_t, \quad (2.4)$$

where  $\lambda_t \geq 0$  captures the response to asymmetric information revealed by the trade direction  $q_t$ . The admittedly stylized assumption that quantities do not matter for market-maker learning obtains, *e.g.* in a pooling equilibrium of informed with uninformed traders (Kelly and Steigerwald, 2004). It fits the observation that in recent years order-splitting into many small trades has become dominant. Because the estimators we derive rely (at most) on trade direction data, further model detail would not add to our results.

At the beginning of each trading round, additional information about  $p_t^*$  and  $\varepsilon_t$  might be revealed by information diffusion from other sources, *e.g.* other markets. With this information, summarized by  $\omega_t$ , the market maker revises his price expectation for the next period according to

$$p_t^e = \tilde{p}_{t-1}^e + \omega_t. \quad (2.5)$$

In periods in which  $p_{t-1}^*$  becomes public information, (2.5) becomes  $p_t^e = p_{t-1}^* + \tilde{\omega}_t$ . Assuming that the price quotes in logarithmic terms are symmetric around the expected efficient price before

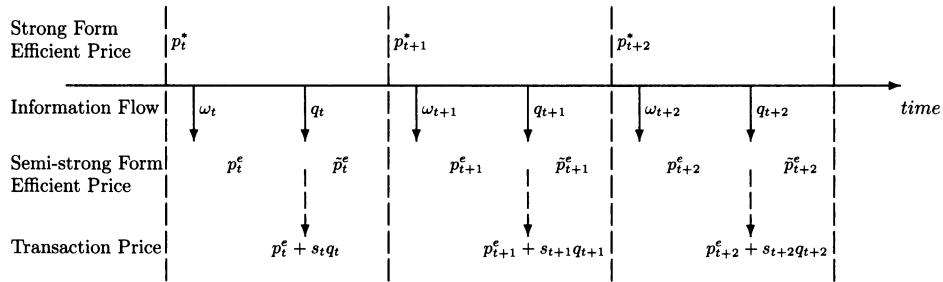


FIGURE 1  
Timing of information and prices

the trade, the observed *transaction price* can be written as

$$p_t = p_t^e + s_t q_t, \quad (2.6)$$

where  $s_t$  is one-half of the spread. In particular, the bid price is  $p_t^{\text{bid}} = p_t^e - s_t$ , the ask price is  $p_t^{\text{ask}} = p_t^e + s_t$ , and the midprice is  $p_t^e$ . These prices and their relationships are illustrated by Figure 1.

Our stylized setup covers three levels of information: full, intermediate (market maker), and public information. Of course, in reality market participants are more heterogeneous with respect to their information sets. Consider, *e.g.*, the difference between traders with and those without access to Nasdaq level II screens. The former traders cannot see the order book, whereas the latter can. We model for concreteness' sake the intermediate price as the market-maker's price. It could, of course, also reflect some other information set, *e.g.* the one of traders with access to a semi-public market information source.

*Strong form efficient returns* in periods  $t = \kappa T$  are therefore

$$\Delta p_t^* \equiv p_t^* - p_{t-1}^* = \sigma \varepsilon_t, \quad (2.7)$$

and 0 in all other periods. *Semi-strong form efficient returns* are

$$\Delta \tilde{p}_t^e \equiv \tilde{p}_t^e - \tilde{p}_{t-1}^e = \lambda_t q_t + \omega_t, \quad (2.8)$$

and *semi-strong form noise* is accordingly

$$\Delta \tilde{u}_t \equiv \Delta p_t - \Delta \tilde{p}_t^e. \quad (2.9)$$

We use the term “*latent price*” as a general term comprising both types of efficient prices. The two latent prices defined here are conceptually very distinct and appeal to distinct audiences. For example, on the one hand, a pure theorist may want to understand the properties of the full-information price, and is thus interested in an estimate of the volatility of the strong form efficient return (2.7). On the other hand, a market maker may need a volatility measure to calculate his risk exposure, thus his relevant price for the asset is  $\tilde{p}_t^e$ , the price at which he keeps the asset on his accounts. It is the volatility of (2.8), and not of (2.7), that affects his balance sheet.

Semi-strong form noise (2.9) differs fundamentally in its cross-correlation properties from (2.3). It is therefore essential for a researcher to be clear what type of latent price the object of interest is, because each requires different procedures to remove MSN appropriately.

Observed *market returns* are

$$\Delta p_t \equiv p_t - p_{t-1} = \Delta p_t^e + s_t q_t - s_{t-1} q_{t-1}.$$

We assume throughout that market conditions are stable so that the noise processes,  $\Delta u_t$  and  $\Delta \tilde{u}_t$ , are covariance stationary.

A convenient estimator of the variance of the strong form efficient return,  $\sigma^2$ , and therefore of the IV of the underlying continuous time process, is the realized volatility (RV) as in Andersen *et al.* (2001b). RV during the time interval  $[0, \bar{T}]$  is defined as the sum of squared market returns over the interval, *i.e.* as

$$\text{Var}(\Delta p_t) = \sum_{t=1}^{\bar{T}} \Delta p_t^2.$$

In the presence of MSN, the RV is generally a biased estimate of  $\sigma^2$ . To see this, decompose the noise into two components, one uncorrelated and one correlated with the latent price, so that  $\Delta u_t = \Delta u_t^u + \Delta u_t^c$ . The uncorrelated component,  $\Delta u_t^u$ , reflects *e.g.* the bid–ask bounce in a market populated with uninformed traders only. The correlated component,  $\Delta u_t^c$ , reflects *e.g.* the effect of asymmetric information. RV can now be decomposed—here shown for the strong form efficient price—as

$$\begin{aligned} \text{Var}(\Delta p_t) &= \text{Var}(\Delta p_t^* + \Delta u_t^u + \Delta u_t^c) \\ &= \sigma^2 + \text{Var}(\Delta u_t^u) + \text{Var}(\Delta u_t^c) + 2\text{Cov}(\Delta p_t^*, \Delta u_t^c). \end{aligned}$$

The bias of RV can stem from any of the last three terms, which are all non-zero in general. IV estimation under the independent noise assumption accounts for the second and third positive terms, but ignores the last term, which is typically negative (Hansen and Lunde, 2006). Correcting the estimates for independent noise only, always reduces the volatility estimate. But because such a correction ignores the last term, which is the second channel through which asymmetric information affects the IV estimate, the overall reduction might be too much. Further, serial correlation of noise, or equivalently a cross-correlation between noise and latent returns at non-zero displacement, requires the use of robust estimators for both the variance and the covariance terms. In this article we determine what correlation and serial correlation market microstructure theory predicts, and how market microstructure theory can be useful for improving IV estimates.

## 2.2. Introducing markets and market makers

Whereas the strong form efficient price (2.1) is an exogenous stochastic process, the semi-strong form efficient price (2.4) and the transaction price (2.6) are an outcome of the market participants' optimizing behaviour. Key mechanisms of the data generator—the financial market—are often observable and allow inferring properties of these price series. This is what we exploit in this article.

Observed transaction prices are determined by the information available about the strong form efficient price and the market participants' response to this information. Three features of the information process matter in particular: First, information content, second, the diffusion speed of information into public knowledge, and third, the duration of its validity.

We focus here on a stylized limit-order market, populated by informed and uninformed traders. Market makers are the counterparty of all trades. Each trading round they quote price  $p_t^e$  and spread

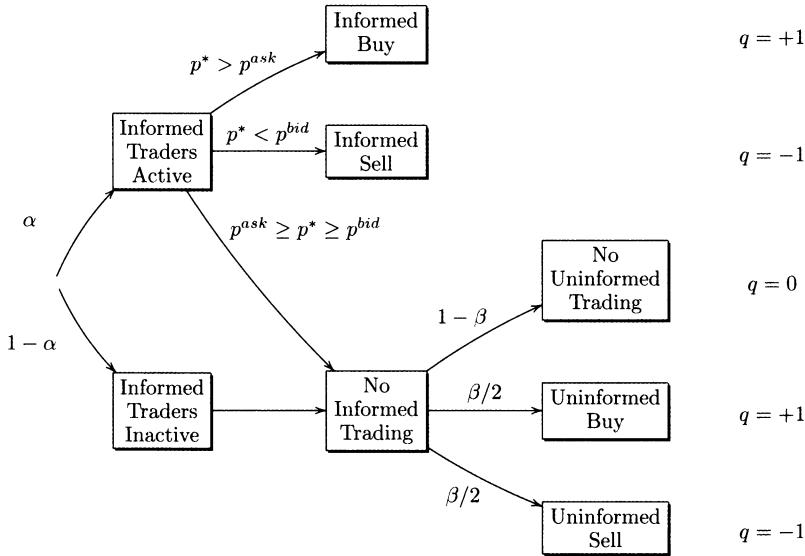


FIGURE 2  
Sequence of informed and uninformed trading decisions

$s_t$  for one unit of the asset. Thereafter, as shown in Figure 2, informed traders screen the market with probability  $\alpha$  for profitable trading opportunities. They buy if  $p_t^* > p_t^{\text{ask}}$ , sell if  $p_t^* < p_t^{\text{bid}}$ , and refuse to trade otherwise. In periods of no informed trade, uninformed traders trade instead with probability  $\beta$ , buying and selling with equal probability.

When trading with an informed trader the market maker always loses. His expected loss is

$$L_n \left[ p_t, F(\cdot; \underline{p}^*, \bar{p}^*) \right] = - \int_{\underline{p}}^{\bar{p}} |(p_t - p_t^*) E(q_t | p_t^*, p_t, s_t)|^n f(p_t^*) dp_t^*, \quad (2.10)$$

where  $n$  reflects the risk aversion of the market maker and  $E(q_t | p_t^* + s_t < p_t^*) = \alpha$ ,  $E(q_t | p_t^* - s_t > p_t^*) = -\alpha$ , and  $E(q_t | p_t^* - s_t \leq p_t^* \leq p_t^* + s_t) = 0$ .  $F(\cdot)$  and  $f(\cdot)$  denote the cumulative distribution function and density with support  $[\underline{p}, \bar{p}]$  of the market-maker's belief about the latent price. Similar to Aghion *et al.* (1991), the market maker faces a tradeoff between avoiding losses today and learning quickly.<sup>3</sup>

Because price quotes are only for limited quantities, the market maker can update his price quote after every trade and his risk exposure is usually small. Accordingly, we assume risk neutrality ( $n=1$ ) throughout the article, and relegate the implications of risk aversion to Section 3.3.2. As shorthand notation for the probability of a trade we define

$$\phi_t = E(q_t^2) = E[\text{Prob}(|q_t| = 1)] = \beta + (1 - \beta)\alpha [1 - F(p_t^* + s_t) + F(p_t^* - s_t)].$$

Note that the model can be recast in tick-time by setting  $\phi_t = 1 \forall t$ . We add the following assumption, which simplifies the model without affecting its basic behaviour.

3. Diebold and Strasser (2010) describe the market setup and market-maker problem in more detail.

**Assumption.** *Ex ante, a buy and a sell is equally likely, so that  $E(q_t) = 0$ . There is no “momentum” in uninformed trading, and thus trades are serially uncorrelated beyond the time of a strong form efficient price change, i.e.  $E(q_{\kappa T+\tau_1} | q_{\kappa T-\tau_2}) = 0 \forall \kappa, \tau_1 \in \mathbb{N}_0, \forall \tau_2 \in \mathbb{N}$ .*

In the following Sections 3 and 4, we look at specializations of this general market-maker problem and examine their effect on the cross-correlation function. For both strong form and semi-strong form efficient returns we first examine the multiperiod case, where private information is not revealed until after many periods. We then specialize to the one-period case, a case where private information becomes public, and worthless, after only one period, where we specifically address the effect of risk aversion.

### 3. RETURN-NOISE CORRELATIONS IN FINANCIAL ECONOMIC ENVIRONMENTS I: STRONG FORM EFFICIENT PRICES

We focus in this article on the *cross-correlation* between latent returns and noise contemporaneously and at all displacements. Throughout, we refer to this quantity simply as the “cross-correlation”. In this section we characterize cross-correlations between strong form efficient returns (2.7) and the corresponding noise (2.3) in various market settings. To study the effect of one efficient price change in isolation, suppose for now that the strong form efficient price changes once every  $T$  periods at a commonly known time at which the previous price becomes public knowledge.

#### 3.1. *The general multi-period case*

The cross-correlations, as shown in Web Appendix A.1.2, follow directly from the price and noise processes. The contemporaneous cross-covariance is

$$\text{Cov}(\Delta p_t^*, \Delta u_t) = \frac{\sigma}{T} [s_0 E(q_0 \varepsilon_0) - \sigma + E(\omega_0 \varepsilon_0)]. \quad (3.11)$$

For cross-covariance at higher displacements  $\tau \in [1; T-1]$  we get

$$\text{Cov}(\Delta p_{t-\tau}^*, \Delta u_t) = \frac{\sigma}{T} [(\lambda_{\tau-1} - s_{\tau-1}) E(q_{\tau-1} \varepsilon_0) + s_\tau E(q_\tau \varepsilon_0) + E(\omega_\tau \varepsilon_0)], \quad (3.12)$$

for cross-covariance at displacement  $T$ , which is when private information becomes public,

$$\text{Cov}(\Delta p_{t-T}^*, \Delta u_t) = \frac{\sigma}{T} \left[ \sigma - s_{T-1} E(q_{T-1} \varepsilon_0) - \sum_{i=0}^{T-2} \lambda_i E(q_i \varepsilon_0) - \sum_{i=0}^{T-1} E(\omega_i \varepsilon_0) \right], \quad (3.13)$$

and for all higher-order displacements  $\tau > T$

$$\text{Cov}(\Delta p_{t-\tau}^*, \Delta u_t) = 0. \quad (3.14)$$

Combining (3.11) with the noise variance derived in the web appendix gives the contemporaneous cross-correlation

$$\text{Corr}(\Delta p_t^*, \Delta u_t) = \frac{s_0 E(q_0 \varepsilon_0) - \sigma + E(\omega_0 \varepsilon_0)}{\sqrt{T} \text{Var}(\Delta u_t)}. \quad (3.15)$$

All other cross-correlations can be obtained analogously.

TABLE 1  
*Cross-correlations between  $\Delta p_t^*$  and MSN in multi-period models*

	$p_t^*$ Martingale	Signal	Traders strategic	$\rho_0$	$\rho_\tau$ $\tau \in [1, T-1]$	$\rho_T$	$\rho_\tau$ $\tau > T$
Roll	Yes	None	N.A.	$\rho_0 < 0$	0	$-\rho_0$	0
Glosten-Milgrom	Yes	Certain/noisy	No	$\rho_0 < 0$	$\rho_{\tau-1} > \rho_\tau > 0$	$\rho_T > 0$	0
Easley-O'Hara	No	Noisy	No	$-\frac{1+e^{-r(T-1)}}{2\sqrt{K(r,T)}}$	$\frac{-e^{-r\tau}+e^{-r(\tau-1)}}{2\sqrt{K(r,T)}}$	$\frac{e^{-r(T-1)}}{2\sqrt{K(r,T)}}$	0
Kyle	Yes	Noisy	Yes	$-\sqrt{\frac{T}{T^2+1}}$	$\sqrt{\frac{1}{T(T^2+1)}}$	$\sqrt{\frac{1}{T(T^2+1)}}$	0

The term  $E(q_\tau \varepsilon_0)$  enters the expressions for the cross-covariance (3.11)–(3.13) linearly but the denominator of the cross-correlation under a square root. Because this term decreases in the share of uninformed trades, the contemporaneous cross-correlation is the smaller, the less informed traders are active. In absence of both informed traders ( $E(q_\tau \varepsilon_0)=0$ ) and of extra information ( $E(\omega_0 \varepsilon_0)=0$ ), the market microstructure reduces to a bid–ask bounce, as in Roll (1984). Even in this case, shown in the first row of Table 1, the latent price and noise are not independent. The contemporaneous cross-correlation (3.15) is negative, the cross-correlations at displacement  $T$  is positive, and all other cross-correlations are 0.

Because of order-splitting, effective spreads have become very small for liquid assets. If no extra information is available and the spread sufficiently small, then the contemporaneous cross-correlation is negative even in presence of informed traders, because  $p_t$  does not react sufficiently to  $\Delta p_t^*$ . It is strictly larger than negative one, because the delayed response of  $\Delta p_t$  to  $\Delta p_{t-\tau}^*$  generates cyclical noise with—absent other market microstructure effects—up to twice the variance of  $\Delta p_t^*$ . Likewise, if the spread roughly matches the adverse selection coefficient, by (3.12) the cross-correlations at displacements 1 up to  $T-1$  are positive, which reflects that the more the market maker learns, the closer  $p_t$  gets to  $p_t^*$ , and the closer noise shrinks to 0. If, additionally, the adverse selection coefficient  $\lambda$  and extra information  $\omega$  in all periods are sufficiently small, i.e. if some private information persists until period  $T$ , then by (3.13) the cross-correlation at displacement  $T$  is positive as well.

In general, however, the sign of the cross-correlations depends on the behaviour of market makers and traders. We now turn to models that allow us to introduce these explicitly.

### 3.2. Special multi-period cases of informed trading

The market maker does not observe the strong form efficient price,  $p_t^*$ , directly, but only signals which allow him to narrow down the range of the current  $p_t^*$  level. He learns over time “by experimentation” the informed traders’ private information about  $p_t^*$  by quoting prices and observing the resulting trades (Aghion *et al.*, 1991, 1993). The market maker has an incentive to find out  $p_t^*$ , because he loses in every trade with an informed trader. His optimization task is to quote prices that minimize his losses while learning about  $p_t^*$  as quickly as possible. We will see that rational behaviour of market participants and the market setup pin down the cross-correlation sign pattern. Only the absolute value of the cross-correlation differs depending on how market participants interact.

The recursive problem of the market maker is hard to solve, and has in general no closed form policy functions for bid and ask prices. Therefore we follow the market microstructure literature by discussing interesting polar cases, which can be solved because  $f(p_t^*)$  is degenerate. In particular, we limit our discussion to the midprice under a constant spread.

**3.2.1. No strategic traders.** Consider first a market in which the market maker observes a noisy signal of whether  $p_t^*$  has changed, and in which traders do not behave strategically. The market maker has to learn both about the quality of the signal and about the latent price. A useful illustration is the stylized model of Easley and O'Hara (1992). As in our general setup in Section 2.2 informed traders are active with probability  $\alpha$ . The strong form efficient price, which is not a martingale here, can assume one of two possible levels:  $p_t^* = \underline{p}^*$  or  $p_t^* = \bar{p}^* > \underline{p}^*$ . These levels, as well as the probability  $\gamma$  of  $p_t^* = \underline{p}^*$ , are publicly known, but the actual realization of  $p_t^*$  is not.<sup>4</sup>

The direction-of-trade signal,  $q_t$ , is thereby noisy in two ways. Not only does the market maker not know if a specific trade originates from informed traders, thereby being informative; the market maker does not even know if there *are* any informed traders. He learns by updating in a Bayesian manner his belief about the probabilities that nobody observed a signal, that informed traders observed  $p_t^* = \bar{p}^*$ , or that they observed  $p_t^* = \underline{p}^*$ , using his information set of all previous quotes and trades.

Easley and O'Hara (1992) show that bid and ask prices, and therefore transaction prices, converge exponentially to the strong form efficient price in calendar time. Market makers sampling in tick-time have the same correlation pattern, but a lower learning rate, because they miss the no-trade periods. These no-trade intervals contain information about  $p_t^*$ , because they lower the probability that informed traders are active.<sup>5</sup>

The following proposition gives the cross-correlations in Easley and O'Hara (1992)-type models. It considers only the dominant exponential learning pattern, and ignores lower order terms which disappear at faster rates as  $\tau$  gets large.

**Proposition 1 (Cross-correlations in the Easley-O'Hara model)** *The contemporaneous cross-correlation in the Easley and O'Hara (1992) model with learning rate  $r$  is*

$$\text{Corr}(\Delta p_t^*, \Delta u_t) = -\frac{1+e^{-r(T-1)}}{2\sqrt{K}} < 0,$$

and the cross-correlations at sufficiently large non-zero displacements follow:

$$\text{Corr}(\Delta p_{t-\tau}^*, \Delta u_t) = \frac{e^r - 1}{2\sqrt{K}} e^{-r\tau} > 0, \quad \forall \tau \in [1, T-1]$$

$$\text{Corr}(\Delta p_{t-T}^*, \Delta u_t) = \frac{e^{-r(T-1)}}{2\sqrt{K}} > 0,$$

where  $K = K(r, T)$ .

*Proof* The proofs to all propositions are collected in Web Appendix A. ||

As before, the contemporaneous correlation is negative. It approaches its minimum for small learning rates and frequent latent price changes. The market-maker learning imposes that the cross-correlation of the strong form efficient price decays geometrically to 0 until  $\tau = T$ .<sup>6</sup>

4. The case of signal certainty, which implies the absence of any uninformed traders, is trivial here: Because  $p_t^*$  can assume only one of two price levels, the first trade reveals the true strong form efficient price. Until the first trade occurs, the expected efficient price is  $\gamma \underline{p}^* + (1-\gamma) \bar{p}^*$ .

5. A variation of this setup is the model of Diamond and Verrecchia (1987), where short selling constraints cause periods of no trading to be a noisy signal of a low latent price.

6. This decay pattern is not unique to the Easley and O'Hara (1992)-model. Glosten and Milgrom (1985) show more generally that if learning is costless, the expectations of market makers and traders necessarily converge as the

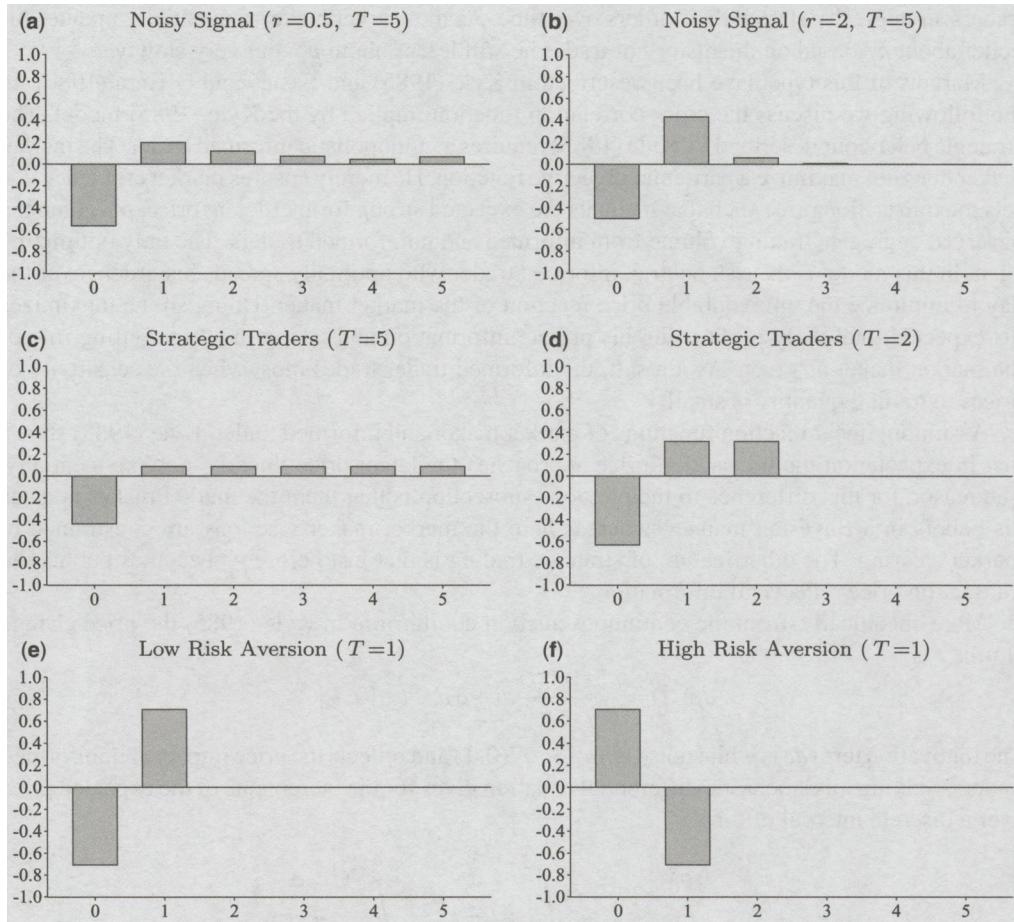


FIGURE 3  
Cross-correlation functions  $\rho_\tau$  of the strong form efficient price

We graph this cross-correlation function in the first row of Figure 3, in the upper left panel for a learning rate of  $r=0.5$ , and in the upper right panel for a faster learning rate of  $r=2$ . Often, optimal learning stops before  $p_t^*$  is reached (Aghion *et al.*, 1991), e.g. if the spread is large or if market-maker risk aversion is small. In that case the cross-correlations cut off at some  $\tau < T$ .

**3.2.2. Strategic traders.** Because the market maker cannot distinguish informed from uninformed trades, informed traders can act strategically. Informed traders aim to make the signals about  $p_t^*$  conveyed by their orders as noisy as possible, while still executing the desired trades. By mimicking uninformed traders they keep the market maker unaware of the change in  $p_t^*$ . Because the market maker observes the order flow and uses it to detect informed trading, the informed

number of trades increases. Because of the uncertainty of whether a trade reflects information or just noise, the market maker faced with a noisy signal adjusts only *partially*. Therefore, whereas the cross-correlations under a noisy signal have the same signs as under signal certainty, their absolute values are dampened towards 0.

traders strategically stretch their orders over time. As the market maker sequentially updates his belief about  $p_t^*$  based on the history of trades he still learns about  $p_t^*$ , but very slowly.

Markets of this type have been described in Kyle (1985) and Easley and O'Hara (1987). In the following we discuss the cross-correlation function implied by the Kyle (1985) model. The strategic behaviour described by Kyle (1985) requires a monopolistic informed trader. The market maker does not maximize a particular objective function. He merely ensures market efficiency, *i.e.* sets the transaction price such that it equals the expected strong form efficient price,  $p_t^e$ , given the observed aggregate trading volume from informed and uninformed traders. The only optimizing agent in this model is the risk neutral, informed trader who optimally spreads his orders over the day to minimize the unfavourable price reaction of the market maker. Doing so, he maximizes his expected total daily profit using his private information and taking the price setting rule of the market maker as given. As a result, the informed trader trades most when the sensitivity of prices to trading quantity is small.

Assuming linear reaction functions of market maker and informed trader, Kyle (1985) shows that in expectation the transaction price approaches the latent price linearly, not exponentially. The reason for this difference to the previous subsection is that there the market maker updates his beliefs in a Bayesian manner, whereas here the market-maker's actions are constrained to market clearing. The other feature of strategic trading is that just before  $p_t^*$  becomes public, the transaction price reflects all information.

More specifically, from the continuous auction equilibrium in Kyle (1985) the price change at time  $t$  is

$$dp^e(t) = \frac{p^* - p^e(t)}{T-t} dt + \sigma dz, \quad t \in [0, T].$$

The innovation term  $dz$  is white noise with  $dz \sim N(0, 1)$  and reflects the price impact of uninformed traders. Solving this stochastic differential equation gives for the increments of the expected price over a discrete interval of time

$$\Delta p_\tau^e = \frac{\Delta p_0^*}{T} + (T - \tau) \int_{\tau-1}^\tau \frac{\sigma}{T-s} dz - \int_0^{\tau-1} \frac{\sigma}{T-s} dz. \quad (3.16)$$

This implies the following cross-correlations:

**Proposition 2 (Cross-correlations in the Kyle model)** *The contemporaneous cross-correlation in Kyle (1985) is*

$$\text{Corr}(\Delta p_t^*, \Delta u_t) = -\sqrt{\frac{T}{T^2+1}},$$

*the cross-correlations at displacements  $\tau \in [1; T]$  are*

$$\text{Corr}(\Delta p_{t-\tau}^*, \Delta u_t) = \sqrt{\frac{1}{T(T^2+1)}},$$

*and all higher order cross-correlations are 0.*

By Proposition 2 the cross-covariance at non-zero displacements is a positive constant. It is positive because of market-maker learning. It is constant because of the strategic behaviour of traders, which spread their informative trades over time. The more periods, the more pronounced is the negative contemporaneous cross-correlation, and the smaller are the cross-correlations at non-zero displacements. The second row of Figure 3 plots this cross-correlation function under

modestly frequent changes in the latent price ( $T=5$ ) in the left panel, and for more frequent changes ( $T=2$ ) in the right panel.

Table 1 compares the cross-correlation patterns of standard multiperiod market microstructure models: The Roll (1984) model in row 1, the Glosten and Milgrom (1985) model in row 2, the Easley and O'Hara (1992) model in row 3, and the Kyle (1985) in row 4, which includes oscillating, linearly decaying and exponentially decaying patterns.

### 3.3. One-period case

In this section we return to the general latent price process, and consider the extreme case that  $p_t^*$  becomes public information at the end of each period, *i.e.*  $\omega_t = p_{t-1}^* - \tilde{p}_{t-1}^e$  and  $T=1$ . This allows us to investigate the impact of risk aversion on the cross-correlation pattern. Because  $p_{t-1}^*$  is now known when the market maker decides on  $p_t$ , it removes any incentive for informed traders to behave strategically. They therefore react immediately, which implies that  $E(q_{t-\tau}\varepsilon_t)=0 \forall \tau \neq 0$  and that all trades are serially uncorrelated, *i.e.*  $E(q_t|q_{t-1})=0$ . For the market maker all periods are identical, and therefore the spread and reaction parameters are both constant over time, *i.e.*  $s_t=s$  and  $\lambda_t=\lambda \forall t$ .

The cross-correlation function inherits its shape from (3.11)–(3.14). At displacement 1 it has the opposite sign and same absolute value as contemporaneously, and it is 0 at displacements larger than 1. In order to pin down the value of the contemporaneous cross-correlation, we now turn to specific models.

**3.3.1. No market maker information.** We start with our baseline assumption that the market maker at time  $t$  has no information whatsoever about  $\Delta p_t^*$ . Plugging  $T=1$ ,  $s_t=s$ , and  $\lambda_t=\lambda$ , and thus  $\phi_t=\phi$ , into the general multiperiod results of Section 3.1 gives

**Proposition 3 (Strong form cross-correlation, one-period model)**

$$\text{Corr}(\Delta p_t^*, \Delta u_t) = \frac{1}{\sqrt{2}} \frac{sE(q_t\varepsilon_t) - \sigma}{\sqrt{\phi s^2 + \sigma^2 - 2s\sigma E(q_t\varepsilon_t)}}, \quad (3.17)$$

$$\text{Corr}(\Delta p_{t-1}^*, \Delta u_t) = -\text{Corr}(\Delta p_t^*, \Delta u_t).$$

If there is trading in every period ( $\beta=1$ , and thus  $\phi=1$ ), then the cross-correlation (3.17) is bounded from above and below by

**Proposition 4 (Bounds of contemporaneous cross-correlation)**

$$-\frac{1}{\sqrt{2}} \leq \text{Corr}(\Delta p_t^*, \Delta u_t) \leq 0.$$

The cross-correlation reaches the lower bound for zero spread. Thus the cross-correlation is highest for midprices, and very small spreads. The contemporaneous cross-correlation for midprices is negative, because  $p_t^e$  does not react instantaneously to the change in the strong form efficient price in the same period. This is an instance of the price stickiness that Bandi and Russell (2006) show to generate “mechanically” a negative contemporaneous cross-correlation. It differs

TABLE 2  
*Cross-correlations between latent prices and MSN in one-period models*

Latent price	$s$	$\lambda$	Loss function	$\rho_0$	$\rho_1$	$\rho_\tau$ $\tau > 1$
$p_t^*$	0	Any	Any	$-\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$	0
	$\geq 0$	Any	Any	$-\frac{1}{\sqrt{2}} \leq \rho_0 < 0$	$-\rho_0$	0
	$\geq 0$	Any	High $n +$ extra info	$\rho_0 > 0$	$-\rho_0$	0
$\tilde{p}_t^e$	$\geq 0$	$\lambda^{opt}$	Quadratic	$-\frac{1}{\sqrt{2}} \leq \rho_0 \leq \frac{1}{\sqrt{2}}$	$-\rho_0$	0
	$\in [0, \lambda[$	$> \frac{\lambda^{opt}}{2}$	Any	$\rho_0 < 0$	$\rho_1 > 0$	0
	$\in [0, \lambda[$	$< \frac{\lambda^{opt}}{2}$	Any	$\rho_0 > 0$	$\rho_1 > 0$	0
	$\lambda$	Any	Any	0	0	0
	$\geq \lambda$	$> \frac{\lambda^{opt}}{2}$	Any	$\rho_0 > 0$	$\rho_1 < 0$	0
	$\geq \lambda$	$< \frac{\lambda^{opt}}{2}$	Any	$\rho_0 < 0$	$\rho_1 < 0$	0

from negative unity because transaction prices move in adjustment to the strong form efficient return of one-period earlier.

We summarize these results in the upper two rows of Table 2. Compared to the multiperiod case in Table 1 the absolute value of the cross-correlation at lag 1 is large, because all information is revealed. Cross-correlations at any displacement beyond 1 are, in contrast, necessarily all 0.

**3.3.2. Incomplete market-maker information and risk aversion.** Throughout this article we assume a risk-neutral market maker. In this subsection we lift this assumption, which can be justified in times of market turbulence. If extreme events occur, strong form efficient prices become highly correlated across assets, in particular stocks. Although the market maker is bound by his quote only up to a fixed quantity on an individual stock, the total exposure of a market maker that has quotes outstanding in many markets might be non-trivial.

Without information about  $\Delta p_t^*$  risk aversion does not change the market-maker behaviour. With extra information, however, the market maker adjusts his quotes before informed traders can take advantage of it. We show in Web Appendix A.6.1 that under some regularity conditions risk aversion paired with extra information, e.g. about the direction of the change in the latent price,  $\{\text{sgn}(\varepsilon_t)\}$ , can invert the cross-correlation pattern.

The sensitivity of the expected loss,  $L_n[p_t, F(\cdot, \underline{p}^*, \bar{p}^*)]$ , to the support of  $p_t^*$ , i.e., to  $\underline{p}^*$  and  $\bar{p}^*$ , increases with risk aversion,  $n$ . For some values of  $n$ , explicit solutions to the market-maker problem are available. A well-known result is that the optimal choice for a risk-neutral market maker ( $n=1$ ) is to set  $p_t^e$  equal to the median of  $f(\cdot)$ , and for a modestly risk averse market maker ( $n=2$ ) to the mean. An extremely risk averse ( $n \rightarrow \infty$ ) market maker follows the most robust pricing role possible: He minimizes his expected loss at the price in the middle of the support of  $f(\cdot)$ , i.e.  $p_t = \frac{\underline{p}^* + \bar{p}^*}{2}$ .

Figure 4, which plots the transaction price as a function of risk aversion  $n$ , illustrates this increasing sensitivity. For a right-skewed distribution  $f(\cdot)$  with infinite support, namely the halfnormal distribution,  $p^e(n)$  increases in  $n$ , starting from the median for  $n=1$ , monotonically without bound. If, in contrast,  $f(\cdot)$  has finite support, then  $p^e(n)$  increases from the median monotonically towards a finite asymptote  $p^e(\infty)$ . This is shown in the right panel of Figure 4

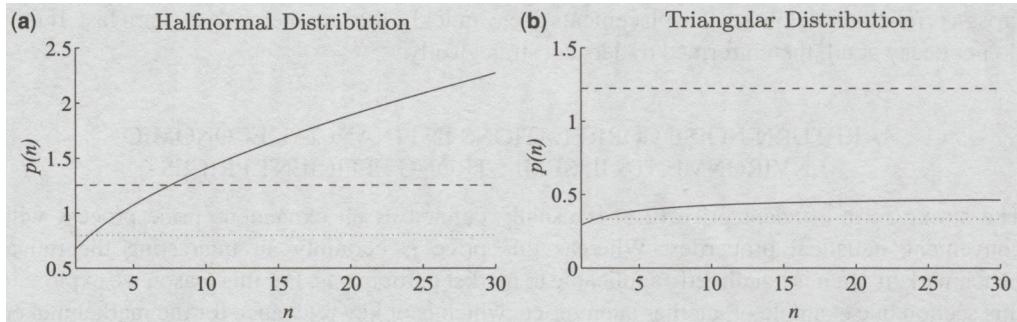


FIGURE 4  
Optimal mid-price for right-skewed expected latent price distributions

for the right-triangular distribution defined on  $[0, 1]$ . This has implications for the possible cross-correlations:

**Proposition 5 (Cross-correlation under market-maker information)** *If the distribution of the expected latent price with ex-ante support  $[p_t^*, \bar{p}_t^*]$  satisfies*

$$\left[ \frac{p_t^* + \bar{p}_t^*}{2} - p_{t-1}^* \right] \text{sgn}(\varepsilon_t) > s + \frac{\sigma}{E(|\varepsilon_t|)}, \quad (3.18)$$

then  $\exists n_0 > 1$  such that  $\forall n > n_0$  it holds that  $\text{Corr}(\Delta p_t^*, \Delta u_t) > 0$ .

Condition (3.18) holds, e.g., for normally distributed, but not for tent distributed  $\Delta p_t^*$ . This is reflected in Figure 4, where the price in the left panel quickly reaches the cutoff  $\frac{\sigma}{E(|\varepsilon|)}$ , plotted as dashed line, whereas in the right panel it never does.

Comparing these results in the third row of Table 2 with the other models, it appears that even though the contemporaneous cross-correlation can be positive for high risk aversion levels, the usual case is that it is negative. For the halfnormal distribution, e.g., we need a rather high risk aversion of  $n \geq 8$ . Nevertheless, changes in risk aversion of the market maker have a distinctive impact on the cross-correlation. Hansen and Lunde (2006) note as their “Fact IV” that “the properties of the noise have changed over time.” Because they base this observation on a comparison of year 2000 with year 2004 it is well possible that the underlying cause is a change in risk aversion.

The link between properties of noise and risk aversion offers itself as a way to estimate the time path of risk aversion from the cross-correlation pattern of transaction prices. In stable periods with low risk aversion the contemporaneous cross-correlation is negative, but as uncertainty shoots up, contemporaneous cross-correlation shoots up with it. In periods of crisis this can lead to the extreme case of an inverted cross-correlation pattern that the lower row of Figure 3 illustrates. It shows the typical cross-correlation pattern of strong form efficient prices in a one-period model with modest risk aversion on the left, and under higher risk aversion on the right.

In summary we have shown in this section that many market properties leave their mark on the cross-correlation pattern: The displacement beyond which correlation is 0 gives an indication of the frequency of information events. The larger the correlation is in absolute value terms the fewer uninformed trades occur in the market. If contemporaneous strong form cross-correlation is positive, then market makers are very risk averse and have access to extra information. If the

cross-correlations at non-zero displacements decay quickly, then market makers learn fast. If they do not decay at all, then informed traders act strategically.

#### 4. RETURN-NOISE CORRELATIONS IN FINANCIAL ECONOMIC ENVIRONMENTS II: SEMI-STRONG EFFICIENT PRICES

The strong form efficient price (2.1) is usually defined as an exogenous price process with convenient statistical properties. Whereas this price is certainly an interesting theoretical benchmark, it often is not directly applicable to market participants. For this reason we explore in this section one example of another latent price, which is of key relevance for the market maker. We call this price,  $\tilde{p}_t^e$  given by (2.4), the semi-strong efficient price, noting that each market participant has his own, depending on his respective information set. Equivalently this setup can be seen as an endogenous latent price process, determined by an exogenous trading process  $q_t$ . It is closely related to the “generalized Roll model” in Hasbrouck (2007).

##### 4.1. Multi-period case

Similar calculations as in the previous sections (see Web Appendix A.1.3) reveal that the cross-correlations for semi-strong efficient prices stem from a gap between the spread,  $s_t$ , and the adverse selection parameter,  $\lambda_t$ . Such a gap can result from processing costs ( $s_t > \lambda_t$ ), from legal restrictions ( $s_t < \lambda_t$ ), or merely from suboptimal behaviour of the market maker.

Noisy signals or strategic behaviour do not affect the semi-strong form cross-correlations, as for example in Easley and O’Hara (1992), where prices are semi-strong efficient by definition. Under semi-strong market efficiency ( $s_t = \lambda_t \forall t$ ) the cross-correlation function is 0 for all displacements.

The Kyle (1985) model assumptions  $\lambda_t = \lambda$  and  $s_t = s \forall t$  give

$$\text{Cov}(\Delta\tilde{p}_{t-\tau}^e, \Delta\tilde{u}_t) = \frac{\lambda(\lambda - s)}{T} \left\{ E(q_{T-\tau} q_{T-1}) + \sum_{i=\tau}^{T-1} [E(q_{i-\tau} q_{i-1}) - E(q_{i-\tau} q_i)] \right\}.$$

If  $\lambda = 0$ , then this cross-correlation function is flat at 0. If instead  $E(q_{i-\tau} q_i)$  is a positive constant between the time of the latent price change and its public announcement, then the cross-correlation function is flat and proportional to  $\frac{\lambda(\lambda - s)}{T}$ . If  $E(q_i q_j) > E(q_{i-\tau} q_j) > 0 \forall i \leq j, \forall \tau > 0$ , the cross-correlation decreases in  $\tau$ .

##### 4.2. One-period case

The simpler case of markets in which all information is revealed after one period without any extra information, *i.e.*

$$\Delta\tilde{p}_t^e = \lambda(q_t - q_{t-1}) + \sigma \varepsilon_{t-1}, \quad (4.19)$$

$$\Delta\tilde{u}_t = (s - \lambda)(q_t - q_{t-1}), \quad (4.20)$$

offers itself again for illustration of these cross-correlation effects. Unlike their strong form counterpart the semi-strong efficient prices are not a martingale. We see in the following proposition that in contrast to the strong form correlations, the absolute value of semi-strong form cross-correlation at displacement 0 and 1 usually differs even in one-period models.

**Proposition 6 (Semi-strong form cross-correlation, one-period model)** *The contemporaneous cross-correlation is*

$$\text{Corr}(\Delta\tilde{p}_t^e, \Delta\tilde{u}_t) = \frac{2\phi\lambda - \sigma E(q_t \varepsilon_t)}{\sqrt{\sigma^2 - 2\sigma\lambda E(q_t \varepsilon_t) + 2\phi\lambda^2}} \frac{\text{sgn}(s - \lambda)}{\sqrt{2\phi}}.$$

*The cross-correlation at displacement 1 equals*

$$\text{Corr}(\Delta\tilde{p}_{t-1}^e, \Delta\tilde{u}_t) = \frac{-\phi\lambda}{\sqrt{\sigma^2 - 2\sigma\lambda E(q_t \varepsilon_t) + 2\phi\lambda^2}} \frac{\text{sgn}(s - \lambda)}{\sqrt{2\phi}}.$$

*All cross-correlations at higher displacements are 0.*

Bounds on the contemporaneous cross-correlation can be obtained by assuming a specific market marker loss function and then solving for the market-maker's optimal  $\lambda$ . Given the quadratic loss function  $E[(\tilde{p}_t^e - p_t^*)^2]$ , e.g., the optimal adverse selection parameter is  $\lambda_{\text{opt}} = \frac{\sigma}{\phi} E(q_t \varepsilon_t) > 0$  (Web Appendix A.8.2). At this  $\lambda_{\text{opt}}$  we have

$$\text{Corr}(\Delta\tilde{p}_t^e, \Delta\tilde{u}_t) = E(q_t \varepsilon_t) \frac{\text{sgn}(s - \lambda^{\text{opt}})}{\sqrt{2\phi}} \leq \frac{1}{\sqrt{2\phi}}, \text{ and}$$

$$\text{Corr}(\Delta\tilde{p}_{t-1}^e, \Delta\tilde{u}_t) = -\text{Corr}(\Delta\tilde{p}_t^e, \Delta\tilde{u}_t) \leq \frac{1}{\sqrt{2\phi}}.$$

Given an uninterrupted flow of trades ( $\phi = 1$ ) the absolute value of cross-correlations is bounded from above by  $\frac{1}{\sqrt{2}}$ .

Proposition 6 shows that the size of the spread matters only relative to the adverse selection parameter. The cross-correlation at displacement 1, e.g., is negative if and only if the spread exceeds the adverse selection cost. The contemporaneous cross-correlation is positive as in Diebold (2006) for  $s > \lambda > \frac{\sigma}{2\phi} E(q_t \varepsilon_t) = \frac{\lambda_{\text{opt}}}{2}$  and for  $s < \lambda < \frac{\lambda_{\text{opt}}}{2}$ . For these parameters again an inverted (compared to the low risk aversion case in Section 3) cross-correlation function obtains as in the lower right panel of Figure 3. Either parametrization reflects a plausible market situation. Small spreads could obtain in some markets from competition or regulatory constraints. Large spreads without violating the market-maker's zero-profit condition can be the result of high risk aversion. By the same reasoning as in Section 3.3.2, there exists a risk aversion level  $n_0$  such that all  $n > n_0$  generate a spread  $s > \lambda$ . Generally, because the spread must cover the order processing cost, it is likely to exceed the adverse selection response.

The lower six rows of Table 2 summarize this differential behaviour of semi-strong compared to strong form efficient prices. Note again that positive contemporaneous cross-correlation for semi-strong efficient prices obtains even in situations where the market maker does not observe a signal.

In summary, positive contemporaneous cross-correlations occur, firstly, for the widely used strong form efficient prices only under high risk aversion if a signal is observed, and, secondly, frequently for latent price processes different from Brownian motion.

## 5. THE RELATIONSHIP BETWEEN PRICE CHANGE FREQUENCY AND SAMPLING FREQUENCY

In this section we discuss the implications that the frequency of price changes in financial markets has for the choice of sampling frequency. We begin with a discussion of the effects of incompletely observed latent price changes, turn then to the effect of sampling frequency, and finally examine the implications of trade frequency for econometric theory.

### 5.1. Frequency of price disclosure and return-noise correlations

For clarity of exposition in most of this article we discuss models, where  $p_{t-1}^*$  becomes public information just before it changes. In general, however, its exact value might never become public. In this case all past  $p_{t-\tau}^*$ ,  $\tau > 0$  contain unrevealed information about  $p_t^*$ .

More specifically, suppose that exact values of the  $\kappa$  most recent latent prices are not fully revealed and therefore partly private information. This changes the market-maker's problem in two ways: First, informed trades now convey the signal  $\{\text{sgn}(p_t^* - p_t)\}$ , distinct from the signal  $\{\text{sgn}(\varepsilon_t)\}$ . Second, the larger  $\kappa$ , the more spread out is *ceteris paribus* the distribution of the market-maker's belief about  $p_t^*$ .

Each signal mixes information on the  $\kappa$  most recent latent price changes,  $\Delta p_{t-iT}^*$ ,  $i \in [0, \kappa]$ , which dampens all cross-correlations towards 0 compared to the models discussed earlier. A potentially wider spread dampens the cross-correlation further.

### 5.2. Sampling frequency and return-noise correlations

We have so far assumed that  $p_t$ ,  $\tilde{p}_t^e$ , and  $p_t^e$  are all updated at the same frequency and chose this as our sampling frequency. Sampling at faster or slower rates will affect the shape of cross-correlation functions.

Consider first the effects of sampling “too fast”, in particular more frequently than trades occur. Suppose we sample  $m$  times during an interval of no changes in market prices, and for that matter, latent prices. The cross-correlation function becomes a spread-out version of the cross-correlation functions derived in the previous sections: after each damped non-zero cross-correlation follow  $m-1$  zero cross-correlations. Zeros in the middle of a cross-correlation function thus indicate overly fast sampling.

A variant of sampling “too fast” is sampling faster than information evolves. That is, sampling at trading frequency, *i.e.* the frequency of  $p_t$ , although the market maker updates  $p_t^e$  only infrequently, *e.g.* only every  $m$ -th trade. Any change of  $p_{im}^e$  ( $i \in \mathbb{N}$ ) now reflects the information about  $\Delta p_0^*$  conveyed by trading activity between  $(i-1)m$  and  $im$ .  $\Delta p_{im}$  is thus more correlated with  $\Delta p_0^*$  than under period-by-period updating. But because the quote is fixed during  $(i-1)m+1$  and  $im$ , the trades in the interim period jointly provide less information than under period-by-period updating. Because further the variance of noise increases due to the delayed accumulated market-maker response, the cross-correlation function oscillates between damped values.

Now consider the effects of sampling “too slowly”, *i.e.* slower than  $p_t^*$  and  $p_t^e$  evolve. Suppose, *e.g.*, that we sample in the one-period model of Section 4.2 only every  $m$ -th tick, where  $\hat{i}$  indexes the  $m$ -tick blocks. Then (4.19) becomes

$$\Delta \tilde{p}_{\hat{i}}^e = \sum_{i=(\hat{i}-1)m+1}^{\hat{i}m} \Delta \tilde{p}_i^e = \lambda(q_{\hat{i}m} - q_{(\hat{i}-1)m}) + \sigma \sum_{i=(\hat{i}-1)m}^{\hat{i}m-1} \varepsilon_i,$$

and the variance increases to  $\text{Var}(\Delta \tilde{p}_{\hat{i}}^e) = m\sigma^2 - 2\sigma\lambda E(q_t \varepsilon_t) + 2\phi\lambda^2$ . Assuming that the statistical properties of the interim periods are the same as the properties of the sampled periods, the expressions for noise (4.20), its variance  $\text{Var}(\Delta u_{\hat{i}})$ , and the covariance  $\text{Cov}(\Delta \tilde{p}_{\hat{i}}^e, \Delta u_{\hat{i}})$  remain unchanged. But increasing the sampling interval averages the initial transaction price reaction with later price changes, thereby again dampening the entire cross-correlation pattern towards 0:

$$|\text{Corr}(\Delta \tilde{p}_{\hat{i}}^e, \Delta u_{\hat{i}})| = \left| \frac{2\phi\lambda - \sigma E(q_t \varepsilon_t)}{\sqrt{2\phi}\sqrt{m\sigma^2 - 2\sigma\lambda E(q_t \varepsilon_t) + 2\phi\lambda^2}} \right| < |\text{Corr}(\Delta \tilde{p}_t^e, \Delta u_t)|.$$

This averaging effect across latent price changes might explain why the negative contemporaneous cross-correlation between returns and noise diminishes as more ticks are combined into one transaction price sample (Hansen and Lunde, 2006).

Standard RV is unbiased if sampling frequency is sufficiently low so that microstructure effects are averaged out. Applying “noise-corrected” RV estimators to data at lower frequencies results in biased estimates, because at lower frequencies slow moving features of the price process are removed, not microstructure noise. Thus they should only be applied to data sampled at frequencies at which microstructure effects can conceivably exist, *e.g.* above 1/100 s.

The upshot is that sampling frequency does not change the sign pattern of cross-correlations but can severely dampen their absolute values. Sampling at a rate detached from the updating frequency of prices and information mutes complications as well as information originating from dependent noise, and effectively changes the properties of the data. Sampling frequency should therefore be chosen based on the price updating frequency of the market.

### 5.3. *Sampling frequency and asymptotic theory*

The previous section has shown that the microstructure of a market implies a natural sampling frequency. In practice, sampling frequency is also central for econometric theory. Infill asymptotic theory, *e.g.*, requires the number of sampling intervals during a fixed time span to go to infinity. Sampling at an infinite frequency is impossible in real financial markets, but as trading keeps becoming faster we can view it as the trading frequency limit in the (infinite) future. Can econometric theory gain anything from examining the developments in financial markets?

Consider the Zhou (1996)-estimator as an example. Its consistency hinges on the ratio of the lag length measured by the number of sample periods to sampling frequency going to 0 as sampling becomes infinitely frequent. That is, under infill asymptotics, the time span that the lag window spans must asymptotically shrink to 0. It is commonly argued that this assumption is “inappropriate” for financial markets (*e.g.* Hansen and Lunde, 2006, p. 139). Effectively, the question comes down to whether MSN decays according to a tick-time or a calendar-time schedule. Linking econometrics to market structure, we argue in the following that tick-time dependence is reasonable in many cases.

When deriving the limiting behaviour of IV estimators, econometric theory commonly assumes that the properties of transaction prices are invariant to the sampling frequency. This might be correct in many instances, but just as often it is not. In the case of financial markets, the maximum feasible sampling frequency is dictated by the trading frequency. As the trading frequency in a given market changes, other features of that market change as well. Therefore asymptotic theory must account for the possibility that price behaviour changes as feasible sampling frequency increases.

To verify the relevance of this possibility, let us revisit the economics of financial markets. The analogue of shrinking the interval length in infill asymptotics is a higher trading frequency in financial markets, which implies a higher feasible sampling frequency. In the following three examples, we examine how a higher feasible sampling frequency affects noise persistence. We consider a slow and a fast market: The slow market is rather illiquid, so that a trade is observed only once during a 5-min interval. The fast market is more liquid, and trades are observed once every minute. The latent price process is the same in both markets. In fact, both slow and fast market might be the very same market at different points in time. The latent price moves more between two trades in the slow market, which means that there the IV over the shortest possible sampling interval is higher.

Consider first a bid–ask bounce. Bid–ask bounces are purely mechanic, and directly linked to observed trades. In the slow market, the possible rebounce occurs 5 min after the original trade,

whereas in the fast market it occurs after only 1 min. Thus the MSN is autocorrelated for 5 min in the slow market, but only for 1 min in the fast market.

Next, consider asymmetric information. If learning of market participants is automated and limited to information extracted from trade signals, then the amount of learning grows in the number of trade signals observed, not in the time that has passed. For a specific example, suppose the market maker needs 10 trades to include half of the latent price change into his price quote. This will take 50 min in the slow, but only 10 min in the fast market. MSN persistence measured in calendar time is thus much shorter in the faster market.

Our third example shows that this applies only to tick-dependent MSN, *i.e.* to situations where private information is revealed by trades only and where the speed of information processing is not a binding constraint. Some properties of MSN, however, might be invariant to sampling frequency. For example, the time that strategic informed traders allocate to fully reveal their information might be exogenous to the trading frequency. Instead, its optimal value might be a function of the speed of information diffusion outside the market, determined by *e.g.* reporting delays, which are fixed in calendar time. Thus the autocorrelation of MSN generated by strategic informed traders in calendar time is the same in the slow and the fast market; it does not shrink as sampling frequency increases.

Overall, the autocorrelation of MSN due to a bid–ask bounce and asymmetric information without strategic traders shrinks in calendar time as the feasible sampling frequency increases. The autocorrelation of MSN due to some types of strategic traders does not.

This has an important implication for the asymptotic theory of IV estimators of the Zhou (1996)- and Hansen and Lunde (2006)-type. When private information is revealed by trades only, the necessary lag length is fixed in terms of ticks, not calendar time. Therefore, the ratio of lag length to sampling frequency approaches 0 when sampling infinitely fast. In these cases the estimators are consistent. They must be modified to ensure consistency when relevant information transmission occurs outside of the financial market, *e.g.* by subsampling (Barndorff-Nielsen *et al.*, 2011b) or kernel-based downweighting of higher order autocovariances (Barndorff-Nielsen *et al.*, 2008).

## 6. PRACTICAL IMPLICATIONS AND EMPIRICAL APPLICATION

We have already drawn some econometric implications insofar as we have shown that market microstructure models predict rich cross-correlation patterns between latent prices and MSN, which have yet to be investigated empirically. Here we go farther, sketching some specific aspects of such empirics, including strategies for using microstructural information to obtain improved “structural” volatility estimators, and comparative aspects of structural and non-structural volatility estimators. We apply our methodology to the stock and the oil futures market.

### 6.1. *Structural volatility estimation via microstructural restrictions*

In the introduction we highlighted the key issue of estimation of integrated volatility (IV) using high-frequency data, the potential problems of the first-generation estimator (simple realized volatility—RV) in the presence of MSN, and subsequent attempts to “correct” for MSN.

In an important development, Barndorff-Nielsen *et al.* (2008) suggest making RV robust to serial correlation via realized kernel estimation methods, which are asymptotically justified under very general conditions. That asymptotic generality is, however, not necessarily helpful in

finite samples. Indeed the frequently unsatisfactory finite-sample performance of non-parametric HAC estimators leads Bandi and Russell (2011) to suggest sophisticated alternative statistical approaches.

Here we explore aspects of a different approach that *specializes* the estimator in accordance with the implications of market microstructure theory. We follow the idea of Aït-Sahalia *et al.* (2005) of modelling MSN explicitly in a fully parametric framework, which makes sampling as often as possible optimal. Here we do not claim optimality; instead we show the practical relevance of tailoring the estimator to the market at hand.

Consider strong form noise given by (2.3), so that  $\Delta p_t = \Delta p_t^* + \Delta u_t$ . Then we have, absent insider information, using the notation  $\gamma_i \equiv E(\Delta p_t \Delta u_{t-i})$  and  $RV \equiv \gamma_0$ , that the variance of strong form efficient returns (2.7) is

$$\sigma^2 = RV + 2 \sum_{i=1}^k \gamma_i - 2E(u_t \Delta u_{t-k}) - 2E(\Delta p_t^* u_{t+k}). \quad (6.21)$$

*Proof* See Web Appendix B.1. ||

If MSN is asymptotically uncorrelated, *i.e.* if  $\lim_{k \rightarrow \infty} E(u_t \Delta u_{t-k}) = 0$  and  $\lim_{k \rightarrow \infty} E(\Delta p_t^* u_{t+k}) = 0$ , then equation (6.21) simplifies to

$$\sigma^2 = RV + 2 \sum_{i=1}^{\infty} \gamma_i. \quad (6.22)$$

This is equivalent to the constant realized kernel estimator discussed in Hansen and Lunde (2006). Without insight in the market microstructure all higher-order autocovariances are potentially important. Empirically most will be noisy estimates of 0 (Barndorff-Nielsen *et al.*, 2008). Without insights in what patterns in transaction prices are caused by MSN, a noise correction like (6.22) will remove all. But actual transaction prices consist not only of a martingale strong form efficient price plus MSN, but also of other disturbances of unknown form. These other disturbances might not be part of any microstructure model. In fact, their existence might not even be known. Lacking better knowledge by any market participant, these must be considered risk, and therefore be part of the volatility estimate of the latent price. A noise correction as equation (6.22) “corrects” price features that are not MSN, but an essential part of the volatility of the latent price process.

The key point we stress in this article is that it is indispensable to sort out the market microstructure before choosing a noise correction. This applies no matter whether MSN is dependent on the latent price or not.

In the following we consider ten potential sources of MSN, five of which are independent, and five are dependent on the latent price. We start with a discussion of two examples of parsimonious noise-robust estimators for realized volatility, both of which are special cases of (6.22), before providing an overview of estimators in Table 3.

Consider first a “bid–ask bounce estimator”, based on a one-period model without extra information and constant spread. From (2.3), (2.5), and (2.6) we obtain  $\Delta u_t = \sigma(\varepsilon_{t-1} - \varepsilon_t) + s(q_t - q_{t-1})$ , and this implies a variance of strong form efficient returns of

$$E[(\Delta p_t^*)^2] = E[(\Delta p_t - \Delta u_t)^2] = E(\Delta p_t^2) + 2s[\sigma E(q_t \varepsilon_t) - \phi s].$$

TABLE 3  
Noise-robust estimators for realized volatility

$\hat{IV}$	Strong-form noise	
	Independent	Dependent
$\gamma_0 + 2\gamma_1$	Measurement error/discrete data $u_t = v_t$ Bid–ask bounce $u_t = \alpha q_t$ , $q_t \in \{-1, +1\}$ iid	Dependent measurement error $u_t = \alpha \varepsilon_t + v_t$ Bid–ask bounce from informed traders $u_t = \alpha q_t, q_t = \begin{cases} -1 & \text{if } \varepsilon_t < 0 \\ +1 & \text{if } \varepsilon_t > 0 \end{cases}$
$\gamma_0 + 2\frac{\gamma_1^2}{\gamma_1 - \gamma_2}$	Autoregressive noise $u_t = \alpha u_{t-1} + v_t$	Non-strategic incompletely informed traders $u_t = \alpha u_{t-1} + \beta(\varepsilon_t + v_t)$
$\gamma_0 + 2\gamma_1 + 2\frac{\gamma_2^2}{\gamma_2 - \gamma_3}$	Autoregressive noise with measurement error $u_t = \sum_{i=0}^{\infty} \alpha^i v_{t-i} + w_t$	Autoregressive noise with one-period private information $u_t = \sum_{i=0}^{\infty} \alpha^i v_{t-i} + \beta \varepsilon_t + w_t$ Non-strategic informed traders $u_t = \alpha \sum_{i=0}^{\infty} \beta^i \varepsilon_{t-i} + v_t$
$\gamma_0 + S(S+1)\gamma_1$	Linear noise decay over $S$ periods $u_t = \alpha \sum_{i=t}^{t-S} \frac{i-t+S}{S} v_i$	
$\gamma_0 + S(3-S)\gamma_1 + S(S-1)\gamma_2$	Strategic informed traders with $S$ -period private information $u_t = \alpha \sum_{i=t}^{t-S} \frac{i-t+S}{S} (\varepsilon_i + v_i)$	
$\gamma_0 - \gamma_0^{*2}$	Market-maker inventory from noise trading $u_t = \alpha \sum_{i=0}^{\infty} q_{t-i},$ $q_t \in \{-1, +1\}$ iid	Market-maker inventory from informed trading $u_t = \alpha \sum_{i=0}^{\infty} q_{t-i}, \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, 1)$ $q_t = \begin{cases} -1 & \text{if } \varepsilon_t < 0 \\ +1 & \text{if } \varepsilon_t > 0 \end{cases}$
$\frac{\pi}{\pi-2} (\gamma_0 - \gamma_0^{*2})$		

*Notes:* The estimators are based on the observable moments  $\gamma_i \equiv E(\Delta p_t \Delta p_{t-i})$  and  $\gamma_i^* \equiv E(\Delta p_t q_{t-i})$ . They are based on the assumption that the latent price changes every period ( $T=1$ ), and remains unobserved for one or more periods, depending on the noise specification. The probability of no latent price change has measure 0. The two white noise processes are  $v_t \stackrel{\text{iid}}{\sim} (0, \eta_v^2)$  and  $w_t \stackrel{\text{iid}}{\sim} (0, \eta_w^2)$ , where  $E(v_s w_t) = 0 \forall s, t$ .

Simple calculations reveal that the last term equals twice the first-order autocorrelation of market returns, so that, even if  $E(q_t \varepsilon_t) \neq 0$ , an unbiased estimator for  $IV = \sigma^2$  is<sup>7</sup>

$$\hat{IV} = RV + 2\gamma_1. \quad (6.23)$$

It is interesting to note the resemblance to estimators of Roll (1984), based on standard asymptotic theory, and Zhou (1996), based on infill asymptotic theory.

7. Hasbrouck (1993) and recently Hansen *et al.* (2008) show how to embed (6.23) into general moving average (MA)-based estimators. Such general MA-estimators are warranted if the researcher has only limited information about the microstructure of the market or has interest different from IV estimation, such as forecasting the latent price process. If, however, the microstructure is known and interest centers on estimating IV, as we assume here, then our estimators may be more appealing.

As another example, consider an estimator for a market with non-strategic incompletely informed traders. Absent any exogenous noise, the transaction price follows an  $\text{MA}(\infty)$  process in the innovations of the latent price:

$$\Delta p_t = (\beta + \sigma) \varepsilon_t + \beta(\alpha - 1) \sum_{i=0}^{\infty} \alpha^i \varepsilon_{t-i-1} \quad (6.24)$$

This parsimonious form of  $\Delta p_t$  accommodates very persistent cross-correlations, similar to the idea behind the examples in Oomen (2006). If our knowledge of the market is this comprehensive, we can obtain an unbiased estimate for IV from (6.24) in a GMM framework using three moments.<sup>8</sup> More specifically, a standard exponential learning model (*e.g.* Easley and O'Hara, 1992) imposes  $\alpha = e^{-r}$  and  $\beta = -\sigma$ , so that

$$\Delta p_t = 0 \cdot \sigma \varepsilon_t + \sigma (e^r - 1) \sum_{i=1}^{\infty} e^{-ri} \varepsilon_{t-i} = \sum_{i=1}^{\infty} [-e^{-ri} + e^{-r(i-1)}] \sigma \varepsilon_{t-i}.$$

The resulting estimate of IV is a scaled version of standard RV

$$\widehat{\text{IV}} = \frac{e^{\widehat{r}} + 1}{e^{\widehat{r}} - 1} \cdot \text{RV} = \frac{\text{RV} + \gamma_1}{\text{RV} - \gamma_1} \cdot \text{RV}, \quad (6.25)$$

where the scaling factor requires a consistent estimate of only one additional parameter, the market-maker's learning rate,  $r$ . It is interesting to note the resemblance to the estimator of Hansen *et al.* (2008), which is also a scaled variant of RV. In contrast to our approach, they do not exploit (*i.e.*, condition on) a specific market microstructure, but attempt to achieve robustness to a wide range of possible microstructures.

Estimator (6.25) offers a structural interpretation to estimates of noise and IV. The learning model predicts that the MSN at all lags decreases with the learning rate. Slow learning implies a very persistent cross-correlation between noise and latent returns, and hence persistent autocorrelation of noise, so that fluctuations in MSN tend to dominate the IV.

Figure 5 provides some perspective. It is based on the noise-to-IV ratios reported by Hansen and Lunde (2006), which are (unfortunately) derived under the assumption of independent noise. The ratio of noise-to-IV shrinks with the number of price-changing quotes per day. If the number of times that the market maker changes his price quote during a trading day is indicative of his speed of learning, then MSN indeed decreases as the learning rate of the market maker increases. This supports the multiperiod learning model.

Furthermore, the recent decline in noise-induced bias of RV (Hansen and Lunde's (2006) fact III) suggests that the learning rate  $r$  has increased. Adding to this Meddahi's (2002) finding that the standard deviation of the bias is large relative to the IV suggests that the learning rate itself may have fluctuated considerably around its increasing trend.

Our example uses a MA process with only two free coefficients, but the large sample sizes typical with high-frequency data can accommodate much richer specifications. Empirical work in market microstructure tends to favour extreme parameterizations, ranging from the very

8. The proof, which we sketch here, is straightforward. Recast the price process (2.1) and (2.2) in continuous time, so that  $\Delta p_t = \Delta p_t^* + \Delta u_t / \sqrt{m}$ , with  $m$  denoting the number of subintervals,  $tm$  equal to one unit of calendar time, and the scale of  $t$  suitably redefined. Then, following the considerations of Section 5.3, under standard assumptions  $r$  is invariant to  $m$  and local infill asymptotic theory can be applied.

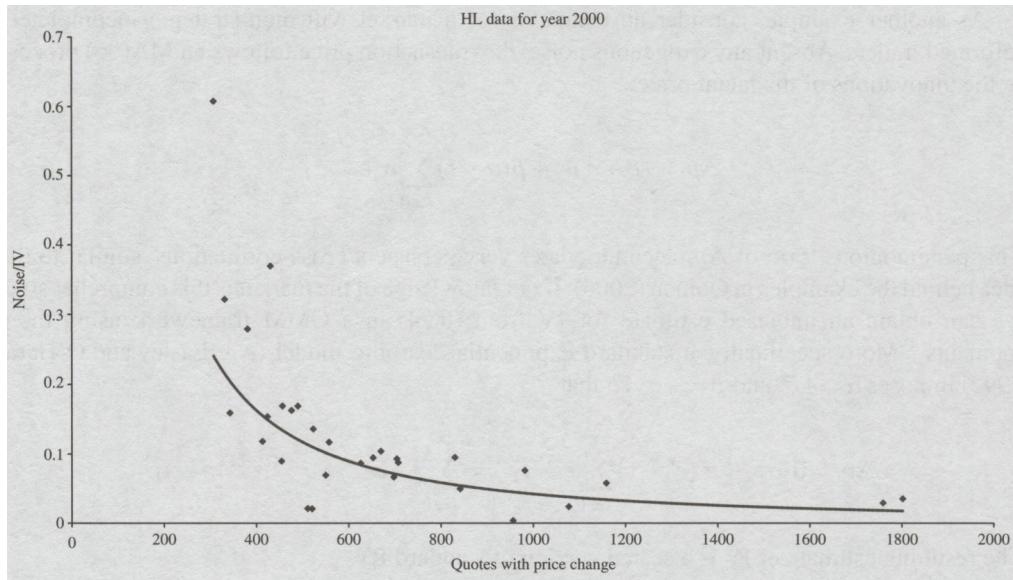


FIGURE 5  
Ratio of noise to IV as a function of quotes per day

*Note:* The vertical axis measures the noise-to-signal ratio as 100 times noise divided by IV under the assumption of independent noise. The horizontal axis gives the number of quotes per day with a price change. Data are for 30 NYSE and NASDAQ equities in 2000, obtained from Hansen and Lunde (2006) Tables 1 and 3. The solid line is a fitted trend.

parsimonious as in the regressions of Glosten and Harris (1988), to the profligate as in the vector autoregressions of Hasbrouck (1996). For RV noise correction the most useful parameterizations may be intermediate, imposing a general correlation pattern but avoiding highly situation-specific assumptions.

Dynamic market microstructure models imply much richer noise structure than the two polar cases of immediate and slow decay that we just discussed. These restrictions can be exploited to construct tailored volatility estimators. In Table 3 we do so by suggesting parsimonious estimators for a variety of market microstructures. Whereas these estimators inevitably also remove price features that are empirically indistinguishable from modelled MSN, their parsimony ensures that this miscorrection is kept to a minimum.

The table is structured as follows: The left column gives the estimator, the middle column an example of independent MSN, to which this estimator applies, and the right column an example of dependent MSN. Interestingly, for many market microstructures that generate dependent noise there is a corresponding market structure with independent noise to which the same estimator fits. Web Appendix B shows that all these estimators are unbiased. They are consistent under the conditions discussed in Section 5.3 or under subsampling (Barndorff-Nielsen *et al.*, 2011b).

We only discuss the dependent noise cases here, because these are—as we have shown in this article—the ones of relevance in actual financial markets. The first row of Table 3 shows that the Zhou (1996)-estimator is the most parsimonious way to deal with a market in which the only MSN stems from the bid–ask bounce, even if trades are driven by private information. The geometric decay of MSN over time under learning is covered by rows two and three, for various exogenous noise processes. The decay becomes linear if traders act strategically, reflected in row four. These three learning estimators specialize to the Zhou (1996)-estimator with  $\gamma_2 = 0$  for non-strategic, and with  $\gamma_1 = \gamma_2$  or  $S = 1$  for strategic informed traders. Likewise, the noise process

we discussed earlier in equation (6.25),  $u_t = \alpha u_{t-1} - \sigma \varepsilon_t$ , is a special case of the non-strategic incomplete informed trader case, with  $\beta = -\sigma$  and  $v_t = 0$ . Finally, the estimator for strategic informed traders collapses to the estimator for linear independent noise decay if  $\gamma_2 = 2\gamma_1$ .

The contribution of the delayed price responses to the learning RV estimators in rows two and three can be expressed by any pair of autocovariances,  $\gamma_i, \gamma_{i+1}, i \geq 2$ . Whereas in the table we show the most parsimonious expression, replacing the last term by an average stabilizes the estimates. For example, in the non-strategic incompletely informed trader case, we can use  $\gamma_0 + 2\gamma_1 \frac{1}{S} \sum_{i=1}^S \frac{\gamma_i}{\gamma_i - \gamma_{i+1}}$ , for any  $S \geq 1$ .

With strategic informed traders choosing the correct length of the private information period is critical for unbiased results, as noted already by Kelly and Steigerwald (2004). In our setup in

$$\text{row four } S \text{ can be estimated by } \hat{S} = \sqrt{\left(\frac{3\gamma_1 - \gamma_2}{2(\gamma_2 - \gamma_1)}\right)^2 + \frac{2}{\gamma_2 - \gamma_1} \sum_{i=1}^{\infty} \gamma_i - \frac{3\gamma_1 - \gamma_2}{2(\gamma_2 - \gamma_1)}}.$$

The MSN in the upper four rows of Table 3 is asymptotically uncorrelated, so IV can be expressed by equation (6.22). This equation does not hold in market-maker inventory models, as the ones in the bottom row. There, MSN follows a unit root process with  $\Delta u_t = \alpha q_t$  so that  $E(u_t \Delta u_{t-i}) = \alpha^2 \forall i$ . In this case autocovariances alone are not sufficient, and  $\widehat{IV}$  must be based on the general equation (6.21).

A common argument for using estimators that, contrary to equation (6.22), downweight autocovariances at non-zero displacements is that it rules out the possibility of a negative volatility estimate. Starting the analysis with such an estimator, however, strips the researcher of the chance to falsify his assumptions on the market microstructure. After all, a negative variance estimate first and foremost indicates that the estimator is misspecified for the microstructure of the market under analysis, and that it should be refined. We therefore suggest starting with microstructure-inspired estimators as the ones in Table 3, and resort to microstructure-free estimators if the market microstructure appears to obey to none of the common models.

Barndorff-Nielsen *et al.* (2009) notice that their IV estimator's ability of detecting properties of volatility depends crucially on the bandwidth: "The 'strength' of this 'microscope' is controlled by the bandwidth parameter, and the realized kernel gradually loses its ability to detect volatility at the local level as ... [the bandwidth] is increased." (Barndorff-Nielsen *et al.*, 2009, p. C27) In effect, there is a tradeoff between the loss of local volatility information and the MSN bias. Utilizing prior knowledge about the market microstructure, Table 3 allows an informed bandwidth choice instead of having to rely exclusively on statistical arguments.

## 6.2. On structural vs. non-structural volatility estimators

Here we emphasize that the more the econometrician knows about the price process of relevance, the more the noise correction can be tailored to it by exploiting microstructure theory. This is important, because, as discussed in Section 2, the price process of interest may differ across users of volatility estimates. Many users are likely to be interested in price processes different from (2.1), which has implications for appropriate volatility estimation. The variance of strong form efficient returns,  $E(\Delta p_t^*) = \frac{\sigma^2}{T}$ , the price under full information, differs both conceptually and numerically from the variance of semi-strong efficient returns,

$$E[(\Delta \tilde{p}_t^e)^2] = \frac{1}{T} \left\{ \sigma^2 + \sum_{i=0}^{T-1} \phi_i \lambda_i^2 + E \left[ \left( \sum_{i=-1}^{-T} \lambda_i q_i \right)^2 \right] - 2\sigma \sum_{i=-1}^{-T} \lambda_i E(q_i \varepsilon_{-T}) \right\}, \quad (6.26)$$

which is the volatility that affects the balance sheet of the market maker. It might therefore be more applicable to studies of market-maker behaviour than  $E[(\Delta p_t^*)^2]$ . To take a simple example, consider again one-period private information,  $T = 1$ , in which case strong form volatility is  $\sigma^2$  and semi-strong volatility (6.26) simplifies to

$$E[(\Delta \tilde{p}_t^e)^2] = \sigma^2 + 2\phi\lambda^2 - 2\sigma\lambda E(q_t \varepsilon_t) \neq \sigma^2. \quad (6.27)$$

The RV estimator of Zhou (1996) is

$$RV_{AC(1)} = E(\Delta p_t^2) + E(\Delta p_{t-1} \Delta p_t) + E(\Delta p_t \Delta p_{t+1}),$$

which is equivalent to equation (6.23). For  $T = 1$  it is

$$E(RV_{AC(1)}) = E\{[s(q_t - q_{t-1}) + \sigma \varepsilon_{t-1}] \times [\sigma(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2}) + s(q_{t+1} - q_{t-2})]\} = \sigma^2.$$

Hence although  $RV_{AC(1)}$  is unbiased for  $\sigma^2$ , it is in general biased with ambiguous direction for  $Var(\Delta \tilde{p}_t^e)$  in (6.27). The same applies to a noise-robust estimator with a large, potentially infinite, lag window, which removes *any* microstructure and other correlation effect. For these estimators to work, the latent return process of *interest* must follow a martingale difference sequence. Semi-strong efficient prices do not; they are serially correlated and inevitably  $RV_{AC(1)}$  is biased relative to  $Var(\Delta \tilde{p}_t^e)$ .

What could an estimator of semi-strong form volatility look like? Consider, *e.g.*, a market where the strong form efficient price becomes public after two periods. From (2.4)–(2.9), we obtain  $\Delta p_t = \Delta \tilde{p}_t^e + \Delta u_t$  with noise given by (4.20). It follows that

$$E[(\Delta \tilde{p}_t^e)^2] = E(\Delta p_t^2) + 2s\phi(\lambda - s).$$

Because  $\tilde{p}_t^e$  is generated by a more complex process than  $p_t^*$ , we need additional market data. Using the autocorrelations of prices and additional market information such as an estimate of the spread  $s$  and of the trade frequency  $\phi$ , an unbiased estimator for IV of the semi-strong efficient price is

$$\widehat{IV} = RV + 2\gamma_1 - 2 \frac{\gamma_1 \gamma_2}{\gamma_2 + s^2 \phi}. \quad (6.28)$$

The obvious difference to the estimators in Table 3 emphasizes the importance of carefully defining the latent price series of interest.<sup>9</sup> This is where market microstructure theory can contribute new insights to IV estimation. By providing distinctive but flexible relationships between MSN and latent returns, and using additional market information, the agnostic statistical noise estimate can be decomposed into its various MSN and fundamental components.

9. To avoid confusion we adhere in this article to the convention that the strong form efficient price follows a martingale. Therefore we introduced  $\tilde{p}_t^e$  as another latent price series of interest. But there is no guarantee that a price with martingale properties exists in a given market. For example, the latent price could itself be the result of learning about random-walk fundamentals, in which case  $p_t^*$  has the properties of the semi-strong form efficient price  $\tilde{p}_t^e$ . Specifically, let fundamentals follow  $\chi_t = \chi_{t-1} + \varepsilon_t$  with  $\varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ . Then the latent price process, known only to the best informed market participants, is

$$\Delta p_t^* = \sigma \sum_{i=1}^T [-e^{-r_1 i} + e^{-r_1(i-1)}] \varepsilon_{t-i}.$$

If market makers are well informed ( $p_t^e = p_t^*$ ) and the bid–ask bounce follows equation (2.6), then mechanically calculating  $RV_{AC(T)}$  gives the variance of the fundamental, not the variance of the strong form efficient price. Obviously, a purely statistical noise correction cannot distinguish between cross-correlation caused by fundamentals and cross-correlation caused by MSN.

TABLE 4  
*Comparison of realized volatility estimators (CTS at 1/s)*

RV	Price	Mid	Bid	Ask
Standard	2.493	1.605	2.733	2.685
Statistical—ACNW(30)	2.146	2.141	2.255	2.257
Bid–ask—AC(1)	2.377	1.547	2.524	2.475
Learning—restricted	2.379	1.548	2.532	2.483
Learning—non-str. noisy	2.268	1.437	2.603	2.395
Learning—non-strategic	2.171	1.435	2.429	2.237
Learning—strategic	2.361	2.363	2.368	2.409

TABLE 5  
*Comparison of realized volatility estimators (TTS at 1/tick)*

RV	Price	Mid	Bid	Ask
Standard	2.494	1.605	2.733	2.685
Statistical—ACNW(30)	2.386	2.511	2.506	2.534
Bid–ask—AC(1)	1.813	1.603	2.313	2.238
Learning—restricted	1.895	1.603	2.343	2.272
Learning—non-str. noisy	1.938	1.602	2.424	2.194
Learning—non-strategic	1.816	1.677	2.208	2.096
Learning—strategic	2.164	2.290	2.284	2.304

### 6.3. Two empirical applications

In this section we provide two illustrative applications of our microstructure-based estimators. We first compare their properties with the standard RV and a statistical IV estimator in a well-known stock market dataset. Then we turn to a current policy debate centering on oil futures volatility.

**6.3.1. Alcoa stock.** As a first application, we compare microstructure-based estimates with statistical estimates of IV of Alcoa Inc. (AA) stock.<sup>10</sup> We use prices on the New York Stock Exchange in the year 2004 cleaned according to the procedure of Barndorff-Nielsen *et al.* (2009). All overnight returns and days with less than 5 h of trading were removed from this dataset, which means that the IV-estimates apply only to the price process *within* trading days. They do not capture the overall riskiness of the stock, because price changes between trading days are excluded.

The estimators in Tables 4–7 are for daily IV, *i.e.*  $E[(\Delta p_t^*)^2]$ , averaged across the year. *RV Standard* is simple realized volatility,  $E(\Delta p_t^2)$ , *RV Bid–ask* is the bid–ask estimator (6.23), *RV Learning—Restricted* the learning estimator (6.25), and *RV Learning—Non-str. Noisy* stands for the estimator for non-strategic, incompletely informed traders. All microstructure-based estimators are defined by Table 3. *RV Statistical* is the consistent flat-top kernel estimator  $RV_{ACNW}(30) = \gamma_0 + 2 \sum_{i=1}^{30} \gamma_i + 2 \sum_{i=1}^{30} \frac{30-i}{30} \gamma_{30+i}$  of Hansen and Lunde (2006). It serves as benchmark, as a statistical estimator that removes all deviations of the transaction price from

10. Gatheral and Oomen (2010) compare 19 IV estimators on simulated data and conclude that a realized kernel and a maximum-likelihood-based estimator perform best in practice. However, they ignore microstructure noise for the most part. Patton (2011) compares four statistical IV estimators of IBM stock prices under time-varying volatility. Absent jumps, they perform better than standard RV sampled 1/5 min.

TABLE 6  
*Comparison of realized volatility estimators (CTS at 1/10 s)*

RV	Price	Mid	Bid	Ask
Standard	2.149	1.585	2.328	2.244
Statistical—ACNW(30)	2.155	2.158	2.160	2.169
Bid–ask—AC(1)	1.970	1.764	2.160	2.102
Learning—restricted	1.977	1.774	2.166	2.107
Learning—non-str. noisy	2.094	1.848	2.190	2.141
Learning—non-strategic	1.983	1.994	2.168	2.132
Learning—strategic	2.211	2.204	2.206	2.214

TABLE 7  
*Comparison of realized volatility estimators (TTS at 1/10 ticks)*

RV	Price	Mid	Bid	Ask
Standard	2.117	1.835	2.216	2.162
Statistical—ACNW(30)	2.364	2.232	2.227	2.242
Bid–ask—AC(1)	2.493	2.273	2.307	2.323
Learning—restricted	2.530	2.333	2.309	2.330
Learning—non-str. noisy	2.494	2.206	2.243	2.272
Learning—non-strategic	2.525	2.429	2.400	2.424
Learning—strategic	2.482	2.354	2.361	2.357

a martingale, which might be different from the IV of the—in our terminology—true latent price process, *i.e.* RV corrected for market-microstructure-induced noise only.

All estimators presented here except the *RV Standard* allow for correlation between noise and latent price. We do not implement the inventory estimators here, because the dataset does not contain signed trades. In the following, we refer to the difference between *RV Standard* and *RV Statistical* as “noise”, in contrast to deviations due to market microstructure effects, which we call MSN.

Table 4 reveals that under 1/s calendar time sampling (CTS) sampling both the restricted learning and bid–ask estimators explain one-third of noise in transaction prices (in the second column). Learning appears to be very fast ( $\hat{r} > 3$ ), which implies that  $\gamma_1$  is small compared to RV. As a result the learning and bid–ask volatility estimates are very similar. More flexible learning estimators capture more of the noise. *RV Learning—Non-strategic*, in particular, captures more than 90% of what *RV Statistical* removes as noise. This means that for Alcoa under CTS indeed most of the noise correction embedded in *RV Statistical* is most likely justified—it is MSN stemming from nonstrategic informed traders. Similarly, *RV Learning—Non-str. Noisy* and *RV Learning—Strategic* capture between two-thirds and all of noise.

Under tick-time sampling (TTS) all microstructure-based estimators estimate IV substantially lower than *RV Standard* and *RV Statistical*. But if microstructure-based estimators remove the most common MSN types at this sampling frequency, then what does *RV Statistical* add back in? What positive cross-correlation between the latent price and noise different from learning can justify the higher estimate? At this point we have to leave this for further research, but also as a warning against a noise correction without a microstructure interpretation in mind.

At lower sampling frequencies the microstructure-based estimators are less tightly linked to the model setup under which we derived them. Whereas *RV Standard* and *RV Statistical* almost coincide that these frequencies, the learning estimators suggest a downward correction under CTS (Table 6) and upward correction under TTS (Table 7).

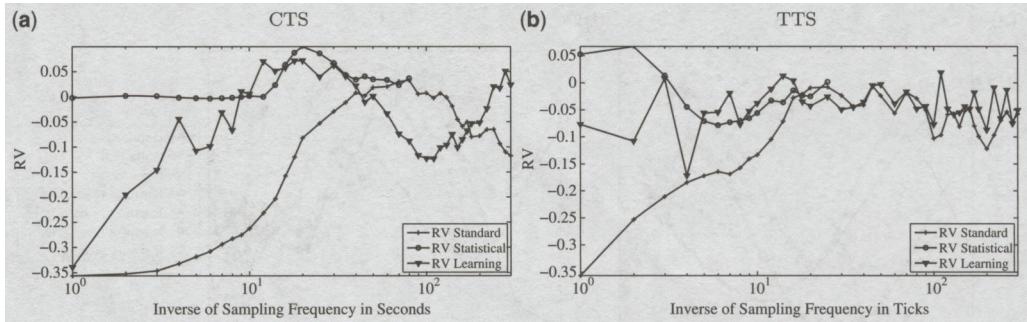


FIGURE 6  
RV estimators and market-maker learning

Examining the structural parameter estimates (not tabulated) provides additional guidance about which microstructure effects are at work at a given frequency. For example, under TTS and transaction prices, the restricted learning estimator fits the data at sampling intervals below 20 ticks (and beyond 130), whereas the strategic learning estimator at intervals up to about 130 ticks. Under CTS, restricted learning fits at frequencies of 1/30 s and slower (and is thus not reliable for the frequencies reported in the tables), and strategic learning at frequencies of 1/30 s and faster.

The IV estimates based on mid quotes are smaller than the other estimates at sampling frequencies of 1/1 s or 1/1 tick, in Tables 4 and 5, respectively. This calls for caution. In the microstructure model setup we discussed, midprices are the least noisy among the four prices. This would even be true if the spread was time-varying, as long as it was independent of future changes in midprices, *e.g.*  $E(\Delta p_t^\text{e} \Delta p_{t-1}^\text{bid}) = E(\Delta p_t^\text{e} \Delta p_{t-1}^\text{ask}) = 0$ . If this was the case, the noise correction would push estimates towards midprice-based estimates, and when lowering sampling rates all IV estimates would converge to these midprice-based values. Tables 6 and 7 reveal that the opposite is the case: At lower sampling frequencies the midprice IV estimates reach the IV estimates of the other three price series. Because none of our microstructure-based IV estimators acceptably corrects the midprice estimates, we conclude that the midprices are subject to a microstructure effect that we did not take into account in deriving the estimators. A plausible explanation is an asymmetrically moving spread, where a change in the bid price, say, is followed by an analogous change in the ask price in a later period, thus temporarily widening the spread. The temporary uncertainty that the wider spread represents is justified, because over longer horizons the latent price is indeed that volatile. The unravelling of uncertainty can be seen as an instance of market-maker learning, so there is reason to hope that a learning estimator such as *RV Learning—Non-strategic* improves the estimate. This is indeed the case. Figure 6 shows the deviation of IV estimates based on midprices from the estimates based on transaction prices, expressed by the ratio  $\frac{RV_{\text{mid}} - RV_{\text{trans}}}{RV_{\text{trans}}}$ . Under TTS, shown on the right panel, the learning estimator does well despite its misspecification. It also improves the estimate under CTS, except at very high frequencies, which is shown on the left panel. Estimators with wide lag windows, such as *RV Statistical* and *RV Learning—Strategic* with estimated learning period are robust to this kind of time-varying spread. However, as said, they remove this part of MSN jointly with non-MSN components.

The volatility signature plots (Andersen *et al.*, 2000) in Figure 7 graph average daily realized volatility as a function of the underlying sampling frequency. One might argue against the use of parsimonious, but microstructure-based, estimators on the practical ground that they do not fully stabilize as the sampling frequency approaches its limit, *i.e.* 1/tick. The volatility signature plots reveal, however, that for the given data *RV Statistical* is not stable either—it moves in a range of

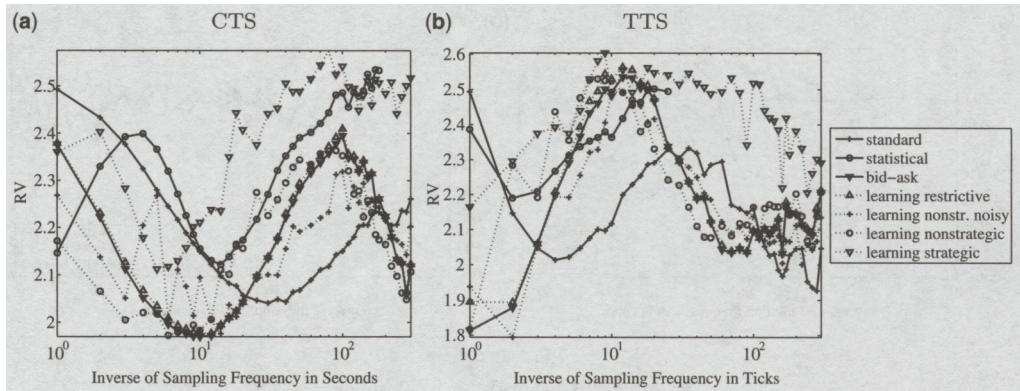


FIGURE 7  
Volatility signature plots for transaction prices

2.2–2.5 for sampling frequencies above 1/100 s or ticks. Most microstructure-based estimators are just as stable.

**6.3.2. Crude oil futures.** In this subsection we apply our estimators to Light Sweet Crude Oil futures (symbol CL) traded on the New York Mercantile Exchange (NYMEX). Our dataset consists of tick-by-tick transaction data from Tick Data, Inc., covering the period from January 2nd, 1987 until September 24th, 2010. It contains trades both within and outside of the main trading hours, which are Monday through Friday from 9:00 a.m. until 2:30 p.m. Eastern Time. The oil future is a standardized contract. One contract covers 1000 bbl with a fixed expiration date, on which oil has to be physically delivered at Cushing, OK. About 66% of trades in our sample are for a single contract, and less than 10% are for more than ten contracts.

Physical delivery is the exception, however, as most market participants roll their positions over to a new contract. We replicate this rollover by constructing a single time series of oil futures prices from the set of futures of different maturities simultaneously traded at each point in time. We switch from one contract to the contract maturing next as soon as the daily volume of the latter exceeds the current contract's volume. In the following analysis, we use TTS and exclude contract rollover and overnight returns.

Comparing the IV estimates from the estimators discussed in this article, the volatility signature plots in Figures 8 and 9 reveal that when sampling at a rate of 1/10 ticks or slower all estimators coincide. At higher sampling frequencies *RV Standard* diverges, which vividly depicts the MSN in oil futures data. The two most restrictive learning estimators, *RV Learning—Restrictive* and *RV Learning—Non-strategic Noisy*, do not stabilize either at higher frequencies, suggesting that MSN in oil futures is more complex than this. In contrast, the estimators *RV Learning—Non-strategic*, *RV Learning—Strategic*, and *RV Statistical* stabilize as sampling frequency reaches 1/1 tick. This convergence pattern in ticks did not change over the years despite a decline in the time between ticks from more than one per minute in 1989 to less than one per second in 2009.

In the remaining analysis we sample at the highest possible frequency, *i.e.* 1/1 tick, and use accordingly only the four estimators that we identified in the volatility signature plot to converge with oil futures data. Prices of oil futures follow a pronounced seasonal volatility pattern. Figure 10 shows that volatility during 1987–2010 is particularly high in January, and reaches its low around

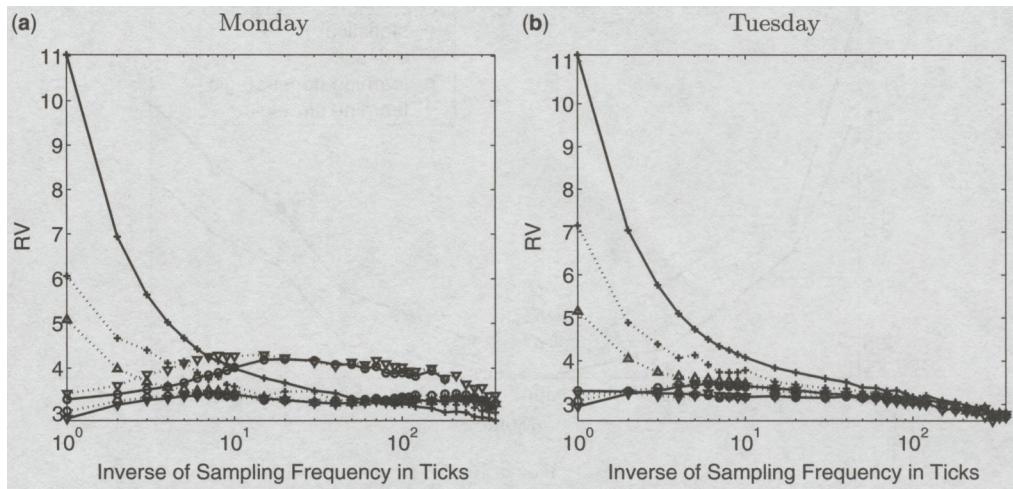


FIGURE 8  
Volatility signature plots of oil futures by day of week  
*Note:* The legend is the same as in Figure 7.

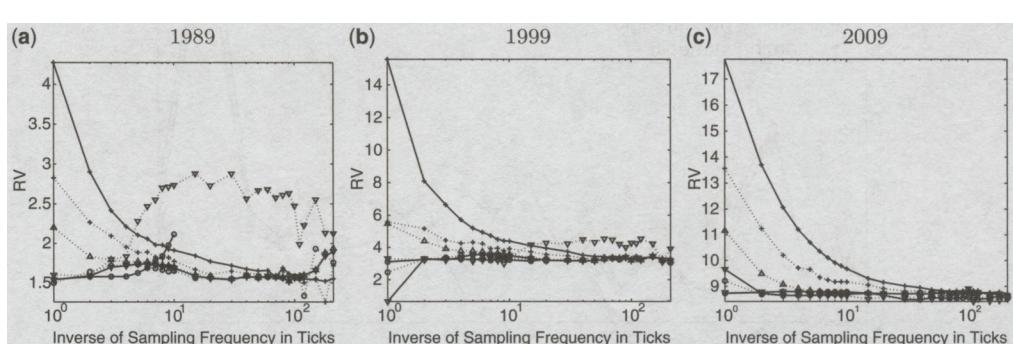


FIGURE 9  
Volatility signature plots of oil futures by year  
*Note:* The legend is the same as in Figure 7.

July. There is no Monday effect. Instead, the volatility peaks on Wednesdays—where it is about 20% higher than on Mondays.

The fluctuations of the IV estimates over the years summarize the recent history of oil prices. In Figure 11 the average daily volatility of oil futures first spikes in 1990, when the world was faced with the Gulf War. After four calm years, 1992–1995, it plateaued at an intermediate level from 1996 to 2007, despite the steep increase in oil prices. During the subsequent financial crisis the volatility of oil futures reached unprecedented levels. As of 2010, the volatility is back to the plateau level from before the financial crisis. Given the seasonal pattern in average daily realized volatility, the 2010 value has to be adjusted upwards by a factor of about 1.5, because our dataset ends just before the Fall of 2010. Even then, however, there is no clear evidence of excess volatility in oil prices at the volatility level of 2010. Based on the data available, regulation of derivatives in the oil market has to be justified with the volatility during the crisis years 2008/2009—not with

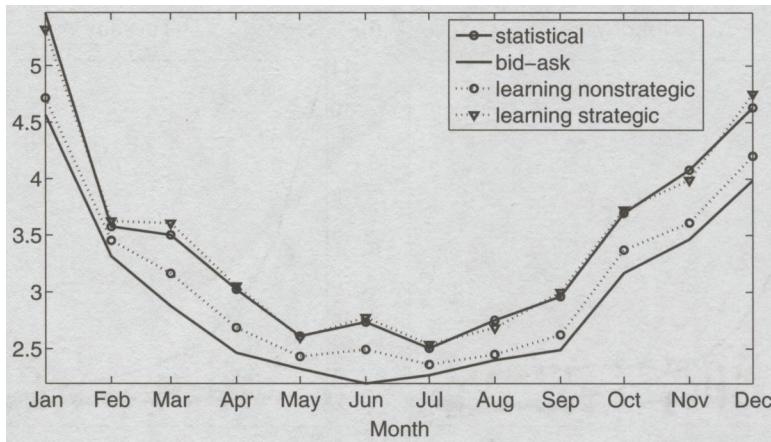


FIGURE 10  
Integrated variance of oil futures by calendar month, 1987–2010

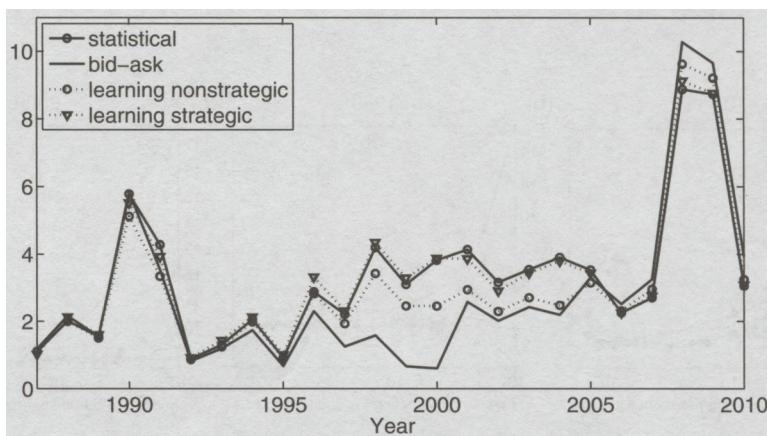


FIGURE 11  
Integrated variance of oil futures by calendar year, 1987–2010

the most recent data—, or with a destabilizing effect of specific groups of traders on the market microstructure. For example, unlike in previous years, in 2009 the IV estimate correcting for strategic learning is smaller than the one for non-strategic learning (Figure 9). This indicates that strategic trading based on private information rather than noise trading increased high-frequency market volatility during this period.

The four estimators show a similar volatility path over time. Numerically, however, they differ considerably. *RV Bid-ask* and *RV Learning—Non-strategic* estimate IV to be lower from the mid 1990s to the mid 2000s, and higher during 2008/2009 than the other two estimators. The switch between *RV Learning—Non-strategic* and *RV Statistical* around 2006 precedes the financial crisis; it suggests that around that time the market structure changed. Noise different from learning first increased volatility, but damped it during the financial crisis. What type of MSN can explain this change is an interesting question for further research. For example, some

links between the strong form efficient price and noise in addition to learning, *e.g.* debt-financed trading, might have been muted during the crisis.

## 7. CONCLUDING REMARKS

The recent realized volatility literature provides statistical insights into MSN and its effects. In this article we have provided complementary *economic* insights, treating MSN not simply as a nuisance, but rather as the result of financial economic decisions, which we seek to understand. In that regard, we derived the predictions of economic theory regarding correlation between MSN and two types of latent price, characterizing and contrasting entire cross-correlation functions in a variety of market environments, with a variety of results. We achieved this not in a new model of market microstructure, but rather in the context of several classic and widely used benchmark models.

Some of our results are generic. For example, cross-correlations between strong form efficient price and MSN at displacements greater than 0 have sign opposite to that of the contemporaneous correlation. Some of our results are not generic but nevertheless quite robust to model choice. For example, all models predict negative contemporaneous correlation between latent price and MSN, so long as the risk aversion of market makers is not too high. Finally, some of our results are highly model-specific. For example, the cross-correlation patterns and absolute magnitudes depend critically on the frequency of latent price changes, the presence of a bid/ask bounce, the timing of information and actions, and the degree of market-maker risk aversion.

We see our results as a first step towards disciplining empirical financial-econometric analyses with microstructure theory. In particular, we have emphasized that benchmark microstructure models can be used to control for MSN in volatility estimation, and that attention to sampling frequency is important for empirical microstructure studies. Moreover, we have shown that microstructure theory enables us to assess the validity of the independence assumption, to offer explanations of empirically observed cross-correlation patterns, to predict the existence of as-yet undiscovered patterns, and to make informed suggestions for improving volatility estimation methods.

Other novel uses of our results may also be possible. For example, the rate of cross-correlation decay might be used to assess the extent to which strategic traders are active in the market, and the sign and size of the contemporaneous correlation might be used to assess the degree of market-maker risk aversion. Indeed market-maker risk aversion might be time-varying, with associated time-varying cross-correlation structure between latent price and MSN. During crises, *e.g.*, market makers may be more risk averse, as borrowing and hedging possibilities are reduced. If so, the “normal pattern” of negative contemporaneous cross-correlation and positive higher-order cross-correlations might switch to a “crisis pattern” of positive contemporaneous cross-correlation and negative higher-order cross-correlations.

We hasten to emphasize, however, that although our results arise most naturally from learning and market-maker risk aversion in the classic benchmark models that we explore, those models are of course very simple, and other mechanisms might generate similar results. In future work beyond the scope of the present article, one might explore richer market microstructure models, incorporating, *e.g.*, strategic behaviour among informed traders with differential information, preference heterogeneity, etc. Ultimately one might attempt to develop and use an encompassing microstructure model that simultaneously includes a variety of such effects. Intertwined future empirical work on noise-corrected volatility estimation might exploit high-frequency data on additional aspects of price determination, such as volume, trade-initiation, and order-book data. Ultimately we hope that our framework may be useful not only for disciplining estimation

with theory, but also for disciplining theory with estimation, providing robust evidence on the comparative merits of various competing theoretical microstructure models.

*Acknowledgments.* For helpful comments and suggestions, far beyond the call of duty, we thank the Co-Editor (Enrique Sentana) and three anonymous referees. We are also grateful to Rob Engle, Joachim Grammig, Peter Hansen, Charles Jones, Eugene Kandel, Gideon Saar, and Frank Schorfheide, as well as participants at the NBER Conference on Market Microstructure, the Oxford-Man Institute Conference on the Financial Econometrics of Vast Data, and the Annual Meeting of the German Finance Association. We thank Ole Barndorff-Nielsen, Peter Hansen, Asger Lunde, and Neil Shephard for sharing their data.

## REFERENCES

- AGHION, P., BOLTON, P., HARRIS, C. and JULLIEN, B. (1991), "Optimal Learning by Experimentation", *Review of Economic Studies*, **58**, 621–654.
- AGHION, P., ESPINOSA, M. P. and JULLIEN, B. (1993), "Dynamic Duopoly with Learning through Market Experimentation", *Economic Theory*, **3**, 517–539.
- ÄIT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2005), "How Often to Sample a Continuous-time Process in the Presence of Market Microstructure Noise", *Review of Financial Studies*, **18**, 351–416.
- ÄIT-SAHALIA, Y., MYKLAND, P. A. and ZHANG, L. (2011), "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise", *Journal of Econometrics*, **160**, 160–175.
- ANDERSEN, T. G., BOLLERSLEV, T., CHRISTOFFERSEN, P. F. and DIEBOLD, F. X. (2006), "Volatility and Correlation Forecasting", in Elliot, G., Granger, C. W. J. and Timmermann, A. (eds) *Handbook of Economic Forecasting*, (Amsterdam: North-Holland) 777–878.
- ANDERSEN, T. G., BOLLERSLEV, T., CHRISTOFFERSEN, P. F. and DIEBOLD, F. X. (2013), "Financial Risk Measurement for Financial Risk Management", in Constantinides, G. M., Harris, M. and Stulz, R. M. (eds) *Handbook of the Economics of Finance*. Vol. 2A of *Handbooks in Economics*, Ch. 17, (Amsterdam: North Holland) 1127–1220.
- ANDERSEN, T. G., BOLLERSLEV, T. and DIEBOLD, F. X. (2010), "Parametric and Nonparametric Volatility Measurement", in Hansen, L. P. and Ait-Sahalia, Y. (eds) *Handbook of Financial Econometrics*, (Amsterdam: North-Holland) 67–137.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and EBENS, H. (2001a), "The Distribution of Realized Stock Return Volatility", *Journal of Financial Economics*, **61**, 43–76.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2000), "Great Realizations", *Risk*, **13**, 105–108.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2001b), "The Distribution of Realized Exchange Rate Volatility", *Journal of the American Statistical Association*, **96**, 42–55.
- ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and LABYS, P. (2003), "Modeling and Forecasting Realized Volatility", *Econometrica*, **71**, 579–625.
- BANDI, F. M. and RUSSELL, J. R. (2006), "Realized Variance and Market Microstructure Noise: Comment", *Journal of Business & Economic Statistics*, **24**, 167–173.
- BANDI, F. M. and RUSSELL, J. R. (2008), "Microstructure Noise, Realized Variance, and Optimal Sampling", *Review of Economic Studies*, **75**, 339–369.
- BANDI, F. M. and RUSSELL, J. R. (2011), "Market Microstructure Noise, Integrated Variance Estimators, and the Accuracy of Asymptotic Approximations", *Journal of Econometrics*, **160**, 145–159.
- BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2008), "Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise", *Econometrica*, **76**, 1481–1536.
- BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2009), "Realized Kernels in Practice: Trades and Quotes", *Econometrics Journal*, **12**, C1–C32.
- BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2011a), "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.
- BARNDORFF-NIELSEN, O. E., HANSEN, P. R., LUNDE, A. and SHEPHARD, N. (2011b), "Subsampling Realised Kernels", *Journal of Econometrics*, **160**, 204–219.
- BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002a), "Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models", *Journal of the Royal Statistical Society B*, **64**, 253–280.
- BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002b), "Estimating Quadratic Variation Using Realized Variance", *Journal of Applied Econometrics*, **17**, 457–477.
- BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2007), "Variation, Jumps, Market Frictions and High Frequency Data in Financial Econometrics", in Blundell, R., Torsten, P. and Newey, W. K. (eds) *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of Econometric Society*, (Cambridge: Cambridge University Press) 328–372.
- DIAMOND, D. W. and VERRECCHIA, R. E. (1987), "Constraints on Short-selling and Asset Price Adjustment to Private Information", *Journal of Financial Economics*, **18**, 277–311.
- DIEBOLD, F. X. (2006), "Realized Variance and Market Microstructure Noise: Comment", *Journal of Business & Economic Statistics*, **24**, 181–183.

- DIEBOLD, F. X. and STRASSER, G. (2010), "On the Correlation Structure of Microstructure Noise: A Financial Economic Approach", (NBER: Working Paper 16469).
- EASLEY, D. and O'HARA, M. (1987), "Price, Trade Size, and Information in Securities Markets", *Journal of Financial Economics*, **19**, 69–90.
- EASLEY, D. and O'HARA, M. (1992), "Time and the Process of Security Price Adjustment", *Journal of Finance*, **47**, 576–605.
- ENGLE, R. F. and SUN, Z. (2007), "When is Noise not Noise – A Microstructure Estimate of Realized Volatility", Manuscript, Stern School, New York University.
- GATHERAL, J. and OOMEN, R. C. A. (2010), "Zero-intelligence Realized Variance Estimation", *Finance and Stochastics*, **14**, 249–283.
- GLOSTEN, L. R. and HARRIS, L. E. (1988), "Estimating the Components of the Bid/Ask Spread", *Journal of Financial Economics*, **21**, 123–142.
- GLOSTEN, L. R. and MILGROM, P. R. (1985), "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders", *Journal of Financial Economics*, **14**, 71–100.
- HANSEN, P. R., LARGE, J. and LUNDE, A. (2008), "Moving Average-based Estimators of Integrated Variance", *Econometric Reviews*, **27**, 79–111.
- HANSEN, P. R. and LUNDE, A. (2006), "Realized Variance and Market Microstructure Noise", *Journal of Business & Economic Statistics*, **24**, 127–161.
- HASBROUCK, J. (1993), "Assessing the Quality of a Security Market: A New Approach to Transaction-cost Measurement", *Review of Financial Studies*, **6**, 191–212.
- HASBROUCK, J. (1996), "Modeling Market Microstructure Time Series", in Maddala, G. S. and Rao, C. R. (eds) *Statistical Methods in Finance*, (Amsterdam: North Holland) 647–692.
- HASBROUCK, J. (2002), "Stalking the "Efficient Price" in Market Microstructure Specifications: an Overview", *Journal of Financial Markets*, **5**, 329–339.
- HASBROUCK, J. (2007), *Empirical Market Microstructure* (New York: Oxford University Press).
- KELLY, D. L. and STEIGERWALD, D. G. (2004), "Private Information and High-frequency Stochastic Volatility", *Studies in Nonlinear Dynamics and Econometrics*, **8**, 1–28.
- KYLE, A. S. (1985), "Continuous Auctions and Insider Trading", *Econometrica*, **53**, 1315–1336.
- MEDDAHI, N. (2002), "A Theoretical Comparison between Integrated and Realized Volatility", *Journal of Applied Econometrics*, **17**, 479–508.
- O'HARA, M. (1995), *Market Microstructure Theory* (Cambridge: Cambridge University Press).
- OOMEN, R. C. A. (2006), "Realized Variance and Market Microstructure Noise: Comment", *Journal of Business & Economic Statistics*, **24**, 195–202.
- PATTON, A. J. (2011), "Data-based Ranking of Realised Volatility Estimators", *Journal of Econometrics*, **161**, 284–303.
- ROLL, R. (1984), "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market", *Journal of Finance*, **39**, 1127–1139.
- ZHANG, L., MYKLAND, P. A. and AÏT-SAHALIA, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High Frequency Data", *Journal of the American Statistical Association*, **100**, 1394–1411.
- ZHOU, B. (1996), "High-frequency Data and Volatility in Foreign-exchange Rates", *Journal of Business & Economic Statistics*, **14**, 45–52.