HOW RELEVANT IS VOLATILITY FORECASTING FOR FINANCIAL RISK MANAGEMENT?

Peter F. Christoffersen and Francis X. Diebold*

Abstract—It depends. If volatility fluctuates in a forecastable way, volatility forecasts are useful for risk management (hence the interest in volatility forecastability in the risk management literature). Volatility forecastability, however, varies with horizon, and different horizons are relevant in different applications. Moreover, existing assessments of volatility forecastability are plagued by the fact that they are joint assessments of volatility forecastability and an assumed model, and the results can vary not only with the horizon but also with the assumed model. To address this problem, we develop a model-free procedure for assessing volatility forecastability across horizons. Perhaps surprisingly, we find that volatility forecastability decays quickly with horizon. Volatility forecastability—although clearly of relevance for risk management at the short horizons relevant for, say, trading desk management—may be much less important at longer horizons.

I. Introduction

Many private-sector firms engage in risk management. In the financial services industry in particular, both interest and capability in risk management are expanding rapidly. Particularly active areas include investment banking, commercial banking, and insurance.[1] Interest has similarly escalated on the regulatory side, as governments around the world seek to impose risk-based capital adequacy standards.[2] It is not an exaggeration to say that risk management has emerged as a major industry in the last ten years, with outlets such as *Risk Magazine* bridging academe and industry.

Portfolio risk depends on the holding period, or *horizon*. But what is the relevant horizon for risk management? This obvious question has no obvious answer. Perusal of the industry literature reveals widespread discussion of the importance of the horizon, disagreement as to the relevant horizon, and, interestingly, an emerging recognition that fairly long horizons are relevant in many applications. Smithson and Minton (1996, p. 39), for example, note that ''nearly all risk managers believe the one-day . . . approach is valid for trading purposes. However, they disagree on the appropriate holding period for the long-term solvency of the institution.'' Chew (1994, p. 65) elaborates, asking whether any short holding period is relevant for risk controllers. McNew (1996, p. 56) makes a precise recommendation, arguing that ''if corporate America were to apply [modern financial risk management techniques] to its asset/liability risk management problem, it is probable that the time horizon would not be less than one quarter and could be significantly longer.'' Locke (1999) reports on the recent development of corporate risk measurement and management systems with a horizon between one and twelve months. Finally, institutional investors in Falloon (1999) argue that the appropriate horizon for investors, as opposed to market makers, is approximately one year, and that the appropriate horizon for pensions funds may be as long as ten years.

The upshot, of course, is that there is no one ''relevant'' horizon, so that thought must be given to the relevant horizon on an application-by-application basis. The relevant horizon will, in particular, likely vary with orientation (such as, public/regulatory versus private/for-profit), position in the firm (such as trading desk versus CFO), asset class (such as equity versus fixed income), and industry (such as banking versus insurance). These considerations lead to an important insight: Although very short horizons may be appropriate for certain tasks (such as managing the risk of a trading desk), much longer horizons may be relevant in other contexts.

There is little doubt that volatility is forecastable on a very high-frequency basis, such as hourly or daily.[3] Interestingly, however, much less is known about volatility forecastability at longer horizons, and, more generally, the pattern and speed of decay in volatility forecastability as we move from short to long horizons. Thus, open and key questions remain for risk management at all but the shortest horizons. How forecastable is volatility at various horizons? With what speed and pattern does forecastability decay as horizon lengthens? Are the recent advances in volatility modeling and forecasting—such as GARCH, stochastic volatility, and related approaches—useful for risk management at longer horizons, or is longer-horizon volatility approximately constant?

One approach to answering these questions involves estimating the path of short-horizon volatility and using it to infer the properties of long-horizon volatility. The simplest implementation of this temporal-aggregation idea is the popular industry practice of ''scaling up'' high-frequency volatility estimates to get a low-frequency volatility estimate (for example, converting a one-day return standard deviation to a thirty-day return standard deviation by multiplying by $\sqrt{30}$). Unfortunately, except under restrictive and routinely violated conditions, scaling is misleading and tends to produce spurious magnification of volatility fluctuations

Received for publication January 19, 1999. Revision accepted for publication August 9, 1999.

* McGill University, and Stern School, NYU, University of Pennsylvania and National Bureau of Economic Research, respectively.

We thank the editor (John Campbell) for helpful comments, as well as seminar participants at Princeton University, the University of Virginia, the Bank of England, the Bank of Italy, the Federal Reserve Bank of New York, Goldman Sachs, London Business School, the NBER Conference on Market Microstructure, and the Newton Institute Workshop on Econometrics and Financial Time Series. Rob Engle, Greg Hopper, Eric Jacquier, Bruce Lehman, and Keith Sill provided helpful discussion. All remaining inadequacies are ours alone. We thank the National Science Foundation, FCAR, and SSHRC for support.

[1] See Santomero (1995, 1997) and Babbel and Santomero (1997).

[2] See, for example, Kupiec and O'Brien (1995).

[3] See, for example, Bollerslev et al. (1992).

The Review of Economics and Statistics, February 2000, 82(1): 12–22

with horizon, as shown by Diebold, Hickman, Inove, and Schuermann (1998).

A more appropriate temporal-aggregation strategy is to fit a model to high-frequency data and, conditional upon the truth of the fitted model, use it to infer the properties of lower-frequency data. Drost and Nijman (1993), for example, provide temporal-aggregation formulae for the weak GARCH(1,1) process. That approach has at least two drawbacks, however. First, the aggregation formulae assume the truth of the fitted model (when in fact the fitted model is simply an approximation), and the best approximation to h-day volatility dynamics is not likely to be what one gets by aggregating the best approximation (let alone a mediocre approximation) to one-day volatility dynamics.[4] Second, temporal-aggregation formulae are presently available only for restrictive classes of models; the literature has progressed little since Drost and Nijman.

An alternative strategy is simply to fit volatility models directly to returns at various horizons of interest, thereby avoiding temporal-aggregation entirely. The idea of working directly at the horizons of interest is a good one, but, unfortunately, different families of parametric volatility models may produce different conclusions about forecastability, as in Hsieh (1993). What we really want, then, is a way to assess volatility forecastability directly from observed returns at various horizons without conditioning on an assumed model. In this paper, we propose a method for doing so, and we use it to assess patterns of volatility forecastability in equity, foreign exchange, and bond markets, with surprising results. We proceed as follows. In section II, we describe in detail our framework for model-free evaluation of volatility forecastability, and then, in section III, we use our methods to assess the volatility forecastability for returns on four major equity indexes, four major dollar exchange rates, and the U.S. ten-year Treasury bond, at horizons ranging from one through twenty trading days. In section IV, we offer concluding remarks and directions for future research.

II. Methods

In this section, we describe and assemble the tools necessary for a workable strategy of model-free assessment of volatility forecastability in risk management contexts. First, we sketch the intuition and give a precise statement of our methods. In particular, we show that recently developed tests of conditional calibration of interval forecasts can be used to provide model-free assessments of volatility forecastability. Next, we develop a formal test of volatility forecastability. Finally, we propose a natural and complementary measure of the strength of volatility forecastability, and we sketch a strategy for its estimation and inference.

[4] See Findley (1983), Weiss (1991), and Tiao and Tsay (1994).

A. *Model-Free Assessment of Volatility Forecastability*

Our strategy for assessing volatility forecastability is intimately connected to assessing the adequacy of interval forecasts. Christoffersen (1998) develops a framework for evaluating the adequacy interval forecasts, and our methods build directly on his. Suppose that we observe a sample path $\{y_t\}_{t=1}^{T}$ of the time series y_t and a corresponding sequence of one-step-ahead interval forecasts, $\{(L_{t|t-1}(p), U_{t|t-1}(p))\}_{t=1}^{T}$, where $L_{t|t-1}(p)$ and $U_{t|t-1}(p)$ denote the lower and upper limits of the interval forecast for time t made at time $t-1$ with desired coverage probability p. We define the hit sequence I_t as

$$I_t = \begin{cases} 1, & \text{if } y_t \in [L_{t|t-1}(p), \quad U_{t|t-1}(p)] \\ 0, & \text{otherwise,} \end{cases}$$

for $t = 1, 2, \ldots, T$. We say that a sequence of interval forecasts has correct *unconditional coverage* if $E[I_t] = p$ for all t; that is the standard notion of "correct coverage."

Correct unconditional coverage is appropriately viewed as a necessary condition for adequacy of an interval forecast. It is not sufficient, however. In particular, in the presence of conditional heteroskedasticity, it is important to check for adequacy of conditional coverage, which is a stronger concept. We say that a sequence of interval forecasts has *correct conditional coverage with respect to an information set* Ω_{t-1} if $E[I_t|\Omega_{t-1}] = p$ for all t. Correct conditional coverage trivially implies correct unconditional coverage; correct unconditional coverage is simply correct conditional coverage with respect to an empty information set. Christoffersen (1998) shows that, if $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, \ldots, I_1\}$, then correct conditional coverage is equivalent to $\{I_t\} \overset{\text{i.id.}}{\sim}$ Bernoulli(p), which can readily be tested.

Having given some background on interval forecast evaluation, now let us proceed to our ultimate goal: the development of tools for model-free assessment of volatility forecastability. Assume that the process y whose volatility forecastability we want to assess is covariance stationary, and, without loss of generality, assume a zero mean. Pick a constant interval symmetric around zero, $[-c, c]$.[5,6] The key insight is that—although the interval $[-c, c]$ is *unconditionally* correctly calibrated at some unknown confidence level, p—it is not *conditionally* correctly calibrated if volatility is forecastable. More precisely, if we measure volatility by the conditional variance, then we know that, if the conditional variance adapts to the evolving information set given by $\{y_{t-1}, y_{t-2}, \ldots, y_1\}$, a fixed-width confidence interval could

[5] Any value of c could be chosen, but typical values would be in range of one or two unconditional standard deviations of y. One could also use an asymmetric interval, but we shall not pursue that idea here.

[6] From a risk management perspective, it might seem curious to use a two-sided interval and thus score a hit when either an extreme left or an extreme right-tail event occurs. We do this deliberately, however, to enhance our power to detect volatility clustering, which is an inherently symmetric phenomenon. Using a test based on a one-sided interval would reduce power substantially.

not be correctly conditionally calibrated, because it fails to widen when the conditional variance rises and to narrow when the conditional variance falls.

The implied strategy for evaluating volatility forecastability is obvious: We know that confidence intervals of the form $[-c, c]$ are correctly unconditionally calibrated at some level, but we don't know whether they are correctly conditionally calibrated, which is to say we don't know whether volatility is forecastable. If the $[-c, c]$ intervals are not only correctly *unconditionally* calibrated, but also correctly *conditionally* calibrated, then volatility is not forecastable, and the hit sequence is i.i.d.[7]

B. *Assessing Independence of the Hit Sequence: A Runs Test*

We have seen that nonforecastability of volatility corresponds to an i.i.d. hit sequence; we now describe a convenient and powerful model-free runs test for testing independence of the hit sequence. The runs test dates at least to Wolfowitz (1943) and David (1947). It has been applied extensively in quality control engineering (for example, Grant and Leavenworth (1988)), and it can be viewed as an application of categorical data analysis (for example, Andersen (1994)).

Define a run as a string of consecutive zeros or ones in the hit sequence.[8] Let r be the number of runs, and let n_0 and n_1 be the total number of zeros and ones in the sequence. Then $T = n_0 + n_1$, and if R is the maximum number of runs possible, then

$$R = \begin{cases} 2 \min\{n_0, n_1\}, & \text{if } n_0 = n_1 \\ 2 \min\{n_0, n_1\} + 1, & \text{otherwise.} \end{cases}$$

Under the null hypothesis that $\{I_t\}_{t=1}^{T}$ is a random sequence, the distribution of the number of runs, r, given n_1 and n_0, is (for $\min\{n_0, n_1\} > 0$)

$$\Pr(r|n_0, n_1) = \frac{f_r}{\binom{T}{n_0}}, \qquad \text{for } r = 2, 3, \ldots, R,$$

where

$$f_{r=2s} = 2\binom{n_0 - 1}{s - 1}\binom{n_1 - 1}{s - 1} \qquad \text{and}$$

$$\begin{aligned} f_{r=2s+1} &= \binom{n_0 - 1}{s}\binom{n_1 - 1}{s - 1} \binom{n_0 - 1}{s - 1}\binom{n_1 - 1}{s} \\ &= \frac{f_{2s}(T - 2s)}{2s}. \end{aligned}$$

This distributional result provides a handy test of independence of the hit sequence; notice that it does not depend on the nominal coverage of the intervals, p. Moreover, the runs test is exact, and it is uniformly most powerful against a first-order Markov alternative.[9]

We conduct a small Monte Carlo experiment to assess the nominal size and power of the runs test in a realistic setting. We generate 1,000 daily return samples of size 6,350, which matches the returns series studied in our subsequent empirical work. We then aggregate the one-day returns to h-day nonoverlapping returns, $h = 2, 3, \ldots, 20$, and we assess the independence of each of the h-day returns series using the runs test. We use four data-generating processes. The first is simply i.i.d. Gaussian noise, which allows us to check whether the test is correctly sized. The remaining three have forecastable volatility: GARCH(1, 1) with Gaussian innovations; GARCH(1, 1) with Student's-t innovations; and the IGARCH process from J. P. Morgan's RiskMetrics. We use highly persistent GARCH processes ($\alpha + \beta = 0.99$, in the standard GARCH notation), as the volatility forecastability will otherwise be trivially negligible at the longer horizons.

The results, shown in figure 1, show that the test is correctly sized, with very high power at short horizons. The power does of course drop with horizon, but, even at a twenty-day horizon, corresponding to four weeks of trading, the power is reasonable. In figure 1, the width of the unconditional interval, c, is set to two (unconditional) standard deviations. In figure 2, we show power functions for different values of c, using the GARCH(1, 1)-t as the data-generating process throughout. We let c vary from one to two standard deviations in increments of a quarter. Power varies moderately with c, with highest power when c is approximately one and one-half standard deviations at each horizon. Nevertheless, we will set c to two standard deviations for most of this paper, as it yields an unconditional coverage of greater relevance to risk managers, with only a slight reduction in power.[10]

C. *Measuring Volatility Forecastability: Markov Transition Matrix Eigenvalues*

We now define a forecastability measure based on a first-order Markov alternative, which naturally complements the runs test of independence. Let the hit sequence be first-order Markov with arbitrary transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix},$$

where $\pi_{ij} = \Pr(I_t = j | I_{t-1} = i)$. The eigenvalues are solutions to the equation

$$\left| \lambda I - \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix} \right| = 0;$$

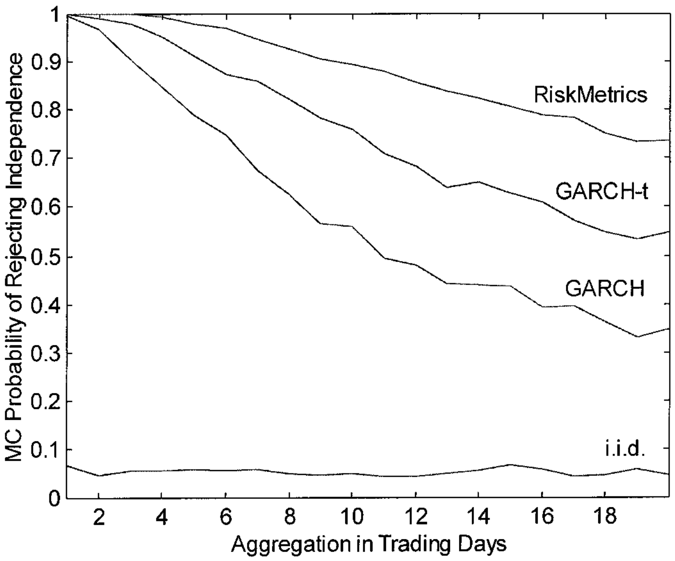
[7] It is interesting to note that tests based on the i.i.d. property of the hit sequence have power not only against volatility forecastability, but also against more-general forms of forecastability in the tail thickness of the conditional distribution, such as those modeled by Hansen (1994). This type of forecastability is equally important for risk managers, whose ultimate concern is not volatility per se, but rather the likelihood of tail events.

[8] For example, the sequence $\{I_t\}_{t=1}^{10} = \{0, 0, 1, 1, 1, 0, 1, 0, 0, 0\}$ has five runs.

[9] See Lehmann (1986) for details.

[10] Furthermore, the choice of c does not change the qualitative results of our subsequent empirical analysis, as we shall demonstrate.

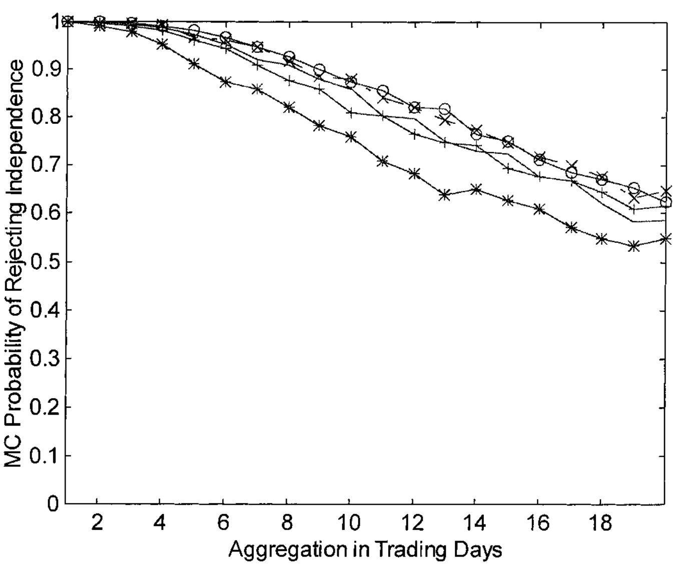
FIGURE 1.—SIZE AND POWER OF 5% RUN TESTS OF INDEPENDENCE FOR ±2 STANDARD DEVIATION INTERVALS UNDER VARIOUS DGPS



The graph shows Monte Carlo estimates of the probability of rejecting the null hypothesis of independence in a runs test with significance level set to 5%. The width of the interval defininining the hit sequence is $\pm 2\hat{\sigma}$. The data-generating processes for the daily data are the RiskMetrics or IGARCH process with $\beta = 0.94$; the GARCH(1, 1)-$t(d)$ process with $\alpha = 0.06$, $\beta = 0.93$, and $d = 5$; the Gaussian GARCH(1, 1) with $\alpha = 0.06$ and $\beta = 0.93$; and independent Gaussian innovations. All processes have 6,350 daily observations. The number of Monte Carlo replications is 1,000. See text for details.

the first eigenvalue is necessarily unity and therefore conveys no information regarding the forecastability of the hit sequence, and the second eigenvalue is simply $S \equiv \pi_{11} - \pi_{01}$. S is a natural persistence measure; note that under independence, $\pi_{01} = \pi_{11}$, so $S = 0$, and, conversely, under strong positive persistence π_{11} will be much larger than π_{01}, so S will be large.[11]

FIGURE 2.—POWER OF 5% RUN TESTS OF INDEPENDENCE FOR VARIOUS INTERVAL FORECASTS USING GARCH(1, 1)-$t(d)$ DATA



The graph shows Monte Carlo estimates of the probability of rejecting the null hypothesis of independence in a runs test with significance level set to 5%. The width of the interval forecast defininining the hit sequence is $\pm m\hat{\sigma}$, where $m = 1$ (—), $m = 1.25$ (ooo), $m = 1.5$ (xxx), $m = 1.75$ (+++), and $m = 2$ (***). The data-generating process for the daily data is the GARCH(1, 1)-$t(d)$ process with $\alpha = 0.06$, $\beta = 0.93$, and $d = 5$, with 6,350 daily observations. The number of Monte Carlo replications is 1,000. See text for details.

S has an alternative and intuitive motivation: It is the first-order serial correlation coefficient of the hit sequence. To see this, we note that[12]

$$E[I_t] = p = p\pi_{11} + (1 - p)\pi_{01} = \frac{\pi_{01}}{1 + \pi_{01} - \pi_{11}}$$

$$\text{Var}\,[I_t] = p(1 - p) = \frac{\pi_{01}(1 - \pi_{11})}{1 + \pi_{01} - \pi_{11}}$$

$$\begin{aligned}\text{Cov}\,(I_t, I_{t-1}) &= E[I_t I_{t-1}] - E^2[I_t] = p\pi_{11} - p^2 \\ &= p(\pi_{11} - p).\end{aligned}$$

Then we form the correlation coefficient and use some algebra to obtain

$$\begin{aligned}\text{Corr}\,(I_t, I_{t-1}) &= \frac{\pi_{11} - p}{1 - p} = \frac{\pi_{11}(1 + \pi_{01} - \pi_{11}) - \pi_{01}}{1 - \pi_{11}} \\ &= \pi_{11} - \pi_{01} = S.\end{aligned}$$

Thus, just as in the familiar $AR(1)$ case for which the root of the autoregressive lag-operator polynomial is the first-order serial correlation coefficient, so too in the first-order Markov case is the (nontrivial) eigenroot.[13]

D. Estimating the Markov Model

The discussion of forecastability measurement has thus far been in population; in practice, of course, one must estimate the relevant Markov models. Maximum-likelihood estimation is particularly simple. For a hit sequence $\{I_1, \ldots, I_T\}$, the likelihood function is immediately[14]

$$\begin{aligned}&L(\pi_{01}, \pi_{11}; I_1, I_2, \ldots, I_T) \\ &\quad = (1 - \pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1 - \pi_{11})^{n_{10}}\pi_{11}^{n_{11}},\end{aligned}$$

where n_{ij} is the number of observations with value I followed by j. The maximum-likelihood estimators of π_{01} and π_{11} are therefore $\hat{\pi}_{01} = n_{01}/(n_{00} + n_{01})$ and $\hat{\pi}_{11} = n_{11}/(n_{10} + n_{11})$. By Slutsky's theorem, the maximum likelihood estimate of the non-unit eigenvalue is then $\hat{S} = \hat{\pi}_{11} - \hat{\pi}_{01}$.

Unlike the exact finite-sample theory available for the runs test of independence, the theory associated with maximum-likelihood estimation of the transition matrix eigenvalue is only asymptotic. Thus, in an attempt to tailor

[11] Analogous use of eigenvalues as mobility measures has been suggested by Shorrocks (1978) and Sommers and Conlisk (1979).

[12] To evaluate the covariance, use the fact that $E[I_t I_{t-1}] = \Pr(I_t = 1 \cap I_{t-1} = 1) = p\pi_{11}$.

[13] See also Hamilton (1994, p. 687).

[14] As is standard, we form the likelihood conditional on the first observation, I_1.

our inference to precise sample sizes relevant for the application at hand, we use simulation methods to assess the significance of our eigenvalue estimates. In particular, for any returns series, we

1. De-mean the returns series.
2. Compute the hit sequence relative to the constant $\pm c$ interval, and then compute the estimate of p, $\hat{p}$, and the estimate of S, $\hat{S}$.
3. Use $\hat{p}$ and the relevant sample size T to
   a. Generate $m = 1, \ldots, M$ samples of i.i.d. Bernoulli ($\hat{p}$) pseudodata.
   b. Compute $\hat{S}_m$.
   c. Compute the 95%-confidence interval for $\hat{S}_m$ and plot it together with $\hat{S}$ computed in step 2.

E. Expanding the Information Set

Our analysis thus far focuses on assessing univariate first-order dependence in the hit sequence. We now broaden our methods to allow for multivariate and higher-order dependence, potentially using the highest-frequency data available (for example, daily), regardless of the return horizon.

Consider nonoverlapping h-day returns y_t, $t = 1, 2, 3, \ldots$. Let the conditional c.d.f. of demeaned h-day returns be

$$\Pr(y_t < c|\Omega_{t-1}) = F(c|\Omega_{t-1}),$$

and define

$$p_t = \Pr(|y_t| < c|\Omega_{t-1}) = F(c|\Omega_{t-1}) - F(-c|\Omega_{t-1}).$$

Assuming that the p.d.f. of y_t is symmetric, we can write

$$p_t = 2F(c|\Omega_{t-1}) - 1.$$

Notice that I_t can be conveniently defined as

$$I_t = 1(|y_t| < c),$$

where $1(\cdot)$ is the indicator function. We therefore have that

$$E[I_t|\Omega_{t-1}] = p_t = 2F(c|\Omega_{t-1}) - 1.$$

Thus, I_t can be viewed as the outcome of a limited dependent-variables regression,

$$I_t = \beta_0 + \beta_1 F(c|\Omega_{t-1}) + e_t,$$

in an unobserved variable, $F(c|\Omega_{t-1})$.

The regression representation is useful for a number of purposes. First, testing the null hypothesis of correct conditional coverage, $E[I_t|\Omega_{t-1}] = p$, for some p, corresponds to testing $\beta_1 = 0$ in the regression above, and thus involves only a simple F-test that all slopes are zero.[15] Second, the regression setup facilitates the inclusion of predictor variables measured at a frequency higher than h, such as lagged squared daily returns. Third, the regression facilitates allowance for higher-order dependence in the indicator sequence via simple inclusion of additional lags of the predictor variables.

III. Volatility Forecastability in Financial Asset Markets

Armed with the tools introduced above, we now proceed to measure volatility forecastability in global foreign exchange, stock, and bond markets. We examine asset return volatility forecastability as a function of the horizon over which the returns are computed, beginning with daily returns and proceeding through nonoverlapping h-day returns, $h = 1, 2, 3, \ldots, 20$.[16]

Because the unconditional volatility of all asset returns rises with the aggregation level, it is natural and appropriate to let the width of our fixed $[-c, c]$ intervals change with the aggregation level. We do so throughout; in fact, we use $c_h = 2\hat{\sigma}_h$ intervals to compute our hit sequences. This yields unconditional coverage in the range of 90% to 95%, which makes for a nice parallel to the value-at-risk (VaR) literature, which typically focuses on VaR in the range of 1% to 10%.

A. Equity and Foreign Exchange Markets

We begin by examining equity and foreign exchange rate returns. We examine returns on four broad-based stock indexes: the U.S. S&P 500, the German DAX, the U.K. FTSE, and the Japanese TPX. We examine returns on four dollar exchange rates: the German Mark, British Pound, Japanese Yen, and French Franc. The sample starts on January 1, 1973, and ends on May 1, 1997, resulting in 6,350 daily observations for each return series.[17]

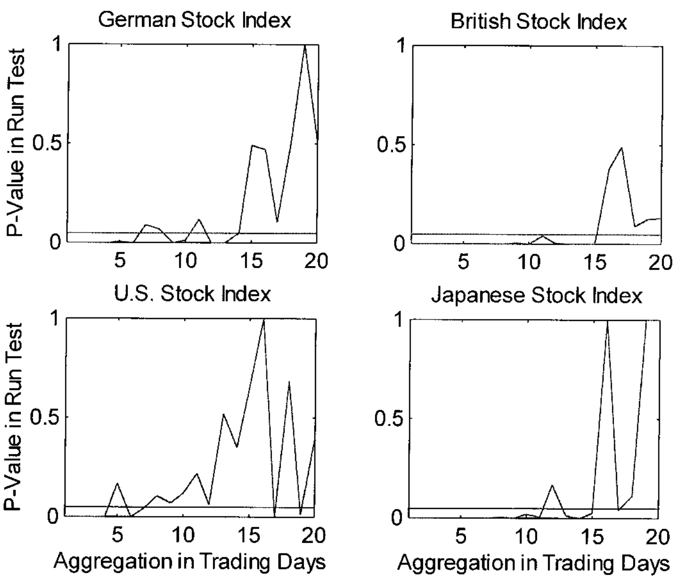
Let us first discuss the runs tests. Figure 3 shows the finite-sample p-values of the runs tests of independence of the hit sequence for equities, as a function of the horizon. It is clear that, for each equity index, the p-values tend to increase with the horizon, although the specifics differ somewhat depending on the particular index examined. As a rough rule of thumb, we summarize the results as saying that for horizons of less than ten trading days we tend to reject independence, which is to say that equity return volatility is

[15] To implement the testing procedure, we could attempt to find a functional form for $F(c|\Omega_{t-1})$, but we prefer a less-parametric approach in which we directly include elements of the information set, such as I_{t-1} or y_{t-1}^2, on the right-hand side. Hence, β_1 should be interpreted as a vector.

[16] Use of nonoverlapping returns eliminates the need to account for the dependence induced by overlapping observations.

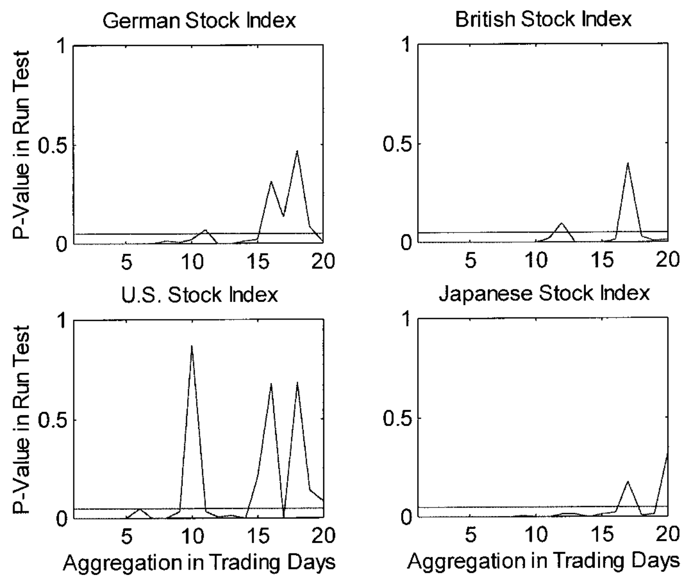
[17] The equity and foreign exchange data are from Datastream International. Equity prices are official local closing prices provided by the local exchanges, and we compute equity returns as logarithmic differences of those prices. Foreign exchange rates are averages of closing London bid and ask quotes; we compute foreign exchange returns as logarithmic differences of those exchange rates. The bond yields are from Bloomberg Financial Services and are averages of closing New York bid and ask quotes.

FIGURE 3.—p-VALUES OF RUNS TESTS FOUR EQUITY INDEXES ± 2 STANDARD DEVIATION INTERVAL FORECASTS



For each series and horizon, we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast. The horizontal line is at 5%. See text for details.

FIGURE 5.—p-VALUES OF RUNS TESTS FOUR EQUITY INDEXES ± 1.5 STANDARD DEVIATION INTERVAL FORECASTS



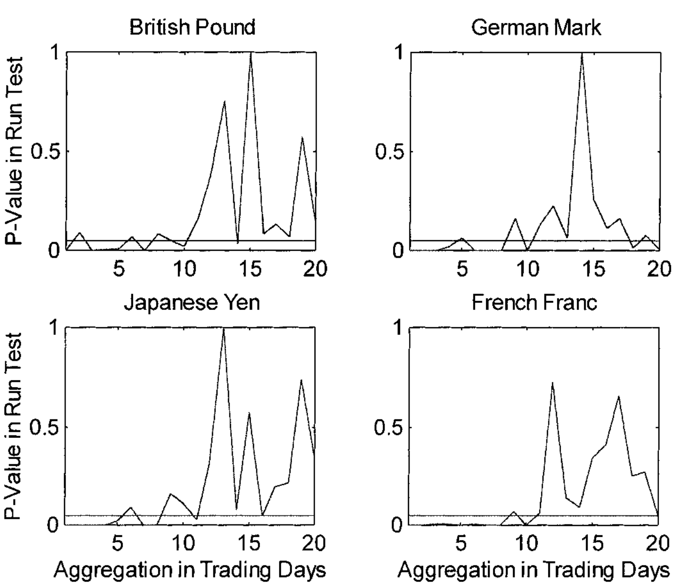
For each series and horizon, we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 1.5\hat{\sigma}$ interval forecast. The horizontal line is at 5%. See text for details.

significantly forecastable, and conversely for horizons greater than ten days. Figure 4 reveals identical patterns for exchange rates.

In our earlier Monte Carlo study, we found that the power of the runs test was slightly higher when the interval width was lowered to one and one-half standard deviations. We therefore now calculate the p-values of the runs test using the narrower intervals and plot the results in figure 5 and 6. Notice that virtually the same qualitative results are obtained, although the stock return volatility in some countries might be forecastable as far as three weeks ahead when looking at the narrower intervals. To save space, we will focus on the wider two-standard-deviation intervals, which are the most relevant for risk management.

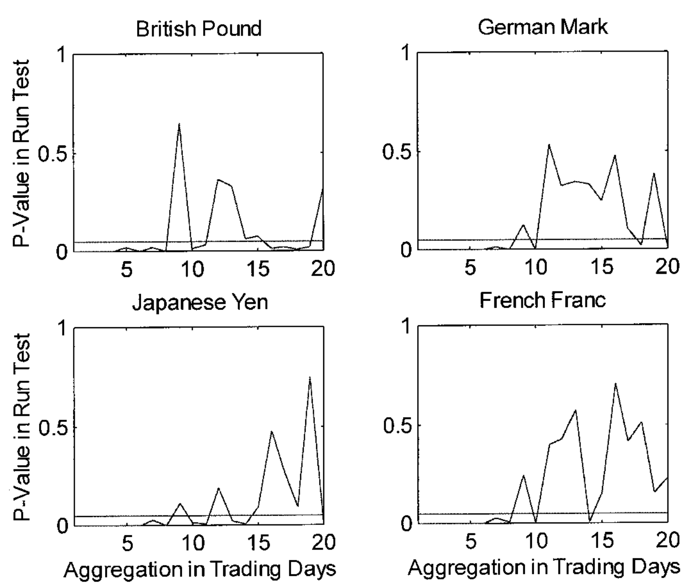
One difficulty with the runs test framework is its exclusive emphasis on testing for volatility forecastability, as opposed to measuring the strength of volatility forecastability. Presumably some volatility forecastability exists even at the longer

FIGURE 4.—p-VALUES OF RUNS TESTS FOUR DOLLAR EXCHANGE RATES ± 2 STANDARD DEVIATION INTERVAL FORECASTS



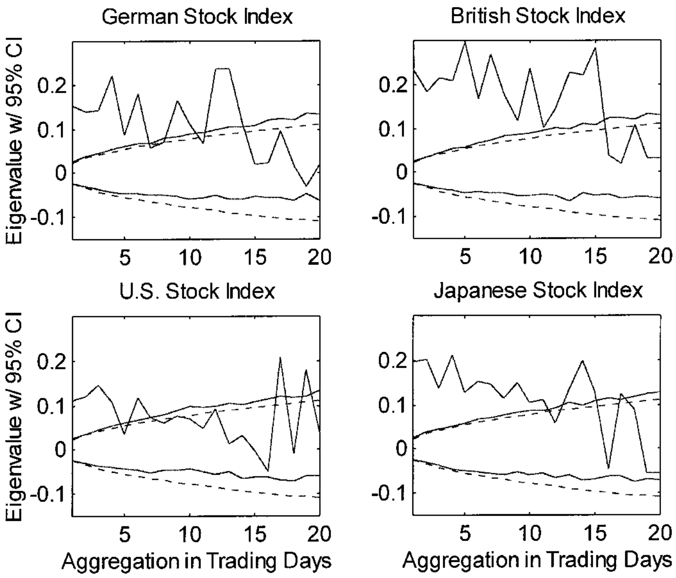
For each series and horizon, we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast. The horizontal line is at 5%. See text for details.

FIGURE 6.—p-VALUES OF RUNS TESTS FOUR DOLLAR-EXCHANGE RATES ± 1.5 STANDARD DEVIATION INTERVAL FORECASTS



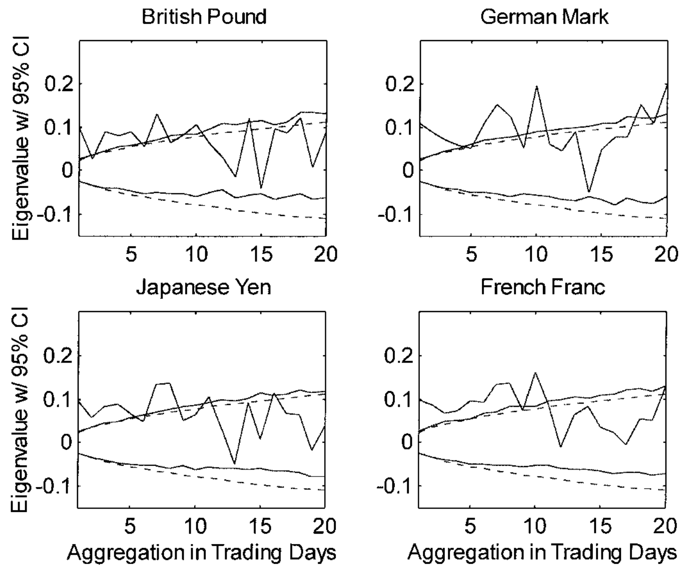
For each series and horizon, we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 1.5\hat{\sigma}$ interval forecast. The horizontal line is at 5%. See text for details.

FIGURE 7.—MARKOV TRANSITION MATRIX EIGENVALUES FOUR EQUITY INDICES



For each series and each horizon, we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast, along with the finite-sample 95%-confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4,000 simulations (solid lines) and asymptotic confidence intervals (dashed lines). See text for details.

FIGURE 8.—MARKOV TRANSITION MATRIX EIGENVALUES FOUR DOLLAR-EXCHANGE RATES



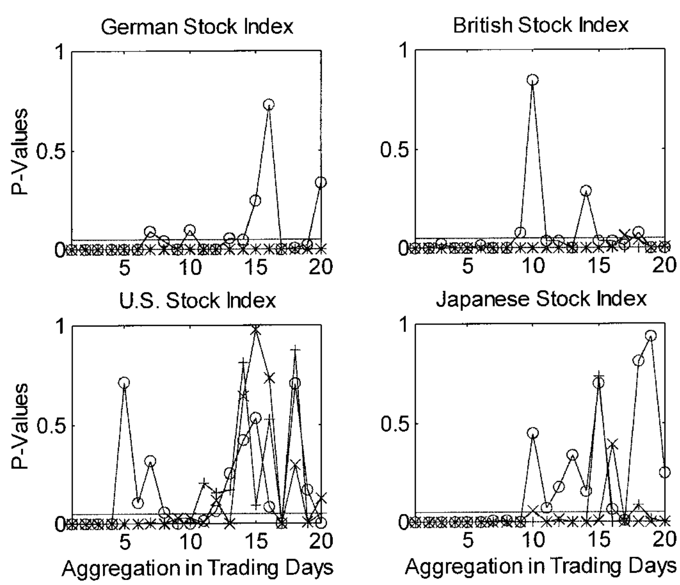
For each series and each horizon we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast, along with the finite-sample 95%-confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4,000 simulations (solid lines) and asymptotic confidence intervals (dashed lines). See text for details.

horizons, and the runs test would detect it if the sample size were larger. But again, our ultimate interest focuses not on the existence of volatility forecastability, but rather on its strength. Hence, we now turn to the estimated transition matrix eigenvalues, which measure the strength of volatility forecastability and are therefore more directly aligned with our ultimate concerns.

We show the estimated transition matrix eigenvalues along with their simulated finite-sample and asymptotic Bartlett 95%-confidence intervals, again as a function of horizon, in figure 7 and 8.[18] A consistent pattern emerges across all equities and exchange rates: At very short horizons, typically from one to ten trading days, the eigenvalues are significantly positive, but they decrease quickly, and approximately monotonically, with the horizon. By the time we reach one-day returns—and often substantially before—the estimated eigenvalues are small and statistically insignificant, indicating that volatility forecastability has vanished. Notice also the deterioration of the validity of the asymptotic confidence intervals as the horizon lengthens and the sample size shrinks.

Recognizing the potential limitations of testing the hit sequence for first-order dependence only, we now turn to a high-frequency, multiple-lag analysis. We first regress the hit sequence at each horizon on each of the following three high-frequency information sets: one to five lags of the hit sequence from one-day returns, one to five lags of the squared daily returns, and one to five lags of the daily RiskMetrics filtered volatility. Figure 9 and 10 plot the p-values from F-tests of the null hypothesis that the high-frequency information is irrelevant (that is, all slopes are zero in the limited dependent variable regression). For the dollar exchange rates, we see that the previous results hold without qualification: The indicator sequences from the returns are not forecastable beyond a ten-day horizon. For

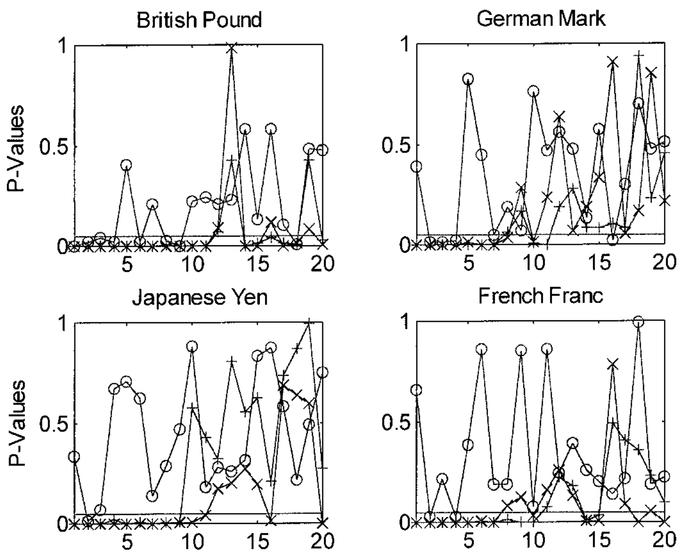
FIGURE 9.—p-VALUES FROM F-TEST OF HIGH-FREQUENCY INFORMATION FOUR EQUITY INDEXES



For each series and each horizon, we plot the p-values of F-tests associated with regressions on the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The high-frequency information sets are one to five lags of daily hits (xxx), one to five lags of daily squared returns (ooo), and one to five lags of daily RiskMetrics volatility (+++). The horizontal line denotes the 5% critical value. See text for details.

[18] The use of asymptotic Bartlett intervals is justified by the fact that the eigenvalue of interest is actually the first autocorrelation of the hit sequence.

FIGURE 10.—p-VALUES FROM F-TESTS OF HIGH-FREQUENCY INFORMATION FOUR DOLLAR-EXCHANGE RATES

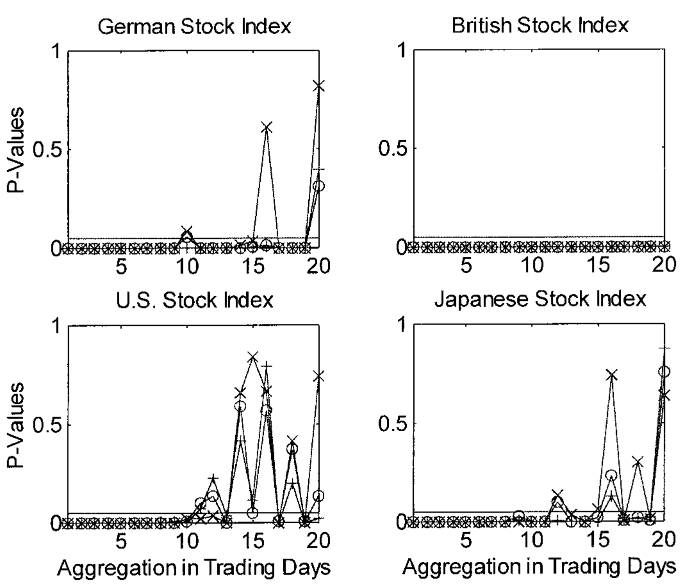


For each series and each horizon, we plot the p-values of F-tests associated with regressions on the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The high-frequency information sets are one to five lags of daily hits (xxx), one to five lags of daily squared returns (ooo), and one to five lags of daily RiskMetrics volatility (+++). The horizontal line denotes the 5% critical value. See text for details.

the equity index returns, however, some forecastability is detectable at longer horizons in the cases of Germany and the U.K. when the high-frequency information is used.

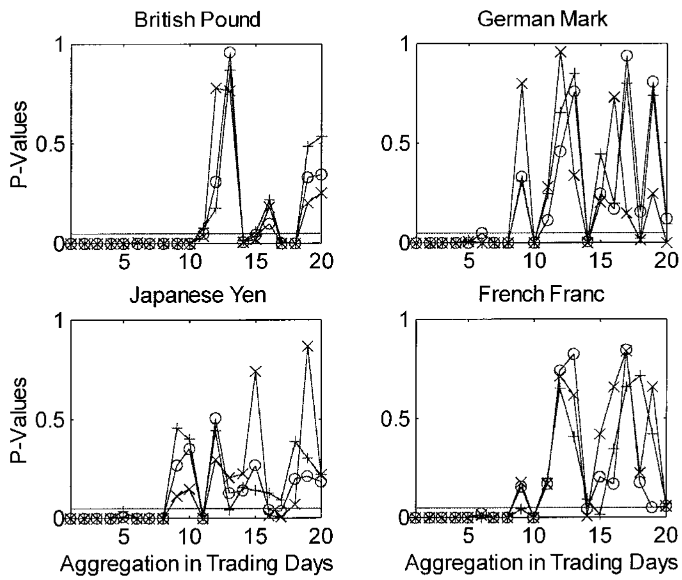
Finally, we test against longer-term dependence in the hit sequence at each horizon by conducting F-tests against the alternative that up to fifteen lags of the hit sequence at the horizon in question have explanatory power. The results are shown in figure 11 and 12. We plot the p-values of the F-test

FIGURE 11.—p-VALUES FROM F-TESTS OF HIGHER-ORDER DEPENDENCE FOUR EQUITY INDEXES



For each series and each horizon, we plot the p-values of F-tests associated with the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The higher-order dependencies in the hit sequence tested are one to five lags (xxx), one to ten lags (ooo), and one to fifteen lags (+++). The horizontal line is drawn at 5%. See text for details.

FIGURE 12.—p-VALUES FROM F-TESTS OF HIGHER-ORDER DEPENDENCE FOUR DOLLAR-EXCHANGE RATES



For each series and each horizon, we plot the p-values of F-tests associated with the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The higher-order dependencies in the hit sequence tested are one to five lags (xxx), one to ten lags (ooo), and one to fifteen lags (+++). The horizontal line is drawn at 5%. See text for details.

corresponding to the null hypothesis that one to five, one to ten, and one to fifteen lags of the hit sequence respectively are irrelevant for predicting the current hit sequence. Again, we see that the dollar exchange rates display the familiar pattern of volatility forecastability decaying quickly, while the stock returns in some cases—again notably Germany and the U.K.—display longer-run forecastability, as evidenced by stronger persistence in hit sequences.

B. Bond Markets

We report results for bonds separately for three reasons, with the first two linked to a priori concerns and the third due to the different nature of the results. First, historical bond market data typically contain only the annual yield and not the price, and it is not possible to calculate exact returns on a bond from yield alone. Thus, to compute bond returns, we are forced to make a potentially inaccurate approximation, which is not required to compute equity and exchange returns. Second, the available historical samples of bond yield data are much more limited. In fact, we analyze the returns of only one bond, the U.S. ten-year Treasury. Third, as we shall show, patterns of bond-market volatility forecastability are, at first sight, different from those in equity and foreign exchange markets and are therefore usefully discussed separately.

Let us begin with a yield-based approximation to a bond's return. Recall that the price of a bond that pays a coupon rate of C every period and \$1 at maturity after n periods is

$$P_{\text{cnt}} = C \sum_{i=1}^{n} \frac{1}{(1 + Y_{\text{cnt}})^i} + \frac{1}{(1 + Y_{\text{cnt}})^n},$$

where Y_{cnt} is the yield per period. Also recall that Macaulay's duration is defined by

$$D_{\text{cnt}} = \frac{\sum_{i=1}^{n} \frac{iC}{(1 + Y_{\text{cnt}})^i} + \frac{n}{(1 + Y_{\text{cnt}})^n}}{P_{\text{cnt}}},$$

which can also be written as[19]

$$D_{\text{cnt}} = -\frac{\Delta P_{\text{cnt}}}{\Delta(1 + Y_{\text{cnt}})} \frac{1 + Y_{\text{cnt}}}{P_{\text{cnt}}}.$$

Assume that the coupon rate is close to the yield, $C \approx Y_{cnt}$, in which case the bond will be priced near par, $P_{\text{cnt}} \approx 1$, resulting in the approximate duration[20]

$$D_{\text{cnt}} \approx \frac{1 - (1 + Y_{\text{cnt}})^{-n}}{1 - (1 + Y_{\text{cnt}})^{-1}}.$$

Finally, use the fact that $\Delta(1 + Y_{\text{cnt}}) = \Delta Y_{\text{cnt}}$ to rewrite the exact duration formula as

$$\frac{\Delta P_{\text{cnt}}}{P_{\text{cnt}}} = -\frac{D_{\text{cnt}} \Delta Y_{\text{cnt}}}{1 + Y_{\text{cnt}}},$$

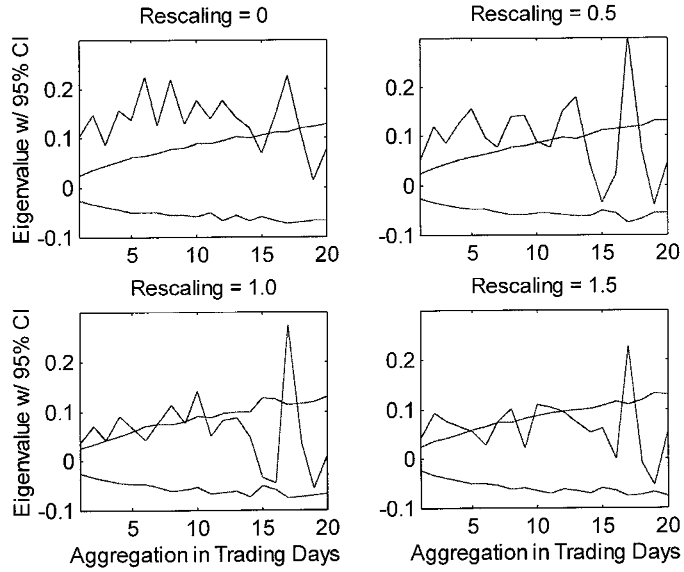
which, when combined with the approximate-duration formula, yields an approximation for returns as a function of only yield and time to maturity,

$$\frac{\Delta P_{\text{cnt}}}{P_{\text{cnt}}} \approx -\frac{\Delta Y_{\text{cnt}}}{1 + Y_{\text{cnt}}} \left(\frac{1 - (1 + Y_{\text{cnt}})^{-n}}{1 - (1 + Y_{\text{cnt}})^{-1}} \right).$$

Having arrived at a workable approximation to bond returns, we now examine the forecastability of bond-return volatility. Limited availability of historical daily international bond yield data forces us to focus exclusively on the ten-year U.S. Treasury bond. As before, the daily sample starts on January 1, 1973, and ends on May 1, 1997. The estimated Markov transition matrix eigenvalues, which appear in the top-left panel of figure 13, indicate substantially more volatility forecastability than in the equity or foreign exchange markets, with some forecastability as far ahead, say, as fifteen to twenty trading days.[21]

It is hard to determine whether the apparently greater bond-market volatility predictability is real. It could be an artifact of the approximation necessary to calculate bond returns. It could also be an artifact of the structural break in

FIGURE 13.—MARKOV TRANSITION MATRIX EIGENVALUES U.S. TEN-YEAR GOVERNMENT BOND RETURNS RESCALED BY VARIOUS POWERS OF THE BOND YIELD



For each horizon, we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast, along with the finite-sample 95%-confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4,000 simulations. The bond returns are rescaled by dividing by the bond yield levels taken to the power of 0, 0.5, 1, and 1.5, respectively. See text for details.

Federal Reserve policy around 1980, which could produce a spurious appearance of high volatility forecastability if not properly accounted for, as suggested by Diebold (1986) and verified by Lamoureux and Lastrapes (1990) and Hamilton and Susmel (1994). At any rate, our finding that volatility is more forecastable in bond markets than elsewhere is consistent with existing evidence, including Engle et al. (1987) and Andersen and Lund (1997).[22]

One intriguing possibility is that bond-return volatility dynamics are linked to those of the short-term interest rate, as in several well-known models of the yield curve, including Brennan and Schwartz (1979) and Cox (1985). Those models imply that a simple rescaling of bond returns by a function of the short-term yield will produce constant-volatility returns. In the spirit of this argument, we rescale the approximate bond returns above by a power of the bond yields,

$$\frac{\Delta P_{\text{cnt}}}{P_{\text{cnt}} Y_{\text{cnt}}^{\gamma}} \approx -\frac{\Delta Y_{\text{cnt}}}{1 + Y_{\text{cnt}}} \left(\frac{1 - (1 + Y_{\text{cnt}})^{-n}}{1 - (1 + Y_{\text{cnt}})^{-1}} \right) \Big/ Y_{\text{cnt}}^{\gamma}.$$

Following Campbell et al. (1997, p. 450), we set γ to 0.5, 1.0, and 1.5, respectively, and then reestimate the eigenvalues of the hit sequences corresponding to each of the rescaled bond returns. The results are shown in figure 13; in particular, when the returns are rescaled by yields to the power of 1 and 1.5, the eigenvalues plotted across horizons exhibit much less persistence than before, and remarkably resemble those found earlier for stock and foreign exchange returns.

[19] See, for example, Campbell et al. (1997, p. 403).

[20] This approximate-duration formula can also be derived as an exact duration in Campbell's approximate log-linear model. See Campbell et al. (1997, p. 408).

[21] The runs test p-values, which we omit to save space, tell the same story.

[22] See also the survey by Bollerslev et al. (1992).

IV. Concluding Remarks and Directions for Future Research

A. *Interpretation of Our Results*

If volatility is forecastable at the horizons of interest, then volatility forecasts are relevant for risk management. But our results indicate that, if the horizon of interest is more than ten or twenty days (depending on the asset class) volatility forecasts may not be of much importance. Our results clash with the assumptions embedded in popular risk management paradigms, which effectively assume highly forecastable volatility. J. P. Morgan's RiskMetrics, for example, is based on forecasts produced by exponentially smoothing squared returns, which are optimal only in the case of integrated volatility dynamics. Our results are, however, consistent with academic studies such as West and Cho (1995), who find that volatility forecasts are not of much importance in foreign exchange markets beyond a five-day horizon.[23] Our results are also in agreement with those of Jacquier et al. (1994, 1999), who use Bayesian methods to estimate stochastic volatility models and find that the posterior distributions have no appreciable mass near the unit root. This result contrasts with many classical studies that use maximum-likelihood estimation techniques and obtain estimates on the boundary of nonstationarity.

We would argue, moreover, that our results are consistent with those of a number of seemingly conflicting recent academic studies which fall into two groups. The first group documents slow decay in long-lag autocorrelations of squared or absolute returns, which indicates long-memory volatility dynamics and would seem to indicate forecastability of volatility at very long horizons (for example, Andersen and Bollerslev (1997)). But that literature tends to work with very high-frequency data—typically five-minute returns—and, although long memory in five-minute returns may well indicate that volatility is highly forecastable many steps into the future, perhaps 100 steps or even 1,000 steps—it does not necessarily indicate forecastability beyond ten or twenty days. For example, 1,000 five-minute steps are just more than three days; even 5,000 five-minute steps are just more than seventeen days.

The second group refutes evidence of the sort provided by Jorion (1995)—which seems to indicate that ARCH models provide poor volatility forecasts—by showing that volatility is much more forecastable when an appropriate measure of realized volatility is used (for example, Andersen and Bollerslev (1998)). That literature, however, focuses on one-day-ahead volatility forecasts, and certainly we agree that short-horizon volatility is highly forecastable. Our analysis, in contrast, focuses on longer-horizon volatility.

[23] The methods of West and Cho (1995), moreover, differ substantially from ours and therefore lend independent confirmation.

B. *What Next?*

We see two particularly interesting directions for future research. The first involves the use of economic, as opposed to statistical, metrics of volatility forecastability. Within the risk management perspective, for example, one might try to assess whether use of volatility forecasts improves the accuracy of calculated VaR measures at various horizons. One could also examine the usefulness of long-horizon volatility forecasts from other perspectives, including asset allocation, as in West et al. (1993) and derivatives pricing, as in Engle et al. (1993) and Christoffersen and Hahn (1999). In doing so, it will be important to use truly ex ante, out-of-sample, forecasts.

The second direction for future research involves addressing the obvious question that emerges from our work: If volatility dynamics are not important for long-horizon risk management, then what is important? It seems to us that all models miss the really big movements such as the U.S. crash of 1987, and ultimately the really big movements are the most important for risk management. This suggests the desirability of directly modeling the extreme tails of return densities, a task facilitated by recent advances in extreme-value theory surveyed by Embrechts et al. (1997) and applied to financial risk management by Danielsson and de Vries (1997). Preliminary ruminations along those lines appear in Diebold, Schuermann, and Stroughair (1998).

REFERENCES

Andersen, E. B., *The Statistical Analysis of Categorical Data,* 3rd edition (Berlin: Springer-Verlag, 1994).

Andersen, T. G., and T. Bollerslev, "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance* 52 (1997), 975–1005.

——— "Answering the Critics: Yes, ARCH Models do Provide Good Volatility Forecasts," *International Economic Review* 39 (1998), 885–905.

Andersen, T., and J. Lund, "Stochastic Volatility and Mean Drift in the Short Term Interest Rate Diffusion: Sources of Steepness, Level and Curvature in the Yield Curve," manuscript, Kellogg School, Northwestern University (1997).

Babbel, D. F., and A. M. Santomero, "Financial Risk Management by Insurers: An Analysis of the Process," *Journal of Risk and Insurance* 64 (1997), 231–270.

Bollerslev, T., R. Y. Chou, and K. F. Kroner, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52 (1992), 5–59.

Brennan, M., and E. Schwartz, "A Continuous-Time Approach to the Pricing of Bonds," *Journal of Banking and Finance* 3 (1979), 133–153.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, *The Econometrics of Financial Markets* (Princeton: Princeton University Press, 1997).

Chew, L., "Shock Treatment," *Risk* 7 (September 1994), 63–70.

Christoffersen, P. F. "Evaluating Interval Forecasts," *International Economic Review* 39 (1998), 841–862.

Christoffersen, P. F., and J. Hahn, "Nonparametric Testing of ARCH for Option Pricing," in Y. S. Abu-Mostafa, B. LeBaron, A. W. Lo, and A. S. Weigend (Eds.), *Proceedings of the Sixth International Conference on Computational Finance* (Cambridge, MA: MIT Press, 1999).

Cox, J., J. Ingersoll, and S. Ross, "A Theory of the Term Structure of Interest Rates," *Econometrica* 53 (1985), 385–408.

Danielsson, J., and C. G. de Vries, ''Extreme Returns, Tail Estimation, and Value-at-Risk,'' manuscript, Financial Markets Group, London School of Economics (1997).

David, F. N., ''A Power Function for Tests of Randomness in a Sequence of Alternatives,'' *Biometrika* 34 (1947), 335–339.

Diebold, F. X., ''Modeling the Persistence of Conditional Variances: Comment,'' *Econometric Reviews* 5 (1986), 51–56.

Diebold, F. X., A. Hickman, A. Inoue, and T. Schuermann, ''Scale Models,'' *Risk* 11 (1998), 104–107.

Diebold, F. X., T. Schuermann, and J. Stroughair, ''Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management,'' Wharton Financial Institutions Center, Working Paper 98-10 (January 1998). Forthcoming in A.-P. N. Refenes, J. D. Moody, and A. N. Burgess (Eds.), *Advances in Computational Finance* (Amsterdam: Kluwer Academic Publishers).

Drost, F. C., and T. E. Nijman, ''Temporal Aggregation of GARCH Processes,'' *Econometrica* 61 (1993), 909–927.

Embrechts, P., C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events* (New York: Springer-Verlag, 1997).

Engle, R. F., C.-H. Hong, A. Kane, and J. Noh, ''Arbitrage Valuation of Variance Forecasts with Simulated Options,'' in D. Chance and R. Tripp (Eds.), *Advances in Futures and Options Research* (Greenwich, CT: JIA Press, 1993).

Engle, R. F., D. M. Lilien, and R. P. Robins, ''Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model,'' *Econometrica* 55 (1987), 391–407.

Falloon, W., ''Growin' Up,'' *Risk* 12 (February 1999), 26–31.

Findley, D. F., ''On the Use of Multiple Models for Multi-Period Forecasting,'' *ASA Proceedings of the Business and Economic Statistic Section* (1983), 528–531.

Grant, E. L., and R. S. Leavenworth, *Statistical Quality Control,* 6th edition (New York: McGraw-Hill, 1988).

Hansen, B., ''Autoregressive Conditional Density Estimation,'' *International Economic Review* 35 (1994), 705–729.

Hamilton, J. D., *Time Series Analysis* (Princeton: Princeton University Press, 1994).

Hamilton, J. D., and R. Susmel, ''Autoregressive Conditional Heteroskedasticity and Changes in Regime,'' *Journal of Econometrics* 64 (1994), 307–333.

Hsieh, D. A., ''Implications of Nonlinear Dynamics for Financial Risk Management,'' *Journal of Financial and Quantitative Analysis* 28 (1993), 41–64.

Jacquier, E., N. G. Polson, and P. E. Rossi, ''Bayesian Analysis of Stochastic Volatility Models,'' *Journal of Business and Economic Statistics* 12 (1994), 371–388.

——— ''Stochastic Volatility: Univariate and Multivariate Extensions,'' manuscript, Finance Department, Boston College (1999).

Jorion, P., ''Predicting Volatility in the Foreign Exchange Market,'' *Journal of Finance* 50 (1995), 507–528.

Kupiec, P., and J. O'Brien, ''Internal Affairs,'' *Risk* 8 (May 1995), 43–47.

Lamoureux, C. G., and W. D. Lastrapes, ''Persistence in Variance, Structural Change and the GARCH Model,'' *Journal of Business and Economic Statistics* 8 (1990), 225–234.

Lehmann, E. L., *Testing Statistical Hypotheses,* 2nd edition (New York: John Wiley, 1986).

Locke, J., ''RiskMetrics Launches New System for Corporate Users,'' *Risk* 12 (May 1999), 54–55.

McNew, L., ''So Near, So VaR,'' *Risk* 9 (October 1996), 54–56.

Santomero, A. M., ''Financial Risk Management: The Whys and Hows,'' *Financial Markets, Institutions and Instruments* 4 (1995), 1–14.

——— ''Commercial Bank Risk Management: An Analysis of the Process,'' *Journal of Financial Services Research* 12 (1997), 83–115.

Shorrocks, A. F., ''The Measurement of Mobility,'' *Econometrica* 46 (1978), 1013–1024.

Smithson, C., and L. Minton, ''Value-at-Risk,'' *Risk* 9 (February 1996), 38–39.

Sommers, P. M., and J. Conlisk, ''Eigenvalue Immobility Measures for Markov Chains,'' *Journal of Mathematical Sociology* 6 (1979), 253–276.

Tiao, G. C., and R. S. Tsay, ''Some Advances in Non-Linear and Adaptive Modeling in Time Series,'' *Journal of Forecasting* 13 (1994), 109–131.

Weiss, A. A., ''Multi-Step Estimation and Forecasting in Dynamic Models,'' *Journal of Econometrics* 48 (1991), 135–149.

West, K. D., H. J. Edison, and D. Cho, ''A Utility-Based Comparison of Some Models of Exchange Rate Volatility,'' *Journal of International Economics* 35 (1993), 23–45.

West, K., and D. Cho, ''The Predictive Ability of Several Models of Exchange Rate Volatility,'' *Journal of Econometrics* 69 (1995), 367–91.

Wolfowitz, J., ''On the Theory of Runs with some Applications to Quality Control,'' *Annals of Mathematical Statistics* 14 (1943), 280–288.