# How Relevant is Volatility Forecasting for Financial Risk Management?

Peter F. Christoffersen
Faculty of Management
McGill University

Francis X. Diebold
Stern School, NYU
University of Pennsylvania
and NBER

This Draft/Print: July 29, 1999

Abstract: It depends. If volatility fluctuates in a forecastable way, then 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.

Acknowledgments: 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. 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.[4] Interest has similarly escalated on the regulatory side, as governments around the world seek to impose risk-based capital adequacy standards.[5] 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

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

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

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 (e.g., public/regulatory vs. private/for-profit), position in the firm (e.g., trading desk vs. CFO), asset class (e.g., equity vs. fixed income), and industry (e.g., banking vs. 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.[6] 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

[6] See, for example, Bollerslev, Chou and Kroner (1992).

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 (e.g., converting a 1-day return standard deviation to a 30-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 with horizon, as shown by Diebold et al. (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 1-day volatility dynamics.[7] 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

[7] See Findley (1983), Weiss (1991), and Tiao and Tsay (1993).

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 2, we describe in detail our framework for model-free evaluation of volatility forecastability, and then in section 3 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. 10-year Treasury bond, at horizons ranging from one through twenty trading days. In section 4 we offer concluding remarks and directions for future research.

## 2. 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.

### 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 1-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

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

for t = 1, 2, ..., 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* $\Omega_{t-1}$ 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 $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, ..., I_1\}$, then correct conditional coverage is equivalent to $\{I_t\} \overset{iid}{\sim} \text{Bernoulli}(p)$, which can readily be tested.

Having given some background on interval forecast evaluation, now let us proceed to our ultimate goal, 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].[8] [9] 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 $\{y_{t-1}, y_{t-2}, \ldots , y_1\}$, then a fixed-width confidence interval could not be correctly conditionally calibrated, because it fails to widen when the conditional variance rises and 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 iid.[10]

Assessing Independence of the Hit Sequence: A Runs Test

---

[8] 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.

[9] 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. Use of a test based on a one-sided interval would reduce power substantially.

[10] It is interesting to note that tests based on the iid 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.

We have seen that non-forecastability of volatility corresponds to an iid 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 (e.g., Grant and Leavenworth, 1988), and it can be viewed as an application of categorical data analysis (e.g., Andersen, 1994).

Define a run as a string of consecutive zeros or ones in the hit sequence.[11] 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 $\min\{n_0,n_1\}>0$)

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

where

$$f_{r=2s} = 2\binom{n_0-1}{s-1}\binom{n_1-1}{s-1} \text{ and } f_{r=2s+1} = \binom{n_0-1}{s}\binom{n_1-1}{s-1} \binom{n_0-1}{s-1}\binom{n_1-1}{s} = \frac{f_{2s}(T-2s)}{2s}.$$

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

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.[12]

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 1-day returns to h-day returns, h = 2, 3, ..., 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 let's us 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 JP Morgan's RiskMetrics. We use highly persistent GARCH processes ($\alpha+\beta=.99$, 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 20-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 a half standard deviations at each horizon. Nevertheless, we will set c to two standard deviations for most of

[12] See Lehmann (1986) for details.

this paper, as it yields an unconditional coverage of greater relevance to risk managers, with only a slight reduction in power.[13]

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

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

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 $\pi_{11}$ will be much larger than $\pi_{01}$, so S will be large.[14]

S has an alternative and intuitive motivation: it is the first-order serial correlation

---

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

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

coefficient of the hit sequence. To see this, we note that[15]

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

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

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

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

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

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 (non-trivial) eigenroot.[16]

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 $\{I_1, ..., I_T\}$, the likelihood function is

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

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

immediately[17]

$$L(\pi_{01}, \pi_{11}; I_1, I_2, ..., I_T) \; = \; (1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}},$$

where $n_{ij}$ is the number of observations with value I followed by j. The maximum likelihood estimators of $\pi_{01}$ and $\pi_{11}$ are therefore $\hat{\pi}_{01} = \frac{n_{01}}{n_{00}+n_{01}}$ and $\hat{\pi}_{11} = \frac{n_{11}}{n_{10}+n_{11}}$. 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 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:

(a) De-mean the returns series

(b) Compute the hit sequence relative to the constant ±c interval, and then compute the estimate of p, $\hat{p}$, and the estimate of S, $\hat{S}$.

© Use $\hat{p}$ and the relevant sample size T to:

(c1) generate m = 1, ..., M samples of iid Bernoulli($\hat{p}$) pseudo-data

(c2) compute $\hat{S}_m$

(c3) compute the 95 percent confidence interval for $\hat{S}_m$ and plot it together with $\hat{S}$

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

computed in (b).

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 (e.g., daily), regardless of the return horizon.

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

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

and define

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

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

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

Notice that $I_t$ can be conveniently defined as

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

where 1(•) is the indicator function. We therefore have that

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

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

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

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

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.[18] 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.

**3. 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 non-overlapping h-day returns, h = 1, 2, 3, ..., 20.[19]

Because the unconditional volatility of all asset returns rises with the aggregation level, it

---

[18] To implement the testing procedure, we could attempt to find a functional form for $F(c \mid \Omega_{t-1})$, but we prefer a less parametric approach in which we directly include elements of the information set, such as $I_{t-1}$ or $y^2_{t-1}$, on the right-hand-side. Hence $\beta_1$ should be interpreted as a vector.

[19] Use of non-overlapping returns eliminates the need to account for the dependence induced by overlapping observations.

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 percent, 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 percent.

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 6350 daily observations for each return series.[20]

Let us first discuss the runs tests. In Figure 3 we show 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 significantly forecastable, and conversely for horizons greater than ten days. Figure 4 reveals identical patterns for exchange

[20] The equity and foreign exchange data are from Datastream International. Equity prices are official local closing prices provided by the local exchanges; 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; they are averages of closing New York bid and ask quotes.

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 1.5 standard deviations. We therefore now calculate the p-values of the runs test using the narrower intervals and plot the results in Figures 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 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 Figures 7 and 8.[21] 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 one

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

reaches 10-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: (1) 1-5 lags of the hit sequence from one-day returns, (2) 1-5 lags of the squared daily returns, (3) 1-5 lags of the daily RiskMetrics filtered volatility. p-values from F-tests of the null hypothesis that the high-frequency information is irrelevant – i.e., all slopes are zero in the limited dependent variable regression – are plotted in Figures 9 and 10. 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 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 Figures 11 and 12. We plot the p-values of the F-test corresponding to the null hypothesis that 1-5, 1-10, and 1-15 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.

Bond Markets

We report results for bonds separately for three reasons; the first two are linked to a priori concerns, and the third concerns the different nature of the results. First, historical bond market data typically contain only the annual yield, 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. 10-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

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

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

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

which can also be written as[22]

[22] See, for example, Campbell, Lo and MacKinlay (1997, p. 403).

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

Assume that the coupon rate is close to the yield, $C \approx Y_{cnt}$, in which case the bond will be priced near par, $P_{cnt} \approx 1$, resulting in the approximate duration[23]

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

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

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

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

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

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 10-year U.S. Treasury bond. As before, the daily

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

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 15-20 trading days.[24]

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 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, Lilien, and Robins (1987) and Andersen and Lund (1997).[25]

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, Ingersoll and Ross (1985). Those models imply that a simple rescaling of bond returns by a function of the short-term yield will constant-volatility returns. In the spirit of this argument, we rescale the approximate bond returns above by a power of the bond yields,

---

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

[25] See also the survey by Bollerslev, Chou and Kroner (1992).

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

Following Campbell, Lo and McKinlay (1997, p. 450), we set $\gamma$ 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.

## 4. Concluding Remarks and Directions for Future Research

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, then 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 5-day horizon.[26] Our results

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

are also in agreement with those of Jacquier, Polson and Rossi (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, which 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 (e.g., Andersen and Bollerslev, 1997). But that literature tends to work with very high-frequency data -- typically 5-minute returns -- and although long memory in 5-minute returns may well indicate that volatility is highly forecastable many steps into the future, perhaps 100 steps or even 1000 steps, it does not necessarily indicate forecastability beyond ten or twenty days. 1000 5-minute steps, for example, are just more than three days; even 5000 5-minute steps are just more than 17 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 (e.g., Andersen and Bollerslev, 1998). That literature, however, focuses on 1-day-ahead volatility forecasts, and certainly we agree that *short*-horizon volatility is highly forecastable. Our analysis, in contrast, focuses on longer-horizon volatility.

## 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, Edison and Cho (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, Klüppelberg and Mikosch (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. (1994), *The Statistical Analysis of Categorical Data*, Third Edition. Berlin: Springer Verlag.

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

Andersen, T. and Bollerslev, T. (1998), "Answering the Critics: Yes, ARCH Models do Provide Good Volatility Forecasts," *International Economic Review*, 39, 885-905.

Andersen, T. and Lund, J. (1997), "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.

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

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

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

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

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

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

Christoffersen, P.F. and Hahn, J. (1999), "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, Mass.: MIT Press.

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

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

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

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

Diebold, F.X., Hickman, A., Inoue, A. and Schuermann, T. (1998), “Scale Models,” *Risk*, 11, 104-107.

Diebold, F.X., Schuermann, T. and Stroughair, J. (1998), “Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management,” Wharton Financial Institutions Center, Working Paper 98-10, http://fic.wharton.upenn.edu/fic/. 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 Nijman, T.E. (1993), “Temporal Aggregation of GARCH Processes,” *Econometrica*, 61, 909-927.

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

Engle, R. F. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, 50, 987-1007.

Engle, R.F., Hong, C.-H., Kane, A. and Noh, J. (1993), "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.

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

Falloon, W. (1999), “Growin’ Up,” *Risk*, 12, February, 26-31.

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

Froot, K.A., Scharfstein, D.S. and Stein, J.C. (1993), “Risk Management: Coordinating Corporate Investment and Financing Policies,” *Journal of Finance*, 48, 1629-1658.

Froot, K.A., Scharfstein, D.S. and Stein, J.C. (1994), “A Framework for Risk Management,” *Harvard Business Review*, November-December, 91-102.

Grant, E.L. and Leavenworth, R.S. (1988), *Statistical Quality Control*, Sixth Edition. New York:

McGraw-Hill.

Hansen, B. (1994), “Autoregressive Conditional Density Estimation,” *International Economic Review*, 35, 705-729.

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

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

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

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

Jacquier, E., Polson, N.G., and Rossi, P.E. (1999), ”Stochastic Volatility: Univariate and Multivariate Extensions,” Manuscript, Finance Department, Boston College.

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

Kupiec, P. and O’Brien, J. (1995), “Internal Affairs,” *Risk*, 8, 43-47.

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

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, Second Edition. New York: John Wiley.

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

McNew, L. (1996), “So Near, So VaR,” *Risk*, 9, October, 54-56.

Modigliani, F. and Miller, M.H. (1958), “The Cost of Capital, Corporation Finance and the Theory of Investment,” *American Economic Review*, 48, 261-297.

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

Santomero, A.M. (1997), “Commercial Bank Risk Management: An Analysis of the Process,” *Journal of Financial Services Research*, 12, 83-115.

Shorrocks, A.F. (1978), “The Measurement of Mobility,” *Econometrica*, 46, 1013-1024.

Smithson, C. and Minton, L. (1996), “Value-at-Risk,” *Risk*, 9, February, 25-27.

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

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

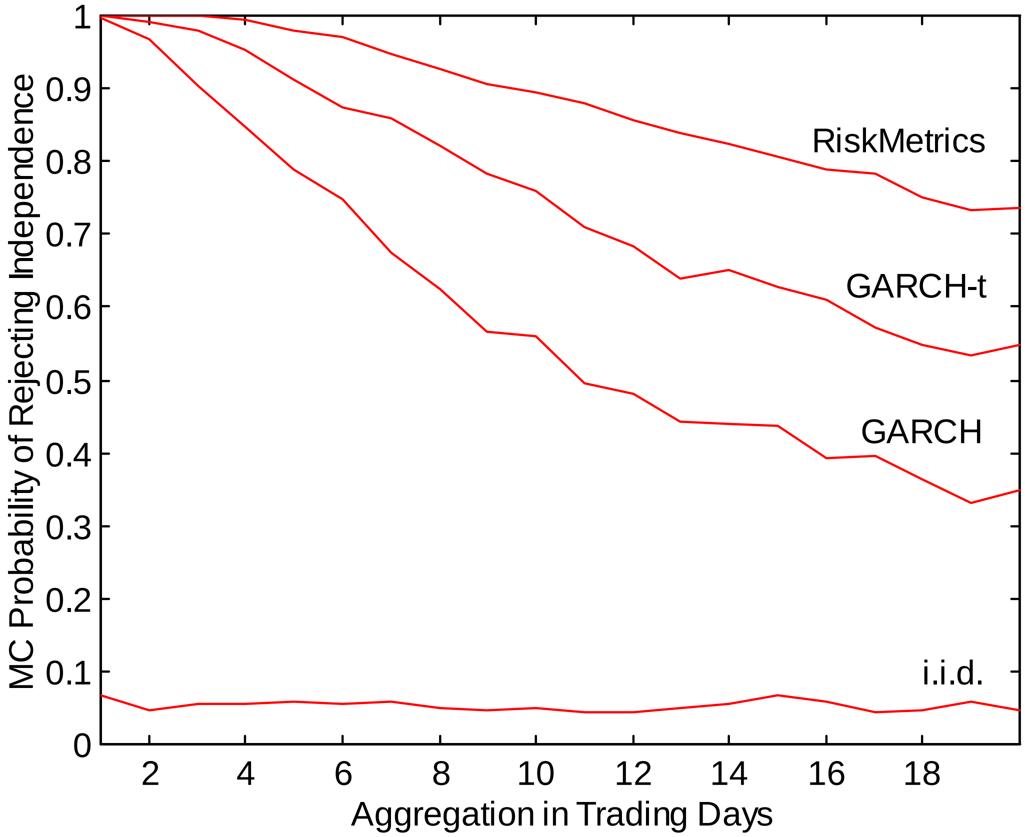
Weiss, A.A. (1991), “Multi-Step Estimation and Forecasting in Dynamic Models,” *Journal of Econometrics,* 48, 135-149.

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

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

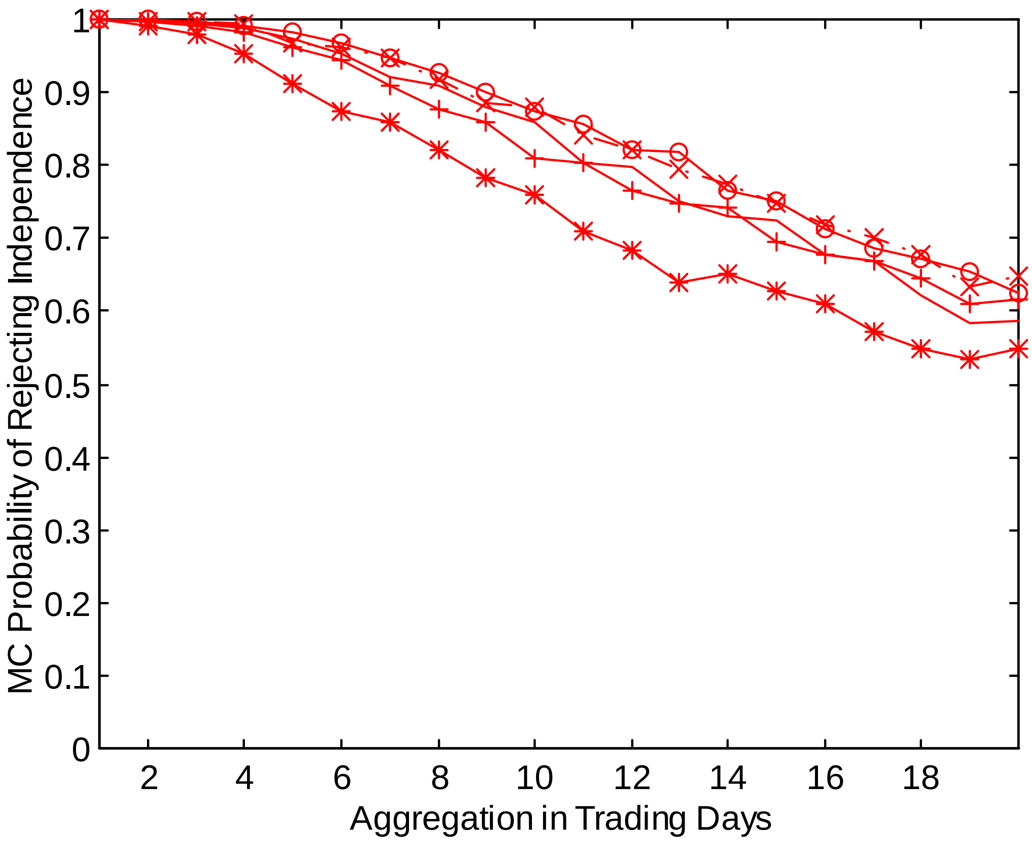
Wolfowitz, J. (1943), “On the Theory of Runs with some Applications to Quality Control,” *Annals of Mathematical Statistics*, 14, 280-288.

**Figure 1**
**Size and Power of 5% Run Tests of Independence**
**for ± 2 Standard Deviation Intervals under Various DGPs**



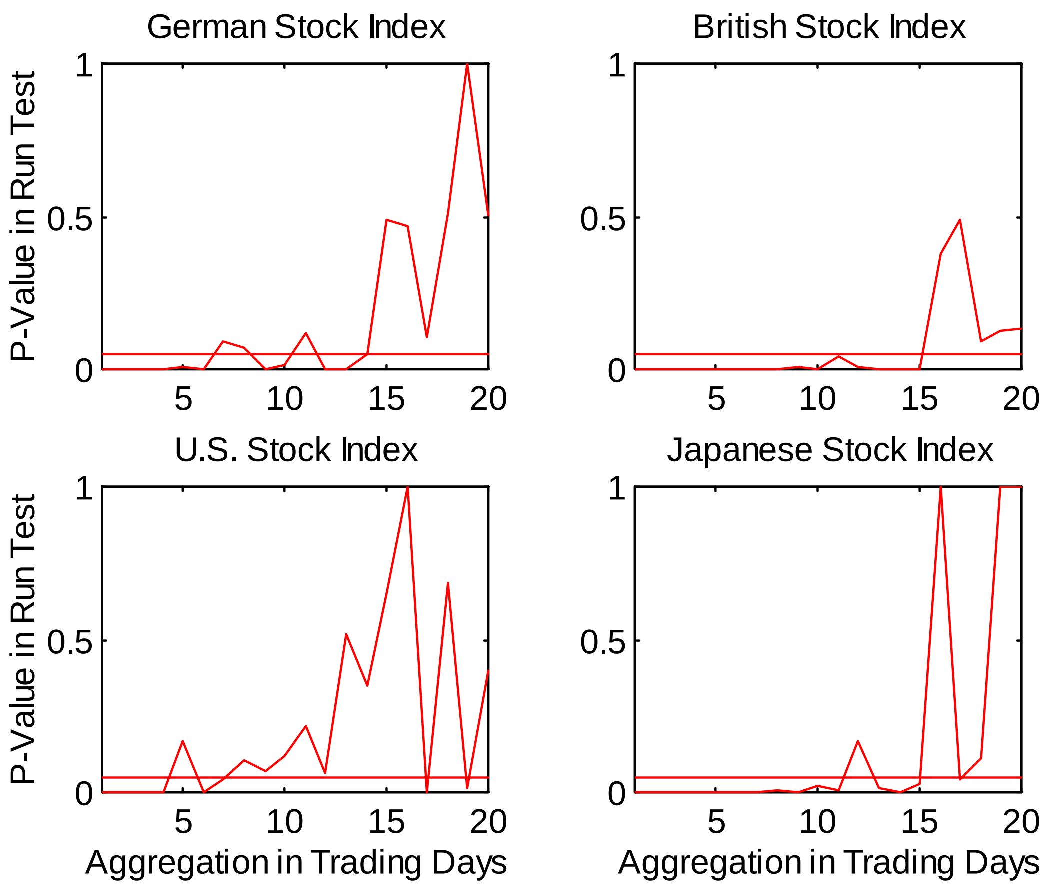

Notes to figure: The graph shows Monte Carlo estimates of the probability of rejecting the null hypothesis of independence in a runs test with significance level set to 5 percent. The width of the interval definining the hit sequence is $\pm 2\hat{\sigma}$. The data generating processes for the daily data are the RiskMetrics or IGARCH process with $\beta$=.94, the GARCH(1,1)-t(d) process with $\alpha$=.06, $\beta$=.93, d=5, the Gaussian GARCH(1,1) with $\alpha$=.06, $\beta$=.93, and independent Gaussian innovations respectively. All processes have 6,350 daily observations. The number of Monte Carlo replications is 1,000. See text for details.

**Figure 2**
**Power of 5% Run Tests of Independence**
**for Various Interval Forecasts Using GARCH(1,1)-t(d) Data**



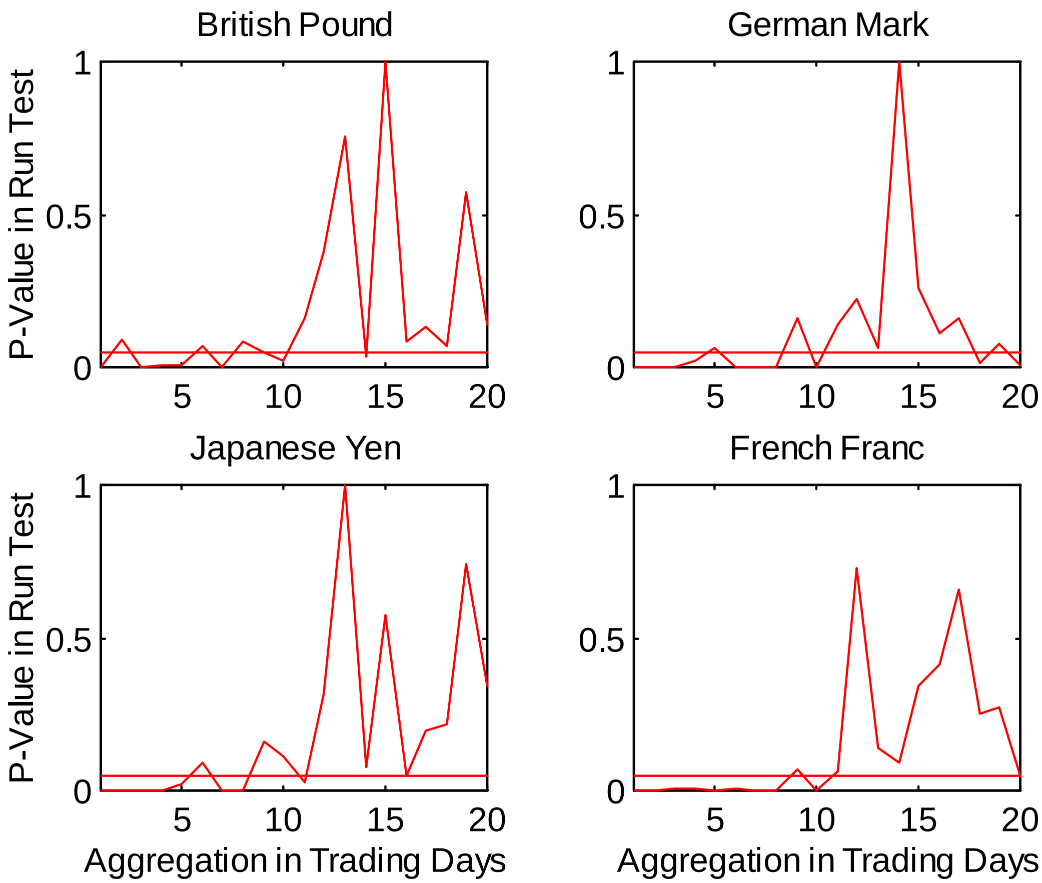

Notes to figure: The graph shows Monte Carlo estimates of the probability of rejecting the null hypothesis of independence in a runs test with significance level set to 5 percent. The width of the interval forecast definining the hit sequence is $\pm m\hat{\sigma}$, where m=1 (——), m=1.25 (ooo), m=1.5 (xxx), m=1.75 (+++), and m=2 (***). The data generating process for the daily data is the GARCH(1,1)-t(d) process with $\alpha$=.06, $\beta$=.93, and d=5, with 6,350 daily observations. The number of Monte Carlo replications is 1,000. See text for details.

**Figure 3**
**p-Values of Runs Tests**
**Four Equity Indexes**
**± 2 Standard Deviation Interval Forecasts**



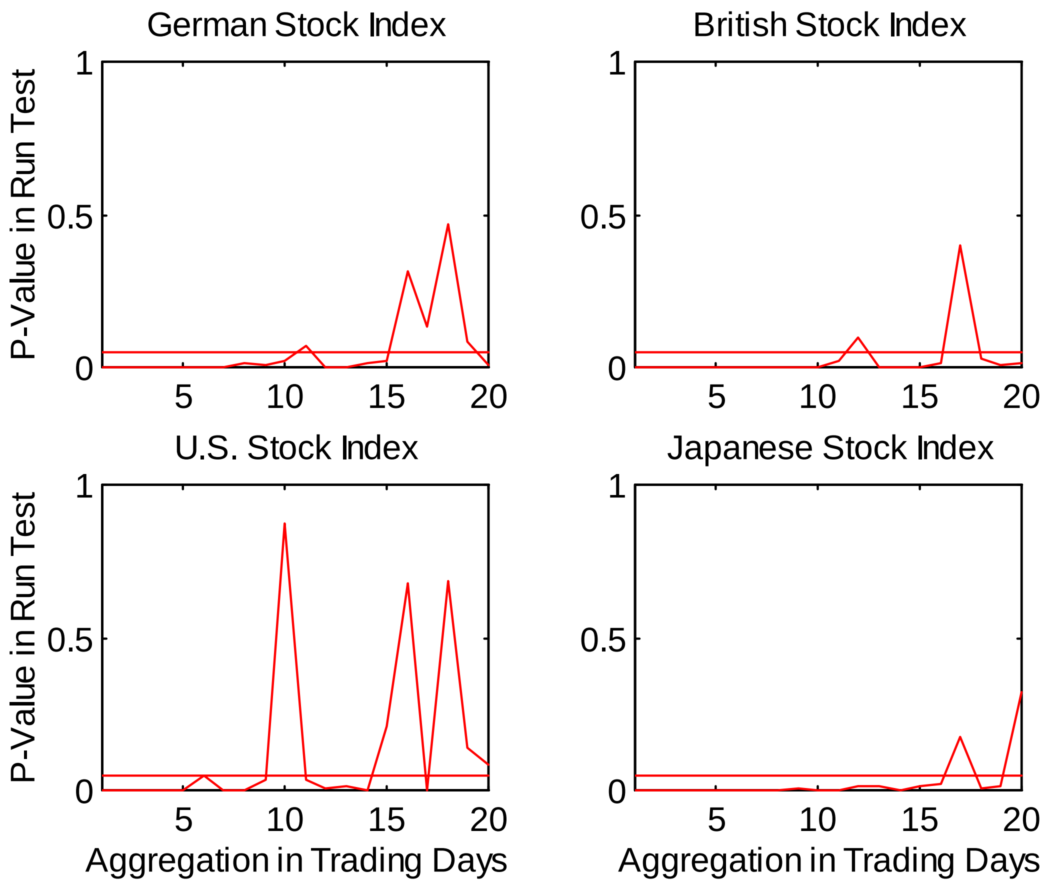

Notes to figure: For each series and horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast. The horizontal line is at 5 percent. See text for details.

**Figure 4**
**p-Values of Runs Tests**
**Four Dollar Exchange Rates**
**± 2 Standard Deviation Interval Forecasts**



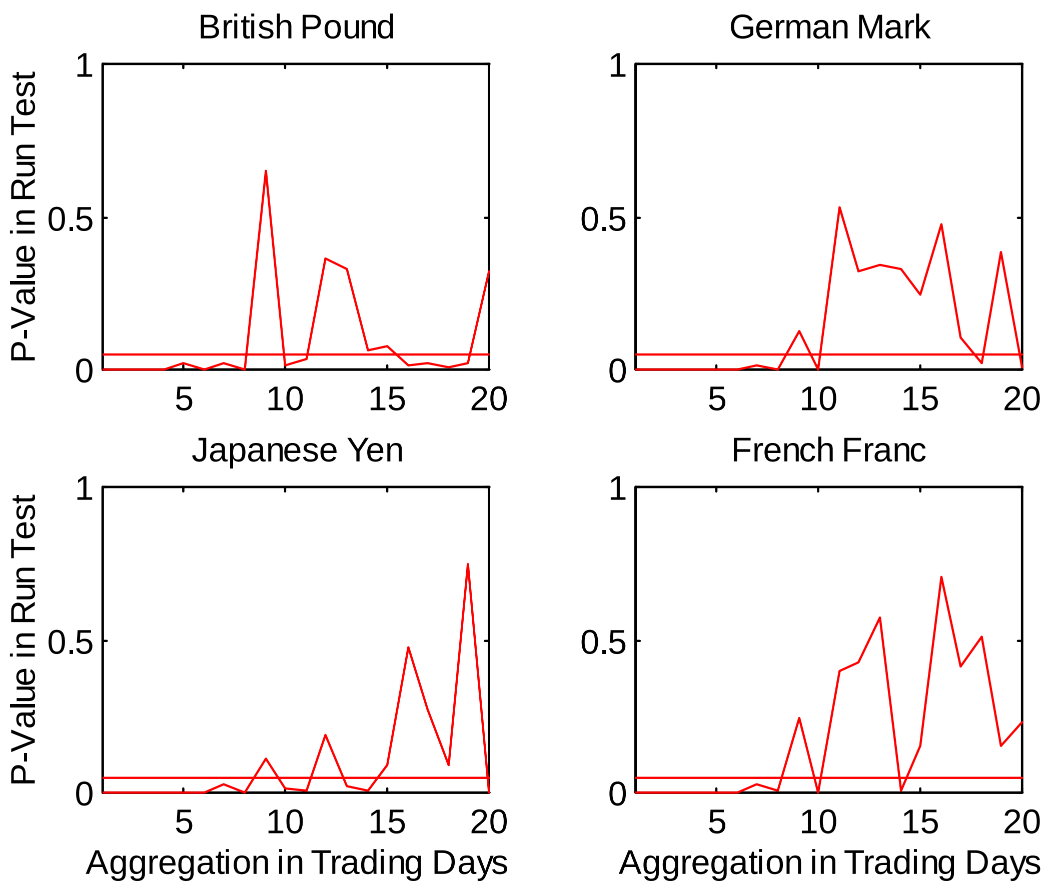

Notes to figure: For each series and horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast. The horizontal line is at 5 percent. See text for details.

**Figure 5**
**p-Values of Runs Tests**
**Four Equity Indexes**
**± 2 Standard Deviation Interval Forecasts**



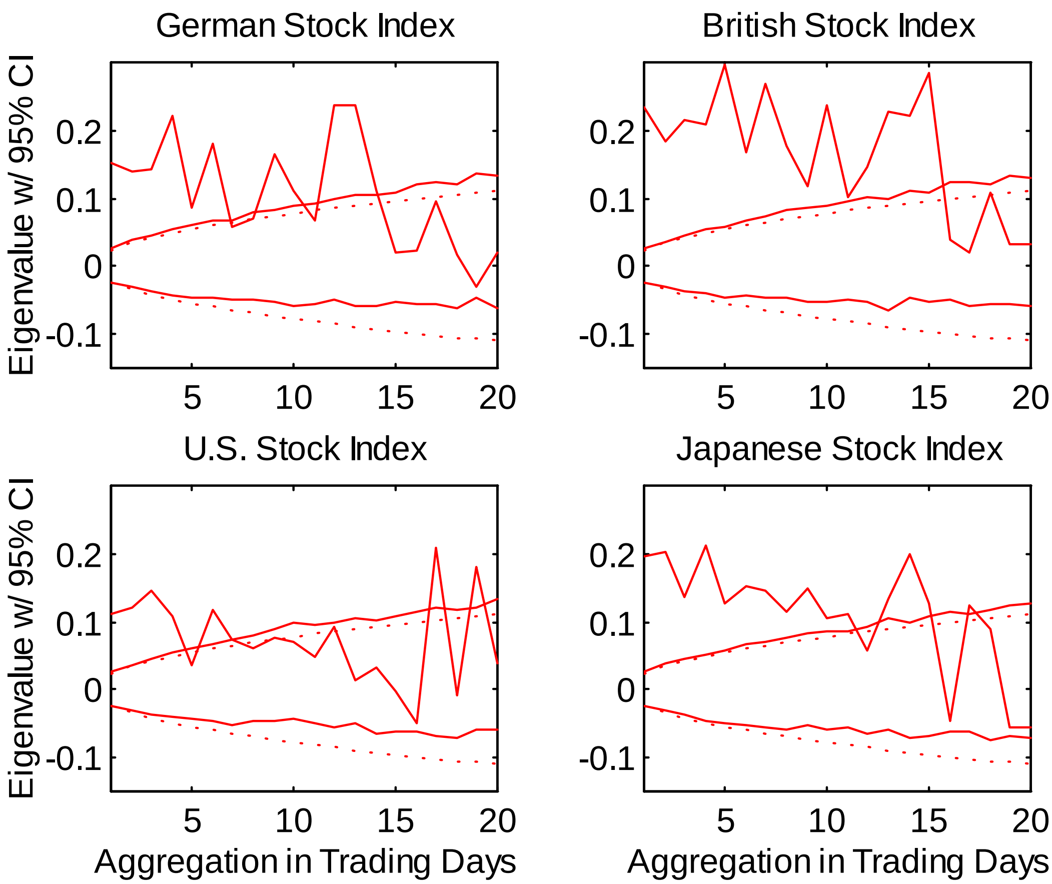

Notes to figure: For each series and horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 1.5\hat{\sigma}$ interval forecast. The horizontal line is at 5 percent. See text for details.

**Figure 6**
**p-Values of Runs Tests**
**Four Dollar Exchange Rates**
**± 1.5 Standard Deviation Interval Forecasts**



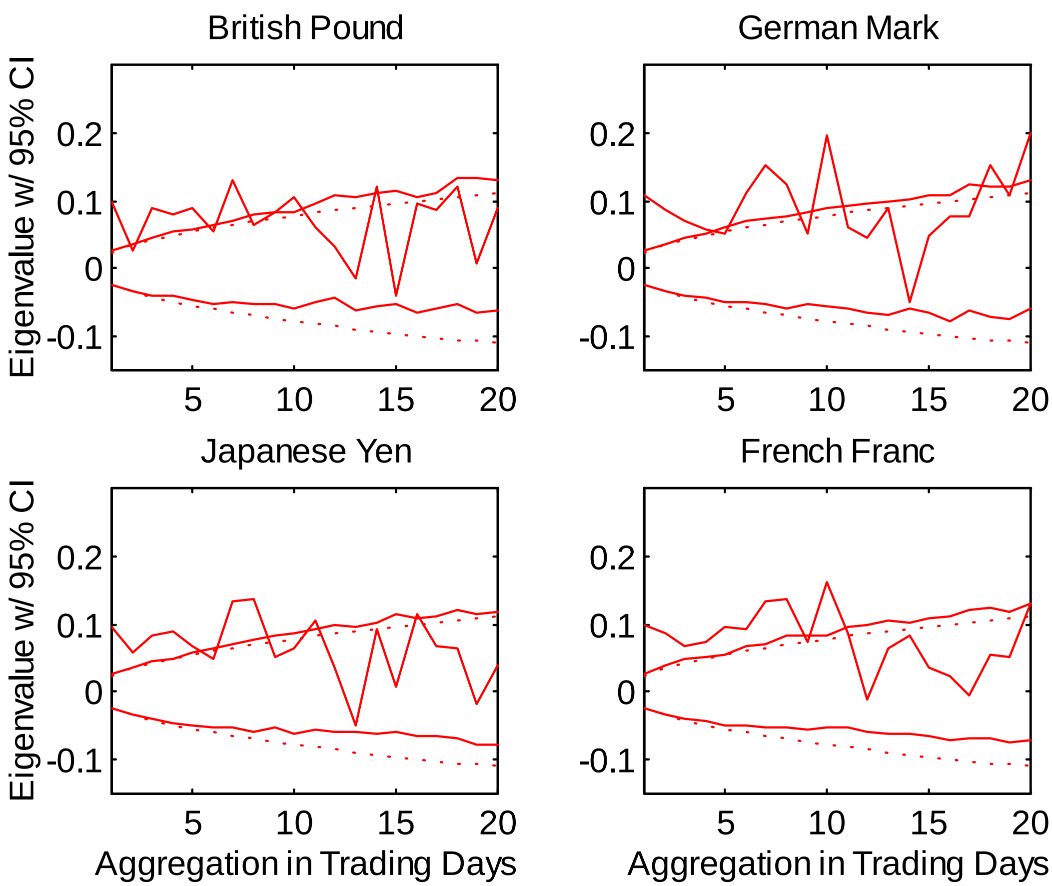

Notes to figure: For each series and horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant $\pm 1.5\hat{\sigma}$ interval forecast. The horizontal line is at 5 percent. See text for details.

**Figure 7**
**Markov Transition Matrix Eigenvalues**
**Four Equity Indices**



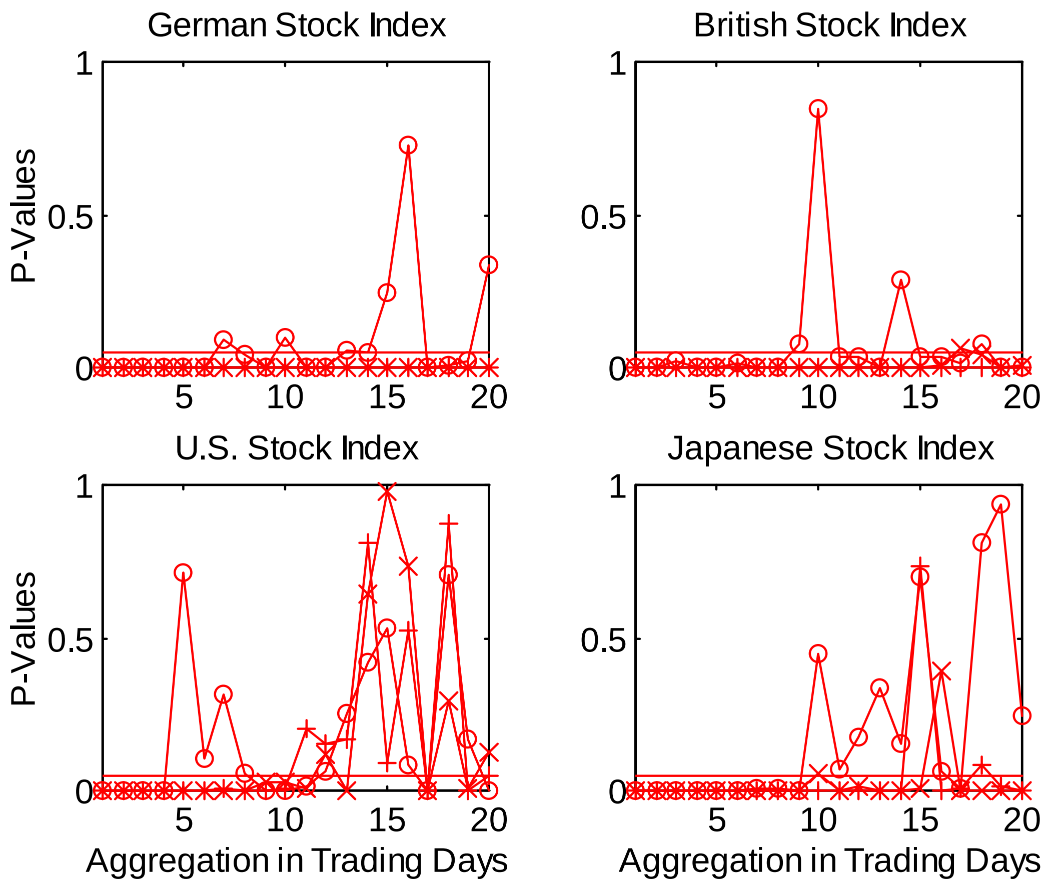

Notes to figure: For each series and each horizon we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast, along with the finite-sample 95 percent confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4000 simulations (solid lines), and asymptotic confidence intervals (dashed lines). See text for details.

**Figure 8**
**Markov Transition Matrix Eigenvalues**
**Four Dollar Exchange Rates**



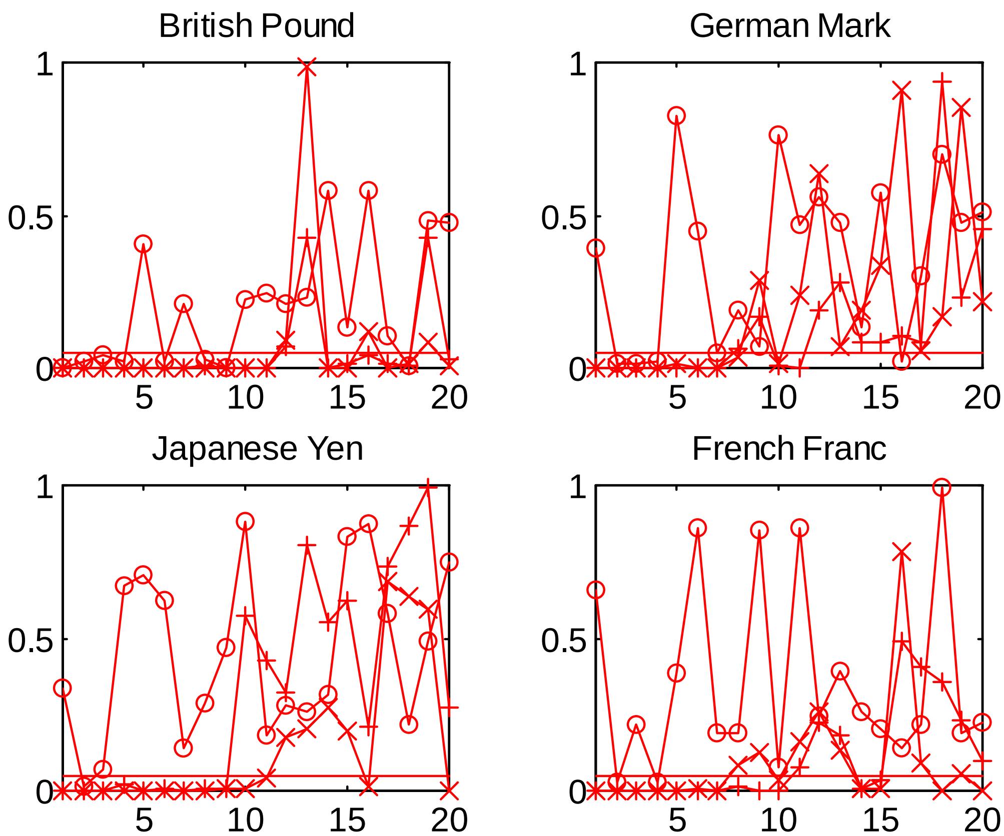

Notes to figure: For each series and each horizon we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant $\pm 2\hat{\sigma}$ interval forecast, along with the finite-sample 95 percent confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4000 simulations (solid lines), and asymptotic confidence intervals (dashed lines). See text for details.

**Figure 9**
**p-Values from F-Test of High-Frequency Information**
**Four Equity Indexes**



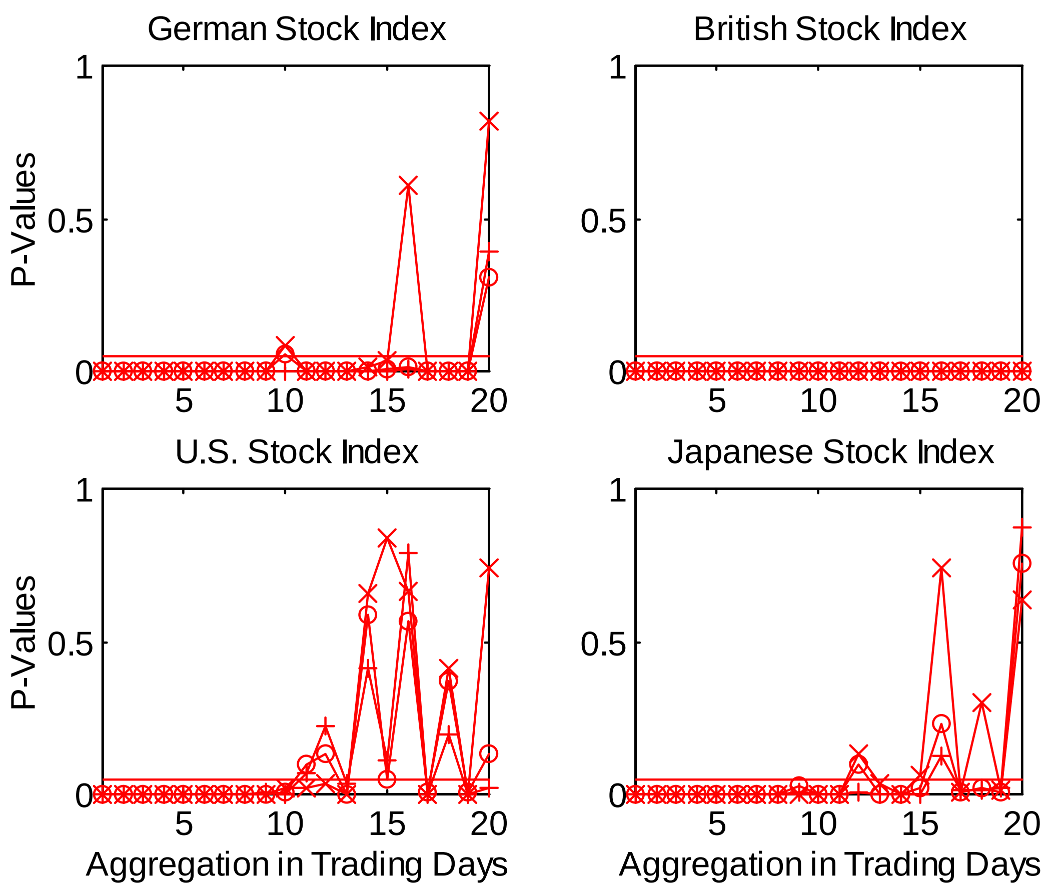

Notes to figure: For each series and each horizon we plot the p-values of F-tests associated with regressions on the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The high frequency information sets are: 1-5 lags of daily hits (x x x), 1-5 lags of daily squared returns ( o o o), and 1-5 lags of daily RiskMetrics volatility (+ + +). The horizontal line denotes the 5% critical value. See text for details.

**Figure 10**
**p-Values from F-Tests of High-Frequency Information**
**Four Dollar Exchange Rates**



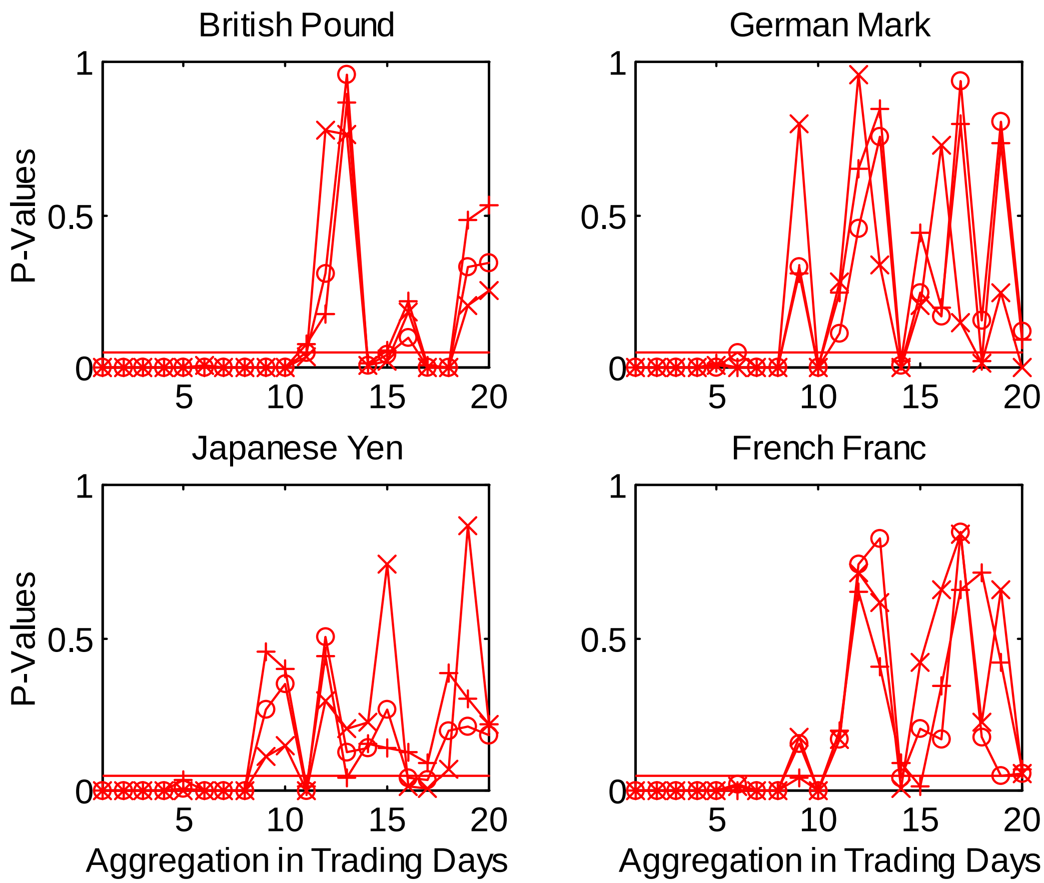

Notes to figure: For each series and each horizon we plot the p-values of F-tests associated with regressions on the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The high frequency information sets are: 1-5 lags of daily hits (x x x), 1-5 lags of daily squared returns ( o o o), and 1-5 lags of daily RiskMetrics volatility (+ + +). The horizontal line denotes the 5% critical value. See text for details.

**Figure 11**
**p-Values from F-Tests of Higher-Order Dependence**
**Four Equity Indexes**



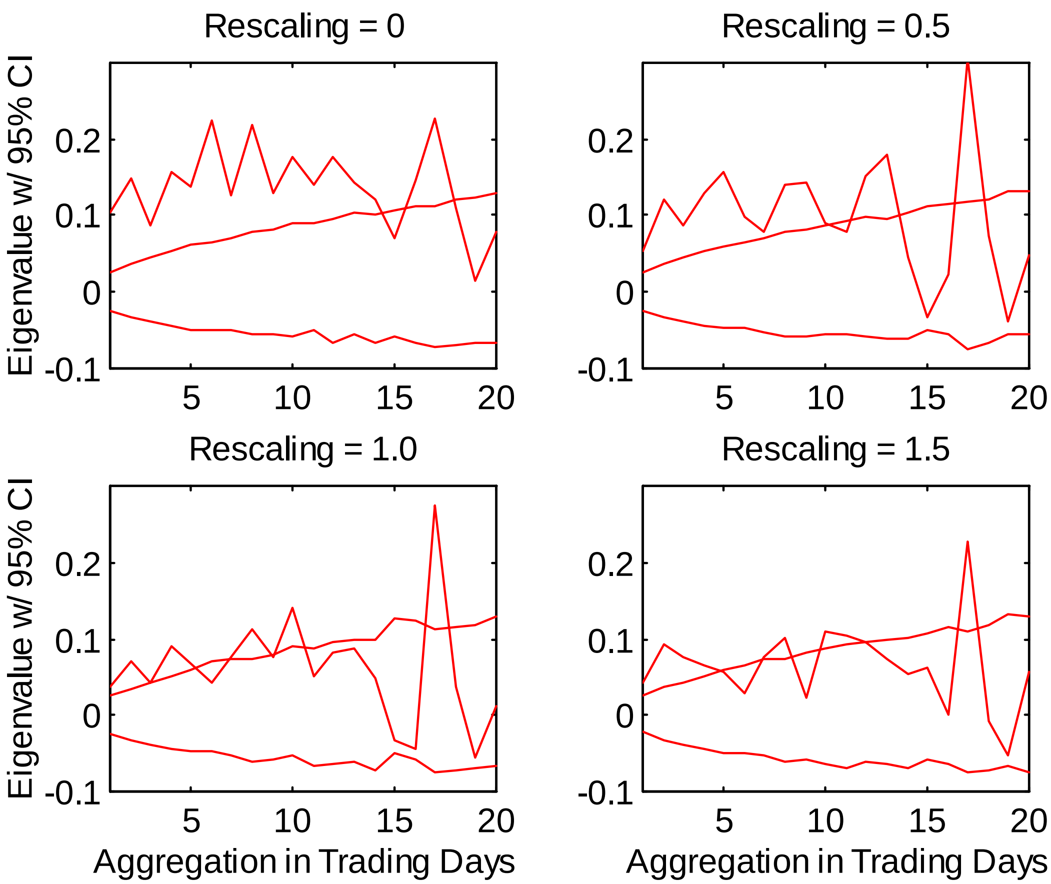

Notes to figure: For each series and each horizon we plot the values of F-tests associated with the hit sequences corresponding to $\pm 2\hat{\sigma}$ interval forecasts. The higher-order dependencies in the hit sequence tested are: 1-5 lags (x x x), 1-10 lags (o o o), and 1-15 lags (+ + +). The horizontal line is drawn at 5 percent. See text for details.

**Figure 12**
**p-Values from F-Tests of Higher-Order Dependence**
**Four Dollar Exchange Rates**




Notes to figure: For each series and each horizon we plot the 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: 1-5 lags (x x x), 1-10 lags (o o o), and 1-15 lags (+ + +). The horizontal line is drawn at 5 percent. See text for details.

**Figure 13**
**Markov Transition Matrix Eigenvalues**
**U.S. 10-Year Government Bond Returns**
**Rescaled by Various Powers of the Bond Yield**




Notes to figure: 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 percent confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4000 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.