How Relevant is Volatility Forecasting for Financial Risk Management?

Peter F. Christoffersen, Francis X. Diebold

October 17, 1997

How Relevant is Volatility Forecasting for Financial Risk Management?

Peter F. Christoffersen

Research Department
International Monetary Fund

Francis X. Diebold

Department of Economics
University of Pennsylvania and NBER

This Draft/Print: October 17, 1997

Send correspondence to Diebold at fdiebold@mail.sas.upenn.edu.

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. Existing assessments are plagued by the fact that they are *joint* assessments of volatility forecastability and an assumed model, and the results vary not only with the horizon, but also with the model. To address this problem, we develop a model-free procedure for measuring 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 very short horizons relevant for, say, trading desk management, may not be important for risk management more generally.

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. [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* chronicling the development.

At first pass, private-sector interest in risk management seems curious. Modigliani and Miller (1958) taught us long ago that the value of a firm is independent of its risk structure; firms should simply maximize expected profits, regardless of the risk entailed; holders of securities can achieve risk transfers via appropriate portfolio allocations. It is clear, however, that the strict conditions required for the Modigliani-Miller theorem are routinely violated in practice. [3] In particular, capital market imperfections, such as taxes and costs of financial distress, cause the theorem to fail and create a role for risk management. Ultimately, firms optimize nonlinear objective functions that are much more complicated than simple expected profits.

The rapid expansion in risk management interest and capability is driven by several

---

[1] See Santomero (1995, 1997), Babbel and Santomero (1996), and Froot and O'Connell (1997).

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

[3] Such imperfections motivate much of the recent literature. See, for example, Froot, Scharfstein, and Stein (1993, 1994) and Oldfield and Santomero (1997).

factors. One obvious factor is the growth in financial derivative markets and products, and the exciting capabilities for risk management that they provide. A second key factor, very much relevant for this paper, is the revolution in modeling and forecasting volatility that began in academics nearly two decades ago (Engle, 1982). As that literature has matured, and as our abilities in computation and simulation have advanced, it has fueled the development of powerful risk-management methods and software. The key insight is that if volatility fluctuates in a forecastable way, then good volatility forecasts can improve financial risk management, which at its core boils down to forecasting the risk associated with holding potentially complicated nonlinear portfolios at various horizons.

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 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."

The upshot, of course, is that there is no one "relevant" horizon. The relevant horizon

will likely vary by position in the firm (e.g., trading desk vs. CFO), motivation (e.g., private vs. regulatory), asset class (e.g., equity vs. fixed income), and industry (e.g., banking vs. insurance); thought must be given to the relevant horizon on an application-by-application basis. But one thing is clear: 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.[4] 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 ARCH, GARCH, stochastic volatility and related models, 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 (e.g., converting 1-day volatility to 30-day volatility by multiplying by 30). Unfortunately, except

---

[4] See, for example, Bollerslev, Chou and Kroner (1992) and Andersen and Bollerslev (1997a).

3

under restrictive and routinely-violated conditions, scaling is misleading and tends to produce spurious magnification of volatility fluctuations with horizon, as shown by Diebold, Inoue, Hickman and Schuermann (1997).

A more appropriate temporal aggregation strategy is to fit a model to the high-frequency data and, conditional upon the truth of the fitted model, use it to infer the properties of the low-frequency data. Drost and Nijman (1993), for example, provide temporal aggregation formulae for the 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 dynamics. [5] 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 portfolio 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 parametric volatility models tend to produce different conclusions, 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 develop methods for doing so, and we use them to assess volatility forecastability in equity, foreign exchange, and bond markets, with surprising results. We proceed as follows. In section 2, we

---

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

4

first sketch the intuition of model-free evaluation of volatility forecastability, and then we

propose tests of correct conditional calibration and measures of the strength of deviations

from correct conditional calibration. In section 3, we use our methods to assess the

forecastability of return volatility for 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 the first subsection, we sketch the intuition and give a precise statement of our

methods. In particular, we argue that recently-developed tests of conditional calibration of

interval forecasts can be used to provide model-free assessments of volatility forecastability.

Then, in the second subsection, we provide a detailed discussion of a formal test of volatility

forecastability. Finally, in the third subsection, we propose a natural and complementary

measure of the strength of forecastability.

<u>Model-Free Assessment of Volatility Forecastability</u>

Shortly we will propose model-free methods for assessing volatility forecastability in

risk management contexts. The first step is to think about the evaluation of interval forecasts,

the adequacy of which is crucially dependent on their ability to capture volatility dynamics.

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, $\left\{\left(L_{t|t-1}(p),\ U_{t|t-1}(p)\right)\right\}_{t=1}^T$, 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* 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 $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, ..., I_1\}$, 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]. [6] The key insight is that although the interval [-c, c] is

---

[6] 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.

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, and 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. Moreover, as we have seen, correct conditional calibration corresponds to an iid

hit sequence.

Assessing Independence of the Hit Sequence: A Runs Test

We have seen that non-forecastability of volatility corresponds to an iid hit sequence;

we now describe a convenient and powerful model-free procedure for testing independence of

the hit sequence. Define a run as a string of consecutive zeros or ones in the hit sequence. [7]

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

---

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

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

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 $\min\{n_0, n_1\} > 0$)

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

where

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

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. [8] Moreover, the runs test is uniformly most powerful against a first-order Markov alternative. [9]

Persistence Measures: Markov Transition Matrix Eigenvalues and First-Order Correlations

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

---

[8] The runs test of randomness of a binary variable traces at least to David (1947).

[9] See Lehmann (1986) for details.

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

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 persistence 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. [10]

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 [11]

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

_____

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

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

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

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

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

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

Thus, just as in the familiar AR(1) case we know that 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.

3. Volatility Forecastability in Financial Asset Markets

We examine asset return volatility forecastability as a function of the horizon over which the returns are calculated. We begin with daily returns and then aggregate to obtain non-overlapping h-day returns, h = 1, 2, 3, ..., 20. Our use of non-overlapping returns ensures that we need not account for the dependence that would arise if we used overlapping observations. We provide an example of the aggregation of returns in Figure 1, in which we show U.S. S&P 500 stock returns for four aggregation levels: 1-day, 5-day, 10-day, and 20-day. We also show ±2 standard deviation intervals, which naturally grow wider as the aggregation level increases. We use these ±2 standard deviation intervals as our [-c, c]

intervals from which we compute the hit sequences. It is natural and appropriate to let c change with the aggregation level. [12]

At each aggregation level, we measure volatility forecastability using the maximum-likelihood estimate of the non-unit eigenvalue of a first-order Markov process fit to the hit sequence. The likelihood function is [13]

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

where n_{ij} is the number of observations with value i followed by j. It is easy to solve analytically for the maximum likelihood estimators of π_{01} and π_{11}, which are

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}$$

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

---

[12] Throughout, when using c equal to two standard deviations, the intervals turn out to have approximately constant unconditional coverage.

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

11

theory associated with maximum-likelihood estimation of the transition matrix eigenvalue is only asymptotic. Thus, to assess the significance of a given eigenvalue estimate, attempting to account for the precise sample size at hand, we use simulation methods. In particular, at each aggregation level, we:

(a) De-mean the returns series and compute a constant ± 2 standard deviation interval.

(b) Compute the indicator sequence, find the estimate of p, $\hat{p}$, and the estimate of S,

$\hat{S}$.

(c) Using $\hat{p}$ and the relevant sample size T,

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

(c2) for each sample, compute $\hat{S}_m$

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

$\hat{S}$ computed in (b).

We now proceed to analyze stock and foreign exchange returns, after which we analyze bond returns.

Equity and Foreign Exchange Markets

We begin with daily stock and foreign exchange market returns. The stock return data are the U.S. S&P 500, the German DAX, the U.K. FTSE, and the Japanese TPX. The foreign exchange rate data are dollar rates for 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. [14]

Let us first discuss the runs tests. In Figure 2 we show the finite-sample p-values of

_____

[14] All data are from Datastream International and Bloomberg Financial Services.

the runs tests of independence of the indicator sequence, as a function of the aggregation level. It is clear that, for each equity index, the p-values tend to rise with the aggregation level, 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 aggregation levels 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 aggregation levels greater than ten days. Figure 3 reveals identical patterns for foreign exchange rates.

Now let us discuss the estimated transition matrix eigenvalues. In Figures 4 and 5 we show the estimated eigenvalues along with their simulated finite-sample 95% confidence intervals, again as a function of the aggregation level. A consistent pattern emerges across all equities and foreign 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 aggregation level. By the time one reaches ten-day returns -- and often substantially before -- the estimated eigenvalues are small and statistically insignificant, indicating that volatility forecastability has vanished.

Bond Markets

We analyze bonds separately for two reasons. 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 potentially erroneous approximations not required to compute equity and foreign exchange returns. Second, as we shall show, patterns of bond-market volatility forecastability appear to differ from those in equity and foreign exchange markets.

First we provide an approximation to the bond 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 [15]

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

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 [16]

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

To obtain a workable expression for bond returns, use the fact that $\Delta(1+Y_{cnt}) = \Delta Y_{cnt}$ to

---

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

[16] 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).

14

rewrite the derivative formula for duration as

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

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

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

Armed with a workable approximation to bond returns, we now examine the forecastability of bond return volatility. Limited availability of historical daily international fixed income data forces us to focus exclusively on the 10 year U.S. Government bond. As before, the sample starts on January 1, 1973 and ends on May 1, 1997, resulting in 6350 daily observations. The results, which appear in Figures 6 (runs test p-values) and 7 (transition matrix eigenvalues), indicate substantially more volatility forecastability than in the equity or foreign exchange markets, with some forecastability out as far, say, as 15-20 trading days. It is hard to determine whether the apparently greater bond market volatility predictability is real, or whether it is an artifact of the approximation used to calculate bond returns. A third possibility -- a structural break in Federal Reserve policy around 1980 -- may also be operative and would produce the 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

15

more forecastable in bond markets than elsewhere is consistent with existing evidence, including Engle, Lilien, and Robins (1987) and Andersen and Lund (1997). [17]

Statistical Power

In any analysis involving the statistical testing of hypotheses, including ours, one must be concerned with power of the tests. Although the power of our tests likely drops with aggregation level, because of the decreasing sample sizes, it is unlikely that the apparent decrease in forecastability with aggregation level is simply due to a power drop, for at least three reasons. First, the point estimates of our forecastability measure, the second eigenvalues of the estimated Markov transition probability matrices fit to hit sequences, also decrease quickly with aggregation level, quite apart from their statistical significance. Second, the tests we use are among the best available; as we noted earlier, the runs test is uniformly most powerful against the first-order Markov alternative. Third, although sample size decreases with aggregation level, it is never small. Even for 20-day returns, the longest horizon examined, the sample size is still larger than 300.

4. Concluding Remarks

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 is effectively not forecastable. Our results clash with the assumptions embedded in popular risk management paradigms, which effectively assume highly persistent volatility dynamics. J.P. Morgan's (1996) RiskMetrics, for example, is built upon exponential smoothing of squared returns,

---

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

which is roughly equivalent to forecasting volatility with an integrated GARCH specification. Our results are, however, consistent with academic studies such as West and Cho (1995), who find that volatility in foreign exchange markets is unforecastable beyond a 5-day horizon. [18]

Our results are also consistent with those of Andersen and Bollerslev (1997b). Andersen and Bollerslev (1997b) study a situation in which the object of interest is the *average* volatility over an interval, as is relevant, in particular, for options pricing, and they appropriately question evidence of the sort provided by Figlewski (1994) and Jorion (1995), which seems to indicate that ARCH models provide poor volatility forecasts. The object of interest in risk management, however, is different; it is the volatility of *end-of-period* portfolio value. We have shown that end-of-period volatility is not forecastable when the holding period is more than ten or twenty days, which does not preclude the improved forecasting of *average* volatility and hence improved options pricing. In short, our focus and results are simply different from, but not inconsistent with, those of Andersen and Bollerslev, despite their superficial disparity.

It would seem to be more difficult to reconcile our results with those of Andersen and Bollerslev (1997a), who find evidence of long memory in very high-frequency (five-minute) exchange rate returns, which indicates that volatility should be forecastable well into the future. Again, however, our results are not necessarily inconsistent. The long memory that Andersen and Bollerslev find at five-minute intervals may well indicate that volatility is highly forecastable many steps into the future, perhaps 100 steps or even 1000 steps. 1000

---

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

five-minute steps, however, are just more than 3 days; even 5000 five-minute steps are just over 17 days. It is entirely possible that the long memory in volatility that seems clearly operative at five-minute intervals may be largely irrelevant for risk management at horizons of ten or twenty days.

If volatility dynamics are not important for risk management beyond horizons of ten or twenty days, 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 and applied to financial risk management by Danielsson and de Vries (1997).

References

Andersen, T.G. and Bollerslev, T. (1997a), "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. (1997b), "Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts," Manuscript, Northwestern University and University of Virginia.

Andersen, T. and Lund, J. (1996), "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. (1996), "Risk Management by Insurers: An Analysis of the Process," Wharton Financial Institutions Center Working Paper 96-16.

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

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*, in press.

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. (1997), "Converting 1-Day Volatility to h-Day Volatility: Scaling by $\sqrt{h}$ is Worse Than You Think," Wharton Financial Institutions Center Working Paper 97-34.

Drost, F.C. and Nijman, T.E. (1993), "Temporal Aggregation of GARCH Processes," *Econometrica*, 61, 909-927.

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

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.

Figlewski, S. (1994), "Forecasting Volatility Using Historical Data," Manuscript, Stern School of Business, New York University.

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. and O'Connell, P.G.J. (1997), "On the Pricing of Intermediated Risks: Theory and Application to Catastrophe Reinsurance" NBER Working Paper No. 6011.

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.

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.

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

J.P. Morgan (1996) "RiskMetrics - Technical Document," 4th Edition, New York.

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

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.

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.

Oldfield, G.S. and Santomero, A.M. (1997), "The Place of Risk Management in Financial Institutions," *Sloan Management Review* , forthcoming.

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* , forthcoming.

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

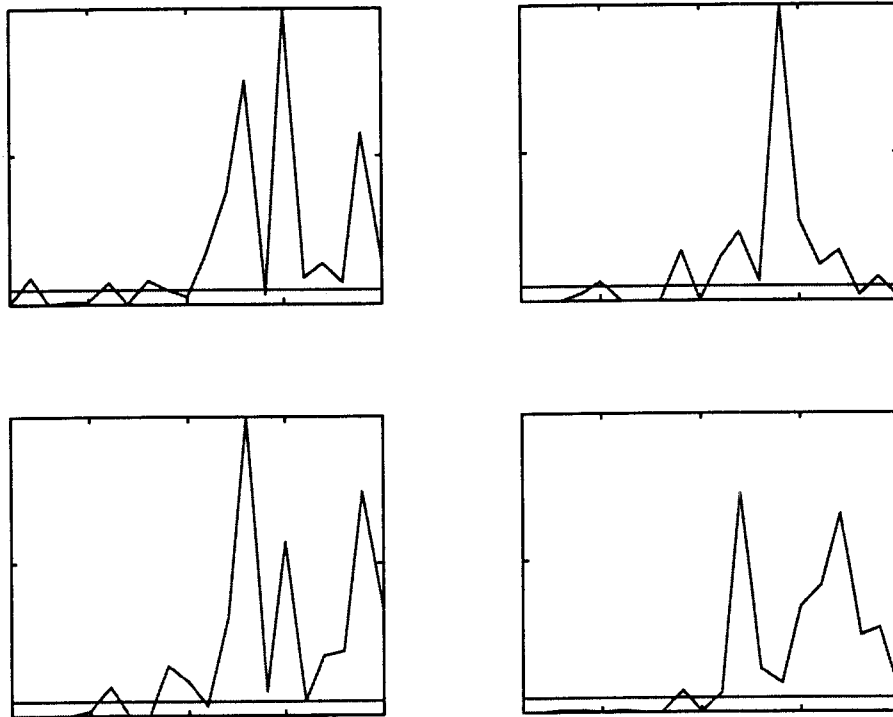**Figure 1**

S&P 500 Returns at Different Aggregation Levels
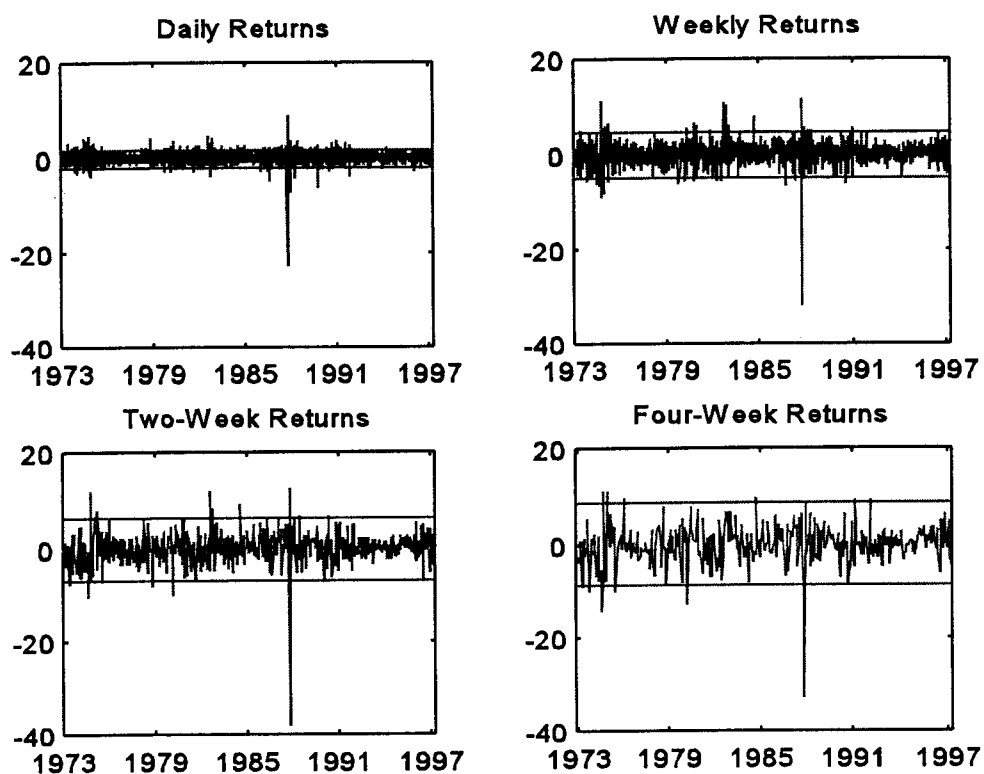
Notes to figure: We plot the daily, weekly, two-week and 4-week non-overlapping percent returns on the S&P 500 index together with unconditional ±2 standard deviation intervals.

Figure 2
P-Values of Runs Tests, Four Equity Markets
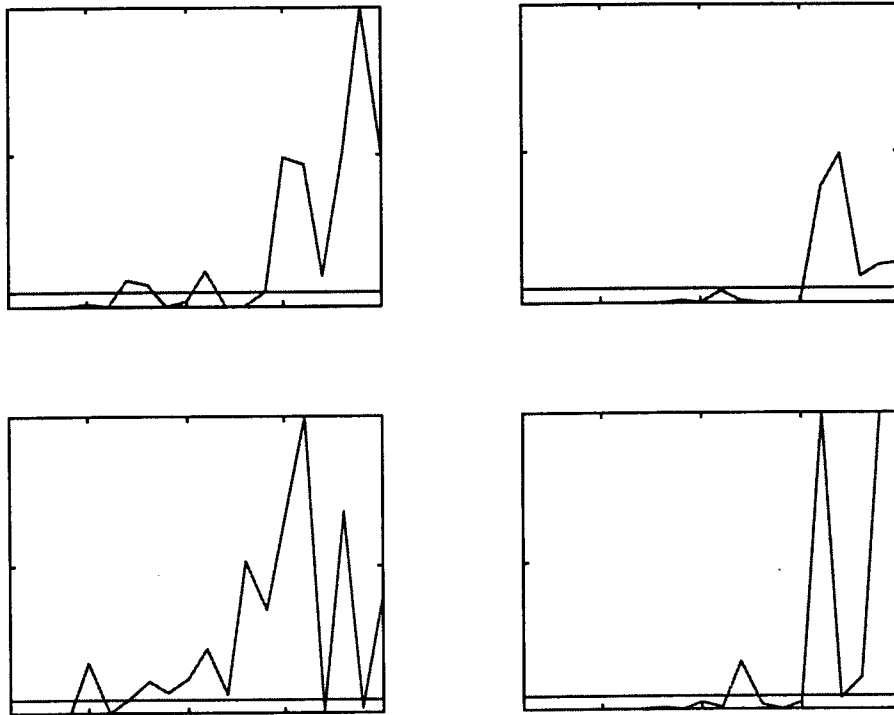
Notes to figure: For each series and horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant ±2 standard deviation interval forecast. The horizontal line is at 5 percent.
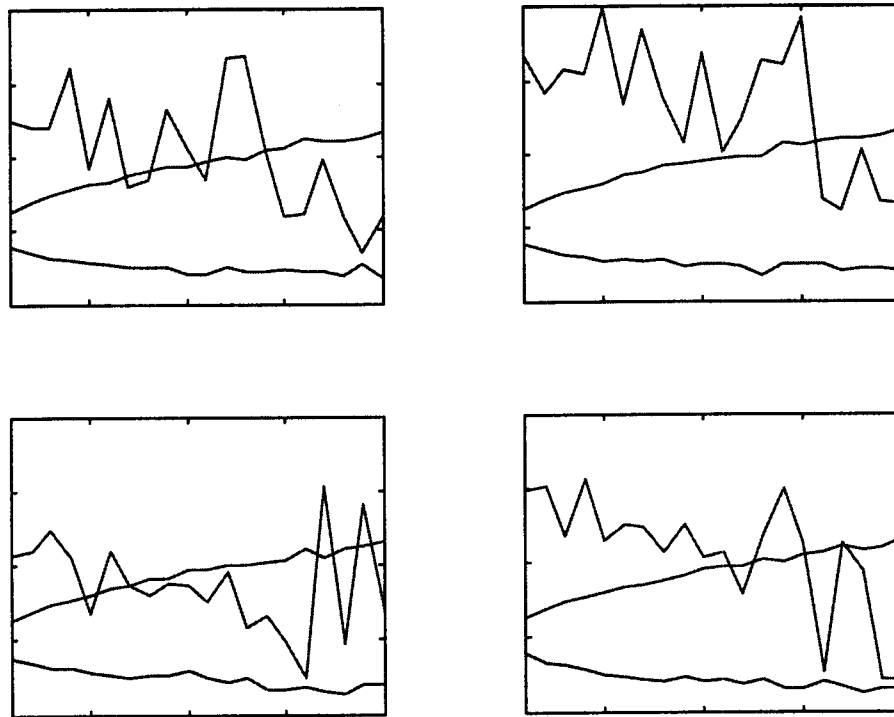
Figure 3
P-Values of Runs Tests, Four Foreign Exchange Markets

es to figure: For each series and each horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant ±2 standard deviation interval forecast. The horizontal line is at 5 percent.
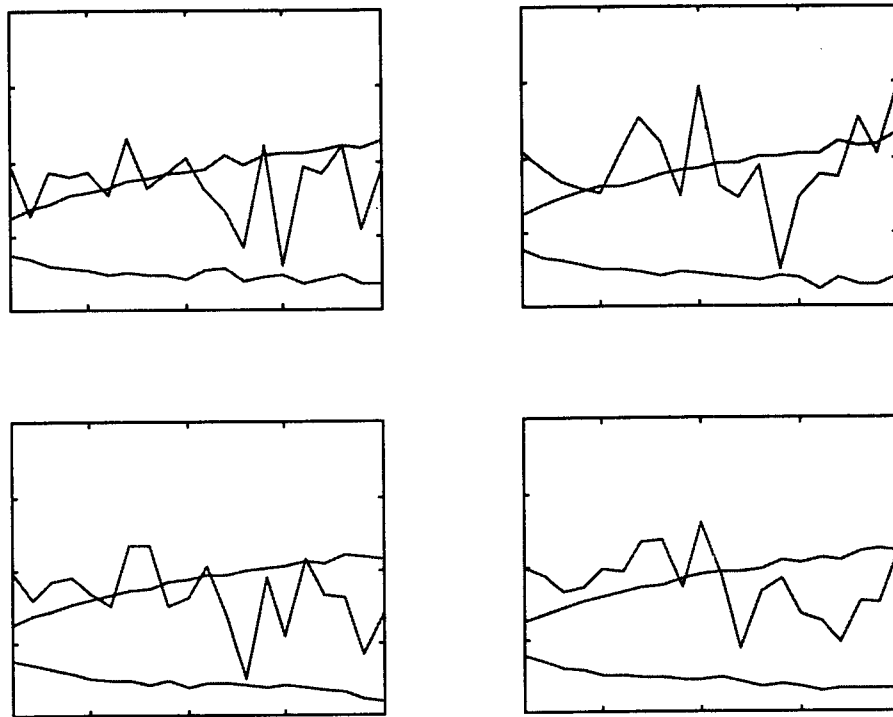
Figure 4
Markov Transition Matrix Eigenvalues, Four Equity Markets
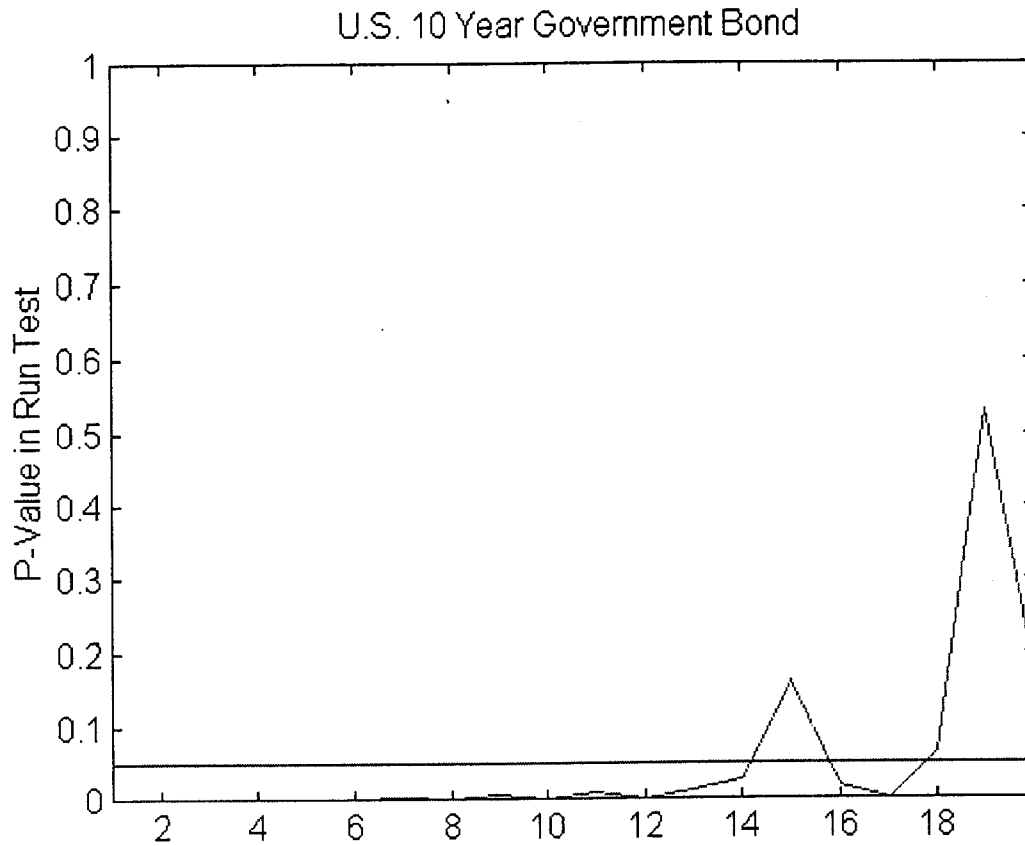

Notes to figure: For each series and each horizon we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant ±2 standard deviation interval forecast, along with the finite-sample 95 percent confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4000 simulations.

Figure 5
Markov Transition Matrix Eigenvalues, Four Foreign Exchange Markets
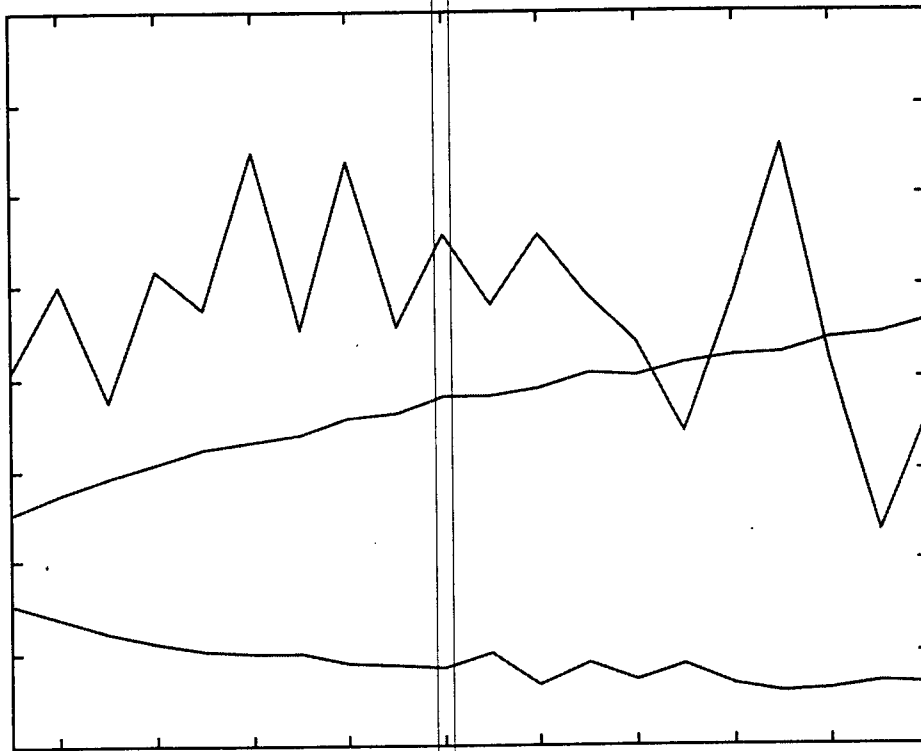

Notes to figure: For each series and each horizon we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant ±2 standard deviation interval forecast, along with the finite-sample 95 percent confidence interval when the eigenvalue is zero. We construct the finite-sample confidence interval from empirical percentiles based on 4000 simulations.

Figure 6
P-Values of Runs Tests, U.S. Ten-Year Government Bond

U.S. 10 Year Government Bond

Notes to figure: For each horizon we plot the finite-sample p-value associated with the runs test on the hit sequence corresponding to a constant ±2 standard deviation interval forecast. The horizontal line is at 5 percent.

Figure 7
Markov Transition Matrix Eigenvalues
U.S. Ten-Year Government Bond

Notes to figure: For each horizon we plot the estimated eigenvalue of the transition matrix estimated from the hit sequence corresponding to a constant ±2 standard deviation 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.