Horizon Problems and Extreme Events in Financial Risk Management

Peter F. Christoffersen, Francis X. Diebold, and Til Schuermann

I. Introduction

There is no one "magic" relevant horizon for risk management. Instead, the relevant horizon will generally vary by asset class (for example, equity versus bonds), industry (banking versus insurance), position in the firm (trading desk versus chief financial officer), and motivation (private versus regulatory), among other things, and thought must be given to the relevant horizon on an application-by-application basis. But one thing is clear: in many risk management situations, the relevant horizons are long—certainly longer than just a few days—an insight incorporated, for example, in Bankers Trust's RAROC system, for which the horizon is one year.

Simultaneously, it is well known that short-horizon asset return volatility fluctuates and is highly forecastable, a phenomenon that is very much at the center of modern risk management paradigms. Much less is known, however, about the forecastability of long-horizon volatility, and the speed and pattern with which forecastability decays as the horizon lengthens. A key question arises: Is volatility forecastability important for long-horizon risk management, or is a traditional constant-volatility assumption adequate?

In this paper, we address this question, exploring the interface between long-horizon financial risk management and long-horizon volatility forecastability and, in particular, whether long-horizon volatility is forecastable enough such that volatility models are useful for long-horizon risk management. In particular, we report on recent relevant work by Diebold, Hickman, Inoue, and Schuermann (1998); Christoffersen and Diebold (1997); and Diebold, Schuermann, and Stroughair (forthcoming).

To assess long-horizon volatility forecastability, it is necessary to have a measure of long-horizon volatility, which can be obtained in a number of ways. We proceed in Section II by considering two ways of converting short-horizon volatility into long-horizon volatility: scaling and formal model-based aggregation. The defects of those procedures lead us to take a different approach in Section III, estimating volatility forecastability directly at the horizons of interest, without making assumptions about the nature of the volatility process, and arriving at a surprising conclusion: Volatility forecastability seems to decline quickly with horizon, and seems to have largely vanished beyond horizons of ten or fifteen trading days.

Peter F. Christoffersen is an assistant professor of finance at McGill University. Francis X. Diebold is a professor of economics and statistics at the University of Pennsylvania, a research fellow at the National Bureau of Economic Research, and a member of the Oliver Wyman Institute. Til Schuermann is head of research at Oliver, Wyman & Company.

If volatility forecastability is not important for risk management beyond horizons of ten or fifteen trading days, then what *is* important? The really big movements such as the U.S. crash of 1987 are still poorly understood, 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 potentially facilitated by recent advances in extreme value theory. We explore that idea in Section IV, and we conclude in Section V.

II. Obtaining Long-Horizon Volatilities from Short-Horizon Volatilities: Scaling and Formal Aggregation[1]

Operationally, risk is often assessed at a short horizon, such as one day, and then converted to other horizons, such as ten days or thirty days, by scaling by the square root of horizon [for instance, as in Smithson and Minton (1996a, 1996b) or J.P. Morgan (1996)]. For example, to obtain a ten-day volatility, we multiply the one-day volatility by $\sqrt{10}$. Moreover, the horizon conversion is often significantly longer than ten days. Many banks, for example, link trading volatility measurement to internal capital allocation and risk-adjusted performance measurement schemes, which rely on annual volatility estimates. The temptation is to scale one-day volatility by $\sqrt{252}$. It turns out, however, that scaling is both inappropriate and misleading.

Scaling Works in iid Environments

Here we describe the restrictive environment in which scaling *is* appropriate. Let v_t be a log price at time t, and suppose that changes in the log price are independently and identically distributed,

$$v_t = v_{t-1} + \varepsilon_t \qquad \varepsilon_t \overset{iid}{\sim} (0, \sigma^2) .$$

Then the one-day return is

$$v_t - v_{t-1} = \varepsilon_t ,$$

with standard deviation σ. Similarly, the h-day return is

$$v_t - v_{t-h} = \sum_{i=0}^{h-1} \varepsilon_{t-i} ,$$

with variance $h\sigma^2$ and standard deviation $\sqrt{h\sigma}$. Hence, the "$\sqrt{h}$ rule": to convert a one-day standard deviation to an h-day standard deviation, simply scale by $\sqrt{h}$. For some applications, a percentile of the distribution of h-day returns may be desired; percentiles also scale by $\sqrt{h}$ if log changes are not only iid, but also normally distributed.

Scaling Fails in Non-iid Environments

The scaling rule relies on one-day returns being iid, but high-frequency financial asset returns are distinctly *not* iid. Even if high-frequency portfolio returns are conditional-mean independent (which has been the subject of intense debate in the efficient markets literature), they are certainly not conditional-variance independent, as evidenced by hundreds of recent papers documenting strong volatility persistence in financial asset returns.[2]

To highlight the failure of scaling in non-iid environments and the nature of the associated erroneous long-horizon volatility estimates, we will use a simple GARCH(1,1) process for one-day returns,

$$y_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$
$$\varepsilon_t \sim NID(0, 1),$$

$t = 1, \ldots, T$. We impose the usual regularity and covariance stationarity conditions, $0 < \omega < \infty$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$. The key feature of the GARCH(1,1) process is that it allows for time-varying conditional volatility, which occurs when α and/or β is nonzero. The model has been fit to hundreds of financial series and has been tremendously successful empirically; hence its popularity. We hasten to add, however, that our general thesis—that scaling fails in the non-iid environments associated with high-frequency asset returns—does not depend in any way on a GARCH(1,1) structure. Rather, we focus on the GARCH(1,1) case because it has been studied the most intensely, yielding a wealth of results that enable us to illustrate the failure of scaling both analytically and by simulation.

Drost and Nijman (1993) study the temporal aggregation of GARCH processes.[3] Suppose we begin with a sample path of a one-day return series, $\{y_{(1)t}\}_{t=1}^{T}$, which

follows the GARCH(1,1) process above.[4] Then Drost and Nijman show that, under regularity conditions, the corresponding sample path of h-day returns, $\{y_{(h)t}\}_{t=1}^{T/h}$, similarly follows a GARCH (1,1) process with

$$\sigma^2_{(h)t} = \omega_{(h)} + \beta_{(h)}\sigma^2_{(h)t-1} + \alpha_{(h)}y^2_{(h)t-1},$$

where

$$\omega_{(h)} = h\omega\frac{1-(\beta+\alpha)^h}{1-(\beta+\alpha)}$$

$$\alpha_{(h)} = (\beta+\alpha)^h - \beta_{(h)},$$

and $|\beta_{(h)}| < 1$ is the solution of the quadratic equation,

$$\frac{\beta_{(h)}}{1+\beta^2_{(h)}} = \frac{a(\beta+\alpha)^h - b}{a(1+(\beta+\alpha)^{2h}) - 2b},$$

where

$$a = h(1-\beta)^2 + 2h(h-1)\frac{(1-\beta-\alpha)^2(1-\beta^2-2\beta\alpha)}{(\kappa-1)(1-(\beta+\alpha)^2)} + 4\frac{(h-1-h(\beta+\alpha)+(\beta+\alpha)^h)(\alpha-\beta\alpha(\beta+\alpha))}{1-(\beta+\alpha)^2}$$

$$b = (\alpha-\beta\alpha(\beta+\alpha))\frac{1-(\beta+\alpha)^{2h}}{1-(\beta+\alpha)^2},$$

and κ is the kurtosis of y_t. The Drost-Nijman formula is neither pretty nor intuitive, but it is important, because it is the key to correct conversion of one-day volatility to h-day volatility. It is painfully obvious, moreover, that the $\sqrt{h}$ scaling formula does not look at all like the Drost-Nijman formula.

Despite the fact that the scaling formula is incorrect, it would still be very useful if it was an accurate approximation to the Drost-Nijman formula, because of its simplicity and intuitive appeal. Unfortunately, such is not the case. As $h \to \infty$, the Drost-Nijman results, which build on those of Diebold (1988), reveal that $\alpha_{(h)} \to 0$ and $\beta_{(h)} \to 0$, which is to say that temporal aggregation produces gradual disappearance of volatility fluctuations. Scaling, in contrast, *magnifies* volatility fluctuations.

A Worked Example

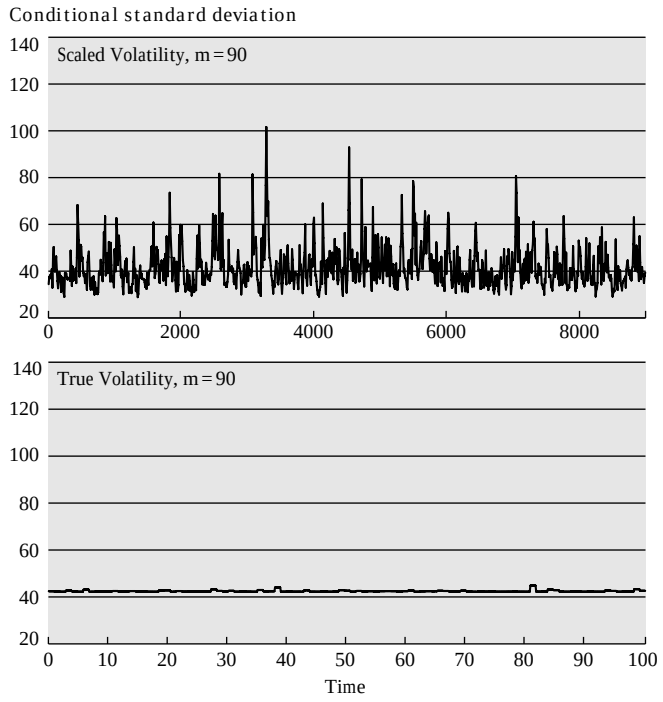
Let us examine the failure of scaling by $\sqrt{h}$ in a specific example. We parameterize the GARCH(1,1) process to be realistic for daily returns by setting $\alpha = 0.10$ and $\beta = 0.85$, which are typical of the parameter values obtained for estimated GARCH(1,1) processes. The choice of ω is arbitrary; we set $\omega = 1$.

The GARCH(1,1) process governs one-day volatility; now let us examine ninety-day volatility. In Chart 1, we show ninety-day volatilities computed in two different ways. We obtain the first (incorrect) ninety-day volatility by scaling the one-day volatility, σ_t, by $\sqrt{90}$. We obtain the second (correct) ninety-day volatility by applying the Drost-Nijman formula.

It is clear that although scaling by $\sqrt{h}$ produces volatilities that are correct on average, it magnifies the volatility fluctuations, whereas they should in fact be damped. That is, scaling produces erroneous conclusions of large fluctuations in the conditional variance of long-horizon returns, when in fact the opposite is true. Moreover, we cannot claim that the scaled volatility estimates are "conservative" in any sense; rather, they are sometimes too high and sometimes too low.

Chart 1

Ninety-Day Volatility, Scaled and Actual



Formal Aggregation Has Problems of Its Own

One might infer from the preceding discussion that formal aggregation is the key to converting short-horizon volatility estimates into good, long-horizon volatility estimates, which could be used to assess volatility forecastability. In general, such is not the case; formal aggregation has at least two problems of its own. First, temporal aggregation formulae are presently available only for restrictive classes of models; the literature has progressed little since Drost and Nijman. Second, the aggregation formulae assume the truth of the fitted model, when in fact the fitted model is simply an approximation, and the best approximation to *h*-day volatility dynamics is not likely to be what one gets by aggregating the best approximation (let alone a mediocre approximation) to one-day dynamics.

III. Model-Free Assessment of Volatility Forecastability at Different Horizons[5]

The model-dependent problems of scaling and aggregating daily volatility measures motivate the model-free investigation of volatility forecastability in this section. If the true process is GARCH(1,1), we know that volatility is forecastable at all horizons, although forecastability will decrease with horizon in accordance with the Drost-Nijman formula. But GARCH is only an approximation, and in this section we proceed to develop procedures that allow for assessment of volatility forecastability across horizons with no assumptions made on the underlying volatility model.

The Basic Idea

Our model-free methods build on the methods for evaluation of interval forecasts developed by Christoffersen (forthcoming). Interval forecasting is very much at the heart of modern financial risk management. The industry standard value-at-risk measure is effectively the boundary of a one-sided interval forecast, and just as the adequacy of a value-at-risk forecast depends crucially on getting the volatility dynamics right, the same is true for interval forecasts more generally.

Suppose that we observe a sample path $\{y_t\}_{t=1}^{T}$ of the asset return series y_t and a corresponding sequence of one-step-ahead interval forecasts, $\{(L_{t|t-1}(p), U_{t|t-1}(p))\}_{t=1}^{T}$, where $L_{t|t-1}(p)$ and $U_{t|t-1}(p)$ denote the lower and upper limits of the interval forecast for time t made at time $t-1$ with desired coverage probability p. We can think of $L_{t|t-1}(p)$ as a value-at-risk measure, and $U_{t|t-1}(p)$ as a measure of potential upside. The interval forecasts are subscripted by t as they will vary through time in general: in volatile times a good interval forecast should be wide and in tranquil times it should be narrow, keeping the coverage probability, p, fixed.

Now let us formalize matters slightly. Define the *hit sequence*, I_t, as

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

for $t = 1, 2, \ldots, T$. We will say that a sequence of interval forecasts has correct *unconditional coverage* if $E[I_t] = p$ for all t, which 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 and other higher order dynamics, it is important to check for adequacy of conditional coverage, which is a stronger concept. We will 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. The key result is that if $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, \ldots, I_1\}$, then correct conditional coverage is equivalent to $\{I_t\} \overset{iid}{\sim}$ Bernoulli(p), which can readily be tested.

Consider now the case where no volatility dynamics are present. The optimal interval forecast is then constant, and given by $\{(L(p), U(p))\}$, $t = 1, \ldots, T$. In that case, testing for correct conditional coverage will reveal no evidence of dependence in the hit sequence, and it is exactly the independence part of the iid Bernoulli(p) criterion that is designed to pick up volatility dynamics. If, however, volatility dynamics *are* present but ignored by a forecaster who

erroneously uses the constant $\{L(p), U(p)\}$ forecast, then a test for dependence in the hit sequence will reject the constant interval as an appropriate forecast: the ones and zeros in the hit sequence will tend to appear in time-dependent clusters corresponding to tranquil and volatile times.

It is evident that the interval forecast evaluation framework can be turned into a framework for assessing volatility forecastability: if a naive, constant interval forecast produces a dependent hit sequence, then volatility dynamics are present.

Measuring and Testing Dependence in the Hit Sequence

Now that we have established the close correspondence between the presence of volatility dynamics and dependence in the hit sequence from a constant interval forecast, it is time to discuss the measurement and testing of this dependence. We discuss two approaches.

First, consider a runs test, which is based on counting the number of strings, or *runs*, of consecutive zeros and ones in the hit sequence. If too few runs are observed (for example, 0000011111), the sequence exhibits positive correlation. Under the null hypothesis of independence, the exact finite sample distribution of the number of runs in the sequence has been tabulated by David (1947), and the corresponding test has been shown by Lehmann (1986) to be uniformly most powerful against a first-order Markov alternative.

We complement the runs test by a second measure, which has the benefit of being constrained to the interval [-1,1] and thus easily comparable across horizons and sequences. Let the hit sequence be first-order Markov with an arbitrary transition probability matrix. Then dependence is fully captured by the nontrivial eigenvalue, which is simply $S \equiv \pi_{11} - \pi_{01}$, where π_{ij} is the probability of a j following an i in the hit sequence. S is a natural persistence measure and has been studied by Shorrocks (1978) and Sommers and Conlisk (1979). 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.

An Example: The Dow Jones Composite Stock Index

We now put the volatility testing framework to use in an application to the Dow Jones Composite Stock Index, which comprises sixty-five major stocks (thirty industrials, twenty transportations, and fifteen utilities) on the New York Stock Exchange. The data start on January 1, 1974, and continue through April 2, 1998, resulting in 6,327 daily observations.

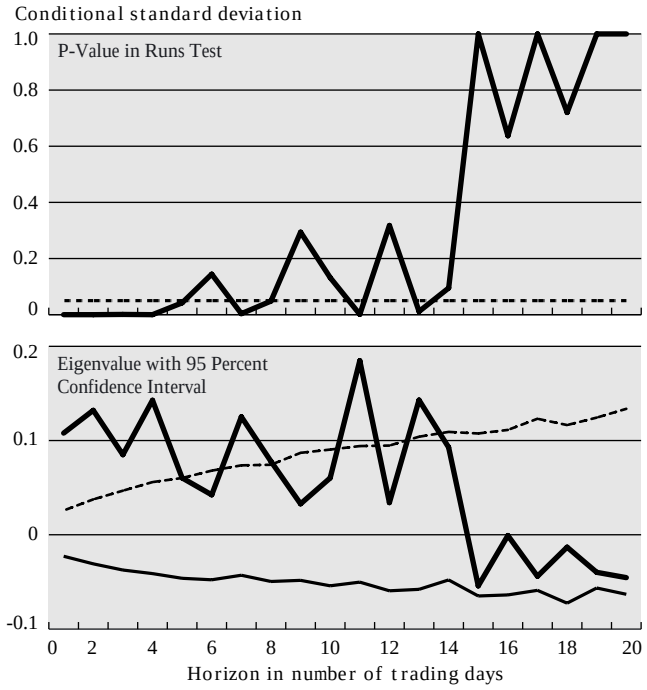
We examine asset return volatility forecastability as a function of the horizon over which the returns are calculated. We begin with daily returns and then aggregate to obtain nonoverlapping h-day returns, $h = 1, 2, 3, \ldots, 20$. We set $\{L(p), U(p)\}$ equal to ± 2 standard deviations and then compute the hit sequences. Because the standard deviation varies across horizons, we let the interval vary correspondingly. Notice that p might vary across horizons, but such variation is irrelevant: we are interested only in dependence of the hit sequence, not its mean.

At each horizon, we measure volatility forecastability using the *P*-value of the runs test—that is, the probability of obtaining a sample that is less likely to conform to the null hypothesis of independence than does the sample at hand. If the *P*-value is less than 5 percent, we reject the null of independence at that particular horizon. The top panel of Chart 2 on the next page shows the *P*-values across horizons of one through twenty trading days. Notice that despite the jaggedness of the line, a distinct pattern emerges: at short horizons of up to a week, the *P*-value is very low and thus there is clear evidence of volatility forecastability. At medium horizons of two to three weeks, the *P*-value jumps up and down, making reliable inference difficult. At longer horizons, greater than three weeks, we find no evidence of volatility forecastability.

We also check the nontrivial eigenvalue. In order to obtain a reliable finite-sample measure of statistical significance at each horizon, we use a simulation-based resampling procedure to compute the 95 percent confidence interval under the null hypothesis of no dependence in the hit sequence (that is, the eigenvalue is zero). In the bottom panel of Chart 2, we plot the eigenvalue at each

Chart 2

Volatility Persistence across Horizons in the Dow Jones Composite Index



Notes: The hit sequence is defined relative to a constant ±2 standard deviation interval at each horizon. The top panel shows the P-value for a runs test of the hypothesis that the hit sequence is independent. The horizontal line corresponds to a 5 percent significance level. The bottom panel shows the nontrivial eigenvalue from a first-order Markov process fit to the hit sequence. The 95 percent confidence interval is computed by simulation.

horizon along with its 95 percent confidence interval. The qualitative pattern that emerges for the eigenvalue is the same as for the runs test: volatility persistence is clearly present at horizons less than a week, probably present at horizons between two and three weeks, and probably not present at horizons beyond three weeks.

Multi-Country Analysis of Equity, Foreign Exchange, and Bond Markets

Christoffersen and Diebold (1997) assess volatility forecastability as a function of horizon for many more assets and countries. In particular, they analyze stock, foreign exchange, and bond returns for the United States, the United Kingdom, Germany, and Japan, and they obtain results very similar to those presented above for the Dow Jones composite index of U.S. equities.

For all returns, the finite-sample *P*-values of the runs tests of independence tend to rise with the aggregation level, although the specifics differ somewhat depending on the particular return 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 return volatility is significantly forecastable, and conversely for aggregation levels greater than ten days.

The estimated transition matrix eigenvalues tell the same story: 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.

IV. Forecasting Extreme Events[6]

The quick decay of volatility forecastability as the forecast horizon lengthens suggests that, if the risk management horizon is more than ten or fifteen trading days, less energy should be devoted to modeling and forecasting volatility and more energy should be devoted to modeling directly the extreme tails of return densities, a task potentially facilitated by recent advances in extreme value theory (EVT).[7] The theory typically requires independent and identically distributed observations, an assumption that appears reasonable for horizons of more than ten or fifteen trading days.

Let us elaborate. Financial risk management is intimately concerned with tail quantiles (for example, the value of the return, *y*, such that $P(Y > y) = .05$) and tail probabilities (for example, $P(Y > y)$, for a large value *y*). *Extreme* quantiles and probabilities are of particular interest, because the ability to assess them accurately translates into the ability to manage extreme financial risks effectively, such as those associated with currency crises, stock market crashes, and large bond defaults.

Unfortunately, traditional parametric statistical and econometric methods, typically based on estimation of

entire densities, may be ill-suited to the assessment of extreme quantiles and event probabilities. Traditional parametric methods implicitly strive to produce a good fit in regions where most of the data fall, potentially at the expense of a good fit in the tails, where, by definition, few observations fall. Seemingly sophisticated nonparametric methods of density estimation, such as kernel smoothing, are also well known to perform poorly in the tails.

It is common, moreover, to require estimates of quantiles and probabilities not only *near* the boundary of the range of observed data, but also *beyond* the boundary. The task of estimating such quantiles and probabilities would seem to be hopeless. A key idea, however, emerges from EVT: one can estimate extreme quantiles and probabilities by fitting a "model" to the empirical survival function of a set of data using only the extreme event data rather than all the data, thereby fitting the tail and only the tail.[8] The approach has a number of attractive features, including:

- the estimation method is tailored to the object of interest—the tail of the distribution—rather than the center of the distribution, and
- an arguably reasonable functional form for the tail can be formulated from a priori considerations.

The upshot is that the methods of EVT offer hope for progress toward the elusive goal of reliable estimates of extreme quantiles and probabilities.

Let us briefly introduce the basic framework. EVT methods of tail estimation rely heavily on a power law assumption, which is to say that the tail of the survival function is assumed to be a power law times a slowly varying function:

$$P(Y > y) = k(y)\ y^{-\alpha},$$

where the "tail index," α, is a parameter to be estimated. That family includes, for example, α-stable laws with $\alpha < 2$ (but not the Gaussian case, $\alpha = 2$).

Under the power law assumption, we can base an estimator of α directly on the extreme values. The most popular, by far, is due to Hill (1975). It proceeds by ordering the observations with $y_{(1)}$ the largest, $y_{(2)}$ the second largest, and so on, and forming an estimator based on the difference between the average of the m largest log returns and the m-th largest log return:

$$\alpha = \left(\left(\frac{1}{m} \sum_{i=1}^{m} ln(y_{(i)}) \right) - ln(y_{(m)}) \right)^{-1}.$$

It is a simple matter to convert an estimate of α into estimates of the desired quantiles and probabilities. The Hill estimator has been used in empirical financial settings, ranging from early work by Koedijk, Schafgans, and de Vries (1990) to more recent work by Danielsson and de Vries (1997). It also has good theoretical properties; it can be shown, for example, that it is consistent and asymptotically normal, assuming the data are iid and that m grows at a suitable rate with sample size.

But beware: if tail estimation via EVT offers opportunities, it is also fraught with pitfalls, as is any attempt to estimate low-frequency features of data from short historical samples. This has been recognized in other fields, such as the empirical finance literature on long-run mean reversion in asset returns (for instance, Campbell, Lo, and MacKinlay [1997, Chapter 2]). The problem as relevant for the present context—applications of EVT in financial risk management—is that for performing statistical inference on objects such as a "once every hundred years" quantile, the relevant measure of sample size is likely better approximated by the number of nonoverlapping hundred-year intervals in the data set than by the actual number of data points. From that perspective, our data samples are terribly small relative to the demands placed on them by EVT.

Thus, we believe that best-practice applications of EVT to financial risk management will benefit from awareness of its limitations, as well as the strengths. When the smoke clears, the contribution of EVT remains basic and useful: it helps us to draw smooth curves through the extreme tails of empirical survival functions in a way that is consistent with powerful theory. Our point is simply that we should not ask more of the theory than it can deliver.

V. 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 fifteen trading days, depending on the asset class, then volatility is effectively not forecastable. Our results question the assumptions embedded in popular risk management paradigms, which effectively assume much greater volatility forecastability at long horizons than appears consistent with the data, and suggest that for improving long-horizon risk management, attention is better focused elsewhere. One such area is the modeling of extreme events, the probabilistic nature of which remains poorly understood, and for which recent developments in extreme value theory hold promise.

ENDNOTES

We thank Beverly Hirtle for insightful and constructive comments, but we alone are responsible for remaining errors. The views expressed in this paper are those of the authors and do not necessarily reflect those of the International Monetary Fund.

1. This section draws on Diebold, Hickman, Inoue, and Schuermann (1997, 1998).

2. See, for example, the surveys of volatility modeling in financial markets by Bollerslev, Chou, and Kroner (1992) and Diebold and Lopez (1995).

3. More precisely, they define and study the temporal aggregation of *weak* GARCH processes, a formal definition of which is beyond the scope of this paper. Technically inclined readers should read "weak GARCH" whenever they encounter the word "GARCH" in this paper.

4. Note the new and more cumbersome, but necessary, notation: the subscript, which keeps track of the aggregation level.

5. This section draws on Christoffersen and Diebold (1997).

6. This section draws on Diebold, Schuermann, and Stroughair (forthcoming).

7. See the recent book by Embrechts, Klüppelberg, and Mikosch (1997), as well as the papers introduced by Paul-Choudhury (1998).

8. The survival function is simply 1 minus the cumulative density function, $1 - F(y)$. Note, in particular, that because $F(y)$ approaches 1 as y grows, the survival function approaches 0.

REFERENCES

Andersen, T., and T. Bollerslev. Forthcoming. "Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts." INTERNATIONAL ECONOMIC REVIEW.

Bollerslev, T., R. Y. Chou, and K. F. Kroner. 1992. "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence." JOURNAL OF ECONOMETRICS 52: 5-59.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. THE ECONOMETRICS OF FINANCIAL MARKETS. Princeton.: Princeton University Press.

Christoffersen, P. F. Forthcoming. "Evaluating Interval Forecasts." INTERNATIONAL ECONOMIC REVIEW.

Christoffersen, P. F., and F. X. Diebold. 1997. "How Relevant Is Volatility Forecasting for Financial Risk Management?" Wharton Financial Institutions Center Working Paper no. 97-45.

Danielsson, J., and C. G. de Vries. 1997. "Tail Index and Quantile Estimation with Very High Frequency Data." JOURNAL OF EMPIRICAL FINANCE 4: 241-57.

David, F. N. 1947. "A Power Function for Tests of Randomness in a Sequence of Alternatives." BIOMETRIKA 34: 335-9.

Diebold, F. X. 1988. EMPIRICAL MODELING OF EXCHANGE RATE DYNAMICS. New York: Springer-Verlag.

Diebold, F. X., A. Hickman, A. Inoue, and T. Schuermann. 1997. "Converting 1-Day Volatility to h-Day Volatility: Scaling by $\sqrt{h}$ is Worse Than You Think." Wharton Financial Institutions Center Working Paper no. 97-34.

———. 1998. "Scale Models." RISK 11: 104-7. (Condensed and retitled version of Diebold, Hickman, Inoue, and Schuermann [1997].)

Diebold, F. X., and J. Lopez. 1995. "Modeling Volatility Dynamics." In Kevin Hoover, ed., MACROECONOMETRICS: DEVELOPMENTS, TENSIONS AND PROSPECTS, 427-72. Boston: Kluwer Academic Press.

Diebold, F. X., T. Schuermann, and J. Stroughair. Forthcoming. "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management." In P. Refenes, ed., COMPUTATIONAL FINANCE. Boston: Kluwer Academic Press.

Drost, F. C., and T. E. Nijman. 1993. "Temporal Aggregation of GARCH Processes." ECONOMETRICA 61: 909-27.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 1997. MODELLING EXTREMAL EVENTS. New York: Springer-Verlag.

Hill, B.M. 1975. "A Simple General Approach to Inference About the Tail of a Distribution." ANNALS OF STATISTICS 3: 1163-74.

Koedijk, K. G., M. A. Schafgans, and C. G. de Vries. 1990. "The Tail Index of Exchange Rate Returns." JOURNAL OF INTERNATIONAL ECONOMICS 29: 93-108.

Lehmann, E. L. 1986. TESTING STATISTICAL HYPOTHESES. 2d ed. New York: John Wiley.

Morgan, J.P. 1996. "RiskMetrics Technical Document." 4th ed.

Paul-Choudhury, S. 1998. "Beyond Basle." RISK 11: 89. (Introduction to a symposium on new methods of assessing capital adequacy, RISK 11: 90-107.)

Shorrocks, A.F. 1978. "The Measurement of Mobility." ECONOMETRICA 46: 1013-24.

Smithson, C., and L. Minton. 1996a. "Value at Risk." RISK 9: January.

———. 1996b. "Value at Risk (2)." RISK 9: February.

Sommers, P. M., and J. Conlisk. 1979. "Eigenvalue Immobility Measures for Markov Chains." JOURNAL OF MATHEMATICAL SOCIOLOGY 6: 253-76.