INTERNATIONAL ECONOMIC REVIEW
Vol. 39, No. 4, November 1998

ANSWERING THE SKEPTICS: YES, STANDARD VOLATILITY MODELS DO PROVIDE ACCURATE FORECASTS*

BY TORBEN G. ANDERSEN AND TIM BOLLERSLEV†[1]

Northwestern University, U.S.A.
Duke University and National Bureau of Economic Research, U.S.A.

A voluminous literature has emerged for modeling the temporal dependencies in financial market volatility using ARCH and stochastic volatility models. While most of these studies have documented highly significant in-sample parameter estimates and pronounced intertemporal volatility persistence, traditional ex-post forecast evaluation criteria suggest that the models provide seemingly poor volatility forecasts. Contrary to this contention, we show that volatility models produce strikingly accurate interdaily forecasts for the latent volatility factor that would be of interest in most financial applications. New methods for improved ex-post interdaily volatility measurements based on high-frequency intradaily data are also discussed.

1. INTRODUCTION

Volatility permeates finance. The variation in economy-wide risk factors is important for the pricing of financial securities, and return volatility is a key input to option pricing and portfolio allocation problems. As such, accurate measures and good forecasts of volatility are critical for the implementation and evaluation of asset and derivative pricing theories as well as trading and hedging strategies. It is also a well-established fact, dating back to Mandelbrot (1963) and Fama (1965), that financial returns display pronounced volatility clustering. However, only over the last decade have financial economists begun to seriously model these temporal dependencies. While the vast majority of the earlier studies relied on the Autoregressive Conditional Heteroskedastic (ARCH) framework pioneered by Engle (1982), there is now a large and diverse time-series literature on volatility modeling. Almost univer-

* Manuscript received January 1998.
† E-mail: *boller@econ.duke.edu*
[1] This work was supported by a grant from the NSF to the NBER. We are grateful to Olsen and Associates for making the intradaily exchange rate quotations available. We would also like to thank Frank Diebold, an anonymous referee, David Hsieh, Jesper Lund, and seminar participants at the Board of Governors of the Federal Reserve System, the International Monetary Fund, the Wharton School, University of South California, Duke, Arizona State, Georgetown, and York Universities, U.S.A., Academia Sinica Taipei, Taiwan, the December 1996 *Triangle Econometrics Workshop*, the *Time Series Analysis of High Frequency Financial Data* conference, San Diego, 1997, the Fall 1997 Meetings of the NBER Asset Pricing Group, the *INFORMS* meeting, Dallas, U.S.A., 1997, and the *Issues in Empirical Finance* conference, London School of Economics, U.K., 1997, for valuable comments. Of course, we remain fully responsible for the content.

sally, reported results point towards a very high degree of intertemporal volatility persistence; see Bollerslev et al., (1992), Bollerslev et al., (1994), Ghysels et al., (1996), and Shephard (1996) for surveys. Yet, in spite of highly significant in-sample parameter estimates, numerous studies find that standard volatility models explain little of the variability in ex-post squared returns; see, Cumby et al., (1993), Figlewski (1997), and Jorion (1995, 1996). This has led to the suggestion that these models may be of limited practical value. In contrast, we show that well-specified volatility models provide strikingly accurate volatility forecasts. There is, in fact, no contradiction between good volatility forecasts and poor predictive power for daily squared returns. In a similar vein, we also demonstrate how high-frequency intraday data may be used constructively in forming more accurate and meaningful ex-post interdaily volatility measurements.

The intuition behind the apparent poor predictive power of well-specified volatility models is straightforward. Let the return innovation be written as $r_t = \sigma_t \cdot z_t$, where z_t denotes an independent mean zero, unit variance stochastic process, while the latent *volatility*, σ_t, evolves in accordance with the particular model entertained.[2] A common approach for judging the practical relevance of any model is to compare the implied predictions with the subsequent realizations. Unfortunately, volatility is not directly observed so this approach is not immediately applicable for volatility forecast evaluation. Still, if the model for σ_t^2 is correctly specified, then $E_{t-1}(r_t^2) = E_{t-1}(\sigma_t^2 \cdot z_t^2) = \sigma_t^2$, which appears to justify the use of the squared return innovation over the relevant horizon as a proxy for the ex-post volatility. However, while the squared innovation provides an unbiased estimate for the latent volatility factor, it may yield very noisy measurements due to the idiosyncratic error term, z_t^2. This component typically displays a large degree of observation-by-observation variation relative to σ_t^2, rendering the fraction of the squared return variation attributable to the volatility process low. Consequently, the poor predictive power of volatility models, when judged by standard forecast criteria using r_t^2 as a measure for ex-post volatility, is an inevitable consequence of the inherent noise in the return generating process.[3]

This motivates a fundamentally different approach. Rather than seeking to perfect the forecast evaluation procedures—taking the noisy observations on volatility provided by fixed-horizon squared returns as given—it may prove fruitful to pursue alternative ex-post volatility measures. Specifically, building on the continuous-time stochastic volatility framework developed by Nelson (1990) and Drost and Werker (1996), we demonstrate how high-frequency data allow for the construction of vastly improved ex-post volatility measurements via cumulative squared intraday returns.

[2] Throughout, we shall refer to σ_t and σ_t^2 interchangeably as volatility. This simplifies terminology and should cause no conceptual confusion, since the measures are linked by a simple monotonic transformation.

[3] This is analogous to the difficulty confronting models of expected returns and risk premia in asset pricing theory based on past and current information. The notoriously low explanatory power for period-by-period returns neither invalidates the theories nor renders them economically irrelevant. Genuine differences between the two scenario do nonetheless become evident later on, as we document that vastly improved empirical measures of ex-post daily volatility are feasible. No such remedies exist for the measurement of expected returns; see also Merton (1980).

In theory, as the observation frequency increases from a daily to an infinitesimal interval, this measure converges to genuine measurement of the latent volatility factor. In practice, this is infeasible because of data limitations and a host of market microstructure features, including nonsynchronous trading effects, discrete price observations, intraday periodic volatility patterns, and bid-ask spreads. Nonetheless, we find that the proposed volatility measures, based on high-frequency returns, provide a dramatic reduction in noise and a radical improvement in temporal stability relative to measures based on daily returns. Further, when evaluated against these improved volatility measurements, we find that daily volatility models perform well, readily explaining about half the variability in the volatility factor.[4] These findings are directly in line with the existing evidence documenting that the standardized residuals from estimated ARCH models are approximately i.i.d. through time; see Hsieh (1989). The 'hit sequence' proposed by Christoffersen (1998) and the integral transform of the series of density forecast analyzed by Diebold et al., (1998) similarly point towards the adequacy of standard interdaily volatility modeling and forecasting techniques.

The plan for the remainder of the paper is as follows. Notation and data sources are set forth in Section 2. Employing a daily sample of Deutschemark–U.S. Dollar (DM-\$) and Japanese Yen–U.S. Dollar (¥-\$) spot exchange rates along with the popular GARCH(1, 1) specification of Bollerslev (1986), Section 3 provides a brief empirical illustration of the highly significant ARCH parameter estimates typically obtained in-sample, and the associated poor out-of-sample forecasting performance vis-a-vis daily squared returns. Section 4 rationalizes the empirical findings in the context of a continuous-time stochastic volatility model. It also initiates the more constructive aspects of our analysis, as we show how the use of high-frequency data may reduce the measurement error involved in quantifying the ex-post latent volatility. Utilizing a one-year sample of five-minute returns, the empirical analysis in Section 5 highlights how the improved high-frequency based volatility measures give rise to radically different conclusions regarding the accuracy of the daily volatility forecasts for the two exchange rates discussed in Section 3. Section 6 concludes with suggestions for future research.

2. NOTATION AND DATA

To set forth notation, let p_t denote the time $t \geq 0$ logarithmic price for a financial asset, with the unit interval corresponding to one day. The discretely observed time series process of continuously compounded returns with m observations per day, or a return horizon of $1/m$, is then defined by,

$$r_{(m),t} \equiv p_t - p_{t-1/m}, \tag{1}$$

[4] In a related context, Hsieh (1991) and Fung and Hsieh (1991) report R^2's between 34 and 55 per cent when modeling volatility by autoregressions for daily sample standard deviations based on 15-minute equity, currency, and bond returns.

where $t = 1/m, 2/m, \dots$. In line with this convention, conditional and unconditional expectations are indexed by the observation frequency of the variables in the information set, and are denoted by $E_{(m),t}(\cdot)$ and $E_{(m)}(\cdot)$, respectively, while the corresponding variance operators are given by $Var_{(m),t}(\cdot)$ and $Var_{(m)}(\cdot)$. We further refer to the continuous-time instantaneous returns process by $r_t \equiv r_{(\infty),t} \equiv dp_t$, while the instantaneous variance is denoted σ_t^2. Likewise, the conditional expectation adapted to the continuous-time sample-path filtration, $\sigma(p_\tau, \tau \leq t)$, is referred to by $E_t(\cdot)$, whereas the corresponding unconditional expectation is denoted $E(\cdot)$. To facilitate comparison, all reported population figures and model estimates are scaled to reflect daily percentage returns.

The model estimates underlying the continuous-time simulations are based on daily returns, or $r_{(1),t}$, for the DM-\$ and the ¥-\$ spot exchange rates from October 1, 1987, through September 30, 1992. Meanwhile, the empirical out-of-sample forecast analysis is based on temporal aggregates of the five-minute returns, or $r_{(288),t}$, for the same two exchange rates from October 1, 1992, through September 30, 1993. These intraday returns are constructed from the linearly interpolated logarithmic midpoint of the continuously-recorded bid and ask quotes that appeared on the interbank Reuters network over the one-year sample. Due to the extremely low market activity over the weekends, the returns from Friday 21:00 Greenwich Mean Time (G.M.T.) through Sunday 21:00 G.M.T. are excluded, resulting in a total of 74,880 five-minute returns spanning 260 days. For a more detailed discussion of the data construction we refer to Andersen and Bollerslev (1997a, 1998), where the identical five-minute DM-\$ return series is analyzed from a different perspective.

3. INTERDAILY VOLATILITY MODELING AND FORECAST EVALUATION

The existence of volatility clustering in speculative returns is ubiquitous. Econometric modeling of this volatility clustering phenomenon has been a very active area of research over the past decade. Many of these studies find that the simple GARCH(1, 1) model provides a good first approximation to the observed temporal dependencies in daily data; see Baillie and Bollerslev (1989), Bollerslev (1987), Engle and Bollerslev (1986), and Hsieh (1989) for some of the early evidence.

3.1. *Daily Volatility Modeling.* In order to formally define the GARCH(1, 1) model, let $\sigma^2_{(m),t}$ denote the conditional variance of $r_{(m),t}$ based on information up through time $t - 1/m$. With a sampling frequency of m observations per day, the volatility model for $r_{(m),t}$ is then given by the following system,

$$r_{(m),t} = \sigma_{(m),t} \cdot z_{(m),t} \tag{2}$$

and

$$\sigma^2_{(m),t} = \psi_{(m)} + \alpha_{(m)} \cdot (\sigma_{(m),t-1/m} \cdot z_{(m),t-1/m})^2 + \beta_{(m)} \cdot \sigma^2_{(m),t-1/m} \tag{3}$$

where $\psi_{(m)} > 0$, $\alpha_{(m)} \geq 0$, $\beta_{(m)} \geq 0$, and $z_{(m),t}$ is i.i.d. with mean zero and variance one.

TABLE 1
DAILY GARCH(1, 1) MODEL ESTIMATES*

	DM-$	¥-$
$\psi_{(1)}$	0.022	0.026
	(0.009)	(0.008)
$\alpha_{(1)}$	0.068	0.104
	(0.016)	(0.026)
$\beta_{(1)}$	0.898	0.844
	(0.023)	(0.032)

*Reports Quasi-Maximum Likelihood estimates (QMLE) for the GARCH(1, 1) model defined in equation (3) based on the daily percentage returns for the Deutschemark–U.S. Dollar and Japanese Yen–U.S. Dollar spot exchange rates from October 1, 1987, through September 30, 1992. The QMLE are obtained under the assumption of conditional normality. Robust standard errors, as described in Bollerslev and Wooldridge (1992), are reported in parentheses.

The parameter estimates for the two daily exchange rates, corresponding to $m = 1$, are reported in Table 1.[5] The estimates for the conditional variance parameters are all highly significant, and the robust Wald tests for no ARCH effects, $\alpha_{(1)} = \beta_{(1)} = 0$, overwhelmingly reject for both rates. The Ljung-Box portmanteau tests for up to thirtieth-order serial correlation in the standardized squared residuals, $\hat{z}^2_{(1),t}$, equal 29.1 and 24.7, respectively, indicating that the GARCH(1, 1) model does a good job of tracking the short-run interdaily volatility dependencies. Consistent with the prior literature, the estimates for $\hat{\alpha}_{(1)} + \hat{\beta}_{(1)}$ are close to unity, as in the IGARCH(1, 1) model of Engle and Bollerslev (1986).[6] This high degree of volatility persistence, coupled with the significant parameter estimates, observed almost universally across different speculative returns, suggests that financial market volatility is highly predictable. Nonetheless, as we confirm in the following section, when judged by standard criteria, the model appears to provide poor forecasts, even over the immediate one-day-ahead horizon.

3.2. *Daily Volatility Forecast Evaluation.* The majority of the volatility forecast evaluations reported in the literature rely on some MSE criteria involving the ex-post squared or absolute returns over the relevant forecast horizon.[7] One particu-

[5] The estimates are quasi-maximum likelihood (QMLE) under the assumption that $z_{(1),t}$ is normally distributed, with robust standard errors in parentheses; see Bollerslev and Wooldridge (1992). The models also allow for intercepts in the conditional mean equations, but these estimates are indistinguishably different from zero and consequently not reported.

[6] Recent evidence suggest that the long-run dependencies in financial market volatility may be better characterized by a fractionally integrated, or FIGARCH, model; see Andersen and Bollerslev (1997b), Baillie et al., (1996), and Bollerslev and Mikkelsen (1996). Since the present analysis is focused exclusively on short-term volatility forecasting, we shall not pursue these more complicated specifications any further here.

[7] Although MSE is a natural choice when evaluating traditional forecasts for the mean, it is less obvious in a heteroskedastic nonlinear environment; see Bollerslev et al., (1994), Engle et al., (1993), Diebold and Mariano (1995), Diebold and Christoffersen (1997), Diebold et al., (1998), Lopez (1995), and West et al., (1993). We explicitly do not pursue any of these more complex forecast evaluation criteria here.

larly popular metric is obtained via the ex-post squared return-volatility regression,

$$r^2_{(m),t+1/m} = a_{(m)} + b_{(m)} \cdot \sigma^2_{(m),t+1/m} + u_{(m),t+1/m}, \tag{4}$$

where $t = 0, 1/m, 2/m, \ldots$. This regression equation provides an analog to a common procedure for evaluating forecasts for the conditional mean, termed the Mincer–Zarnowitz regression following Mincer and Zarnowitz (1969). If the model for the conditional variance is correctly specified and $E_{(m),t}(r^2_{(m),t+1/m}) = \sigma^2_{(m),t+1/m}$, it follows that, in population, $a_{(m)}$ and $b_{(m)}$ equals zero and unity, respectively.[8,9] Of course, in practice the values for $\sigma^2_{(m),t+1/m}$ are subject to estimation error, resulting in a standard errors-in-variables problem and a downward bias in the regression estimate for $b_{(m)}$.[10] Nonetheless, the coefficient of multiple determination, or $R^2_{(m)}$, from the regression in (4) provides a direct assessment of the variability in the ex-post returns, as measured by $r^2_{(m),t+1/m}$, that is explained by the particular estimates of $\sigma^2_{(m),t+1/m}$. The $R^2_{(m)}$ is therefore often interpreted as a simple gauge of the degree of predictability in the volatility process, and hence of the potential economic significance of the volatility forecasts.

The use of this $R^2_{(m)}$ as a guide to the accuracy of volatility forecasts is, however, problematic. Rational financial decision making hinges on the anticipated future volatility and not the subsequent realized squared returns. Under the null hypothesis that the estimated GARCH(1,1) model constitutes the correct specification, the true return variance is, by definition, identical to the GARCH volatility forecast. The regression in (4) relies on the observed squared returns as a measure of realized volatility. This is justified to the extent that they provide unbiased estimators of the underlying latent volatility. However, realized squared returns are poor estimators of day-by-day movements in volatility, as the idiosyncratic component of daily returns is large. It is therefore impossible to interpret the resulting $R^2_{(m)}$, unless we establish a benchmark for the value expected under the null hypothesis of correct model specification.[11]

To illustrate these points, consider the GARCH(1,1) estimates for the daily DM-\$ and ¥-\$ exchange rates. The $R^2_{(1)}$'s from the one-step-ahead return volatility regressions in (4) for the 260 weekday returns over the subsequent year from October 1,

[8] This assumes that the conditional mean of $r_{(m),t}$ is zero. Otherwise, replace $r^2_{(m),t}$ in equation (4) by $(r_{(m),t} - \mu_{(m),t})^2$, where $\mu_{(m),t}$ denotes the conditional mean; Pagan and Ullah (1988) and Pagan and Sabau (1992) analyze the complications that arise when the conditional mean depends on the conditional variance. At the daily horizon, any predictability in the mean is of second order importance.

[9] A closely related regression, $|r_{(m),t+1/m}| = c_{(m)} + d_{(m)} \cdot \sigma_{(m),t+1/m} + v_{(m),t+1/m}$, has been employed in a number of studies; e.g., Jorion (1995). However, unlike $b_{(m)}$, the population value of $d_{(m)}$ hinges on distributional assumptions. For simplicity we therefore concentrate exclusively on the squared return–volatility regression in (4) in the present analysis.

[10] If the forecasts are unbiased in population, the downward bias in the estimate for $b_{(m)}$ is given as $-\mathrm{Var}_{(m)}(\nu_{(m),t}) \cdot [\mathrm{Var}_{(m)}(\nu_{(m),t}) + \mathrm{Var}_{(m)}(\sigma^2_{(m),t})]^{-1}$, where $\nu_{(m),t}$ denotes the measurement error in $\sigma^2_{(m),t}$; e.g., Chow (1983). Christensen and Prabhala (1998) explicitly recognize this bias within the context of evaluating variance forecasts based on implied volatilities from options prices.

[11] The predication on R^2 as a convenient measure for summarizing predictable changes in returns is underscored by Roll (1988) in his 1987 Presidential Address to the American Finance Association, succinctly entitled 'R^2.'

1992 through September 30, 1993, equal 0.047 and 0.026, respectively.[12] These 'disappointingly' low $R^2_{(1)}$'s are in line with the evidence in the extant literature for other speculative returns and sample periods.[13] For instance, on evaluating the predictive power of a GARCH(1, 1) model for weekly returns on the S&P100 stock index from 1983–1989, Day and Lewis (1992) report $R^2_{(1/5)} = 0.039$, while Pagan and Schwert (1990) find $R^2_{(1/22)} = 0.067$ with a GARCH(1, 2) model for monthly aggregate U.S. stock market returns from 1835–1925. Jorion (1996) uses the same GARCH(1, 1) specification as here, but a longer seven-year sample of daily DM-$ returns from 1985–1992, to obtain $R^2_{(1)} = 0.024$. Modeling weekly stock and bond market volatility in the U.S. and Japan from 1977–1990 by an EGARCH model, Cumby et al., (1993) report $R^2_{(1/5)}$'s ranging from 0.003 to 0.106, while West and Cho (1995) find $R^2_{(1/5)}$'s ranging from 0.001 to 0.045 with a GARCH(1, 1) model for five different weekly U.S. dollar exchange rates from 1973–1989. Closely related results have been reported by Akgiray (1989), Boudoukh et al., (1997), Brailsford and Faff (1996), Canina and Figlewski (1993), Dimson and Marsh (1990), Frennberg and Hansson (1996), Figlewski (1997), Heynen and Kat (1994), Jorion (1995), Schwert (1989, 1990a), and Schwert and Seguin (1990). Predictably, these systematically low $R^2_{(m)}$'s reported throughout the literature have led to the perception that standard volatility models may be seriously misspecified and provide poor volatility forecasts, and consequently be of limited, if any, practical use.

To highlight the fallacy of such a conclusion, we derive the population R^2 under the null hypothesis that the returns are generated by a GARCH(1, 1) model as in equations (2) and (3). Letting $\kappa_{(m)} \equiv E_{(m),t}(z^4_{(m),t})$ denote the conditional kurtosis of the standardized innovations, it is straightforward to show that, provided the unconditional kurtosis for $r_{(m),t}$ is finite, or $\kappa_{(m)} \cdot \alpha^2_{(m)} + \beta^2_{(m)} + 2 \cdot \alpha_{(m)} \cdot \beta_{(m)} < 1$ (see Bollerslev 1986), we have

$$\begin{aligned}\mathrm{Var}_{(m)}\left(r^2_{(m),t}\right) = {} & \psi^2_{(m)}\left(\kappa_{(m)} - 1\right) \cdot \left(1 - \beta^2_{(m)} - 2 \cdot \alpha_{(m)} \cdot \beta_{(m)}\right) \\ & \cdot \left(1 - \kappa_{(m)} \cdot \alpha^2_{(m)} - \beta^2_{(m)} - 2 \cdot \alpha_{(m)} \cdot \beta_{(m)}\right)^{-1} \\ & \cdot \left(1 - \alpha_{(m)} - \beta_{(m)}\right)^{-2},\end{aligned}$$

and,

$$\begin{aligned}\mathrm{Var}_{(m)}\left(\sigma^2_{(m),t}\right) = {} & \psi^2_{(m)} \cdot \left(\kappa_{(m)} - 1\right) \cdot \alpha^2_{(m)} \cdot \left(1 - \kappa_{(m)} \cdot \alpha^2_{(m)} - \beta^2_{(m)} - 2 \cdot \alpha_{(m)} \cdot \beta_{(m)}\right)^{-1} \\ & \cdot \left(1 - \alpha_{(m)} - \beta_{(m)}\right)^{-2}.\end{aligned}$$

[12] Consistent with the results in Table 1, all our out-of-sample predictions use daily returns measured at 12:00 G.M.T. While the reported figures do reflect the actual definition of the daily return interval, the qualitative conclusions are robust. For instance, on measuring the daily returns at 0:00 G.M.T., the one-year out-of-sample $R^2_{(1)}$'s equal 0.021 and 0.012, respectively.

[13] Whereas the results for the DM-$ and ¥-$ exchange rates reported here are truly out-of-sample, most of the results reported in the literature rely on in-sample parameter estimates. If anything, this is likely to bias the $R^2_{(m)}$'s upward.

Thus, the (true) population $R^2_{(m)}$ from the regression in equation (4) equals,

$$(5)\quad R^2_{(m)} = \mathrm{Var}_{(m)}\left(\sigma^2_{(m),t}\right)\cdot \mathrm{Var}_{(m)}\left(r^2_{(m),t}\right)^{-1} = \alpha^2_{(m)}\cdot\left(1-\beta^2_{(m)}-2\cdot\alpha_{(m)}\cdot\beta_{(m)}\right)^{-1}.$$

By the implicit assumption of a finite unconditional fourth-order moment underlying the squared return–volatility regression, the coefficient of multiple determination cannot exceed $\kappa^{-1}_{(m)}$. In particular, with conditional Gaussian errors the $R^2_{(m)}$ from a correctly specified GARCH(1,1) model is bounded from above by $\frac{1}{3}$, while with conditional fat-tailed errors the upper bound is even lower. Moreover, with realistic parameter values for $\alpha_{(m)}$ and $\beta_{(m)}$, the population value for the $R^2_{(m)}$ statistic is significantly below this upper bound. In other words, low R^2's are not an anomaly, but rather a direct implication of standard volatility models.

Consider again the daily DM-\$ and ¥-\$ GARCH(1,1) parameter estimates for $\alpha_{(1)}$ and $\beta_{(1)}$ in Table 1. The population $R^2_{(1)}$'s equal 0.064 and 0.096, respectively. While these $R^2_{(1)}$'s are slightly higher than the actual one-year out-of-sample statistics calculated above, the values are in close accordance with the $R^2_{(m)}$'s reported in the extant literature.[14] Thus, even for a correctly specified model with $E_{(m),t}(r^2_{(m),t+1/m}) = \sigma^2_{(m),t+1/m}$, it is naive to expect a 'high' $R^2_{(m)}$ from the squared return–volatility regression in (4).

The fact that the daily GARCH(1,1) model do not explain much of the variability in the squared DM-\$ returns is also evident from Figure 1, which graphs the 260 one-day-ahead volatility forecasts from October 1, 1992 through September 30, 1993, along with the corresponding realized daily squared returns. The variability in $\sigma^2_{(1),t}$ is diminutive compared to the variability in $r^2_{(1),t}$. The next section further explores this issue within the context of a continuous-time stochastic volatility model.

4. CONTINUOUS-TIME VOLATILITY MODELING AND FORECAST EVALUATION

The previous section demonstrates that the low $R^2_{(m)}$ measures are consistent with standard volatility models. However, this finding does not settle the underlying fundamental issue of whether these models actually provide meaningful volatility forecasts. To address this question, we adopt a continuous-time diffusion framework with the property that all discretely sampled time series obey so-called weak-form GARCH(1,1) models. This approach has the added advantage that many asset pricing models and most derivatives pricing theories are cast in a similar framework.

Specifically, we assume that the instantaneous returns are generated by the continuous-time martingale,

$$(6)\qquad dp_t = \sigma_t\cdot dW_{p,t},$$

[14] By ignoring the higher volatility following market closures, the GARCH(1,1) models reported in Table 1 systematically over-estimate volatility on regular trading days, possibly explaining part of the discrepancy between the actual and population $R^2_{(1)}$'s; see Baillie and Bollerslev (1989) and Andersen and Bollerslev (1998) for a detailed analysis of day-of-the-week and holiday effects in the foreign exchange market. For simplicity, we do not pursue this additional complication here.

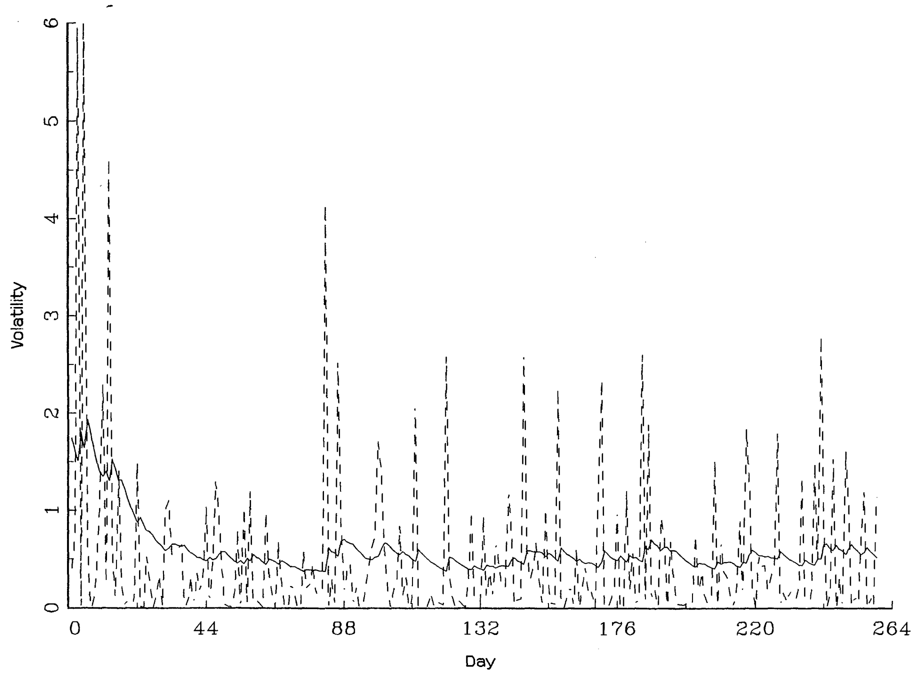

FIGURE 1

DAILY SQUARED RETURNS AND GARCH(1,1) FORECASTS. THE SOLID LINE GRAPHS THE DAILY ONE-STEP-AHEAD GARCH(1,1) VOLATILITY FORECASTS FOR THE DEUTSCHEMARK–U.S. DOLLAR EXCHANGE RATE GIVEN BY EQUATION (3) WITH $m \equiv 1$ AND THE PARAMETER VALUES IN TABLE 1. THE DOTTED LINE REFERS TO THE CORRESPONDING REALIZED DAILY SQUARED RETURNS, $r^2_{(1),t}$. THE SAMPLE PERIOD EXTENDS FROM OCTOBER 1, 1992 THROUGH SEPTEMBER 29, 1993.

where $W_{p,t}$ denotes a standard Wiener process.[15] By Ito's Lemma, the minimum MSE forecast for the conditional variance for the one-day returns, or $r_{(1),t+1} \equiv p_{t+1} - p_t$, is then readily expressed as,

$$(7) \qquad E_t\left(r^2_{(1),t+1}\right) = E_t\left(\int_0^1 r^2_{t+\tau}\, d\tau\right) = E_t\left(\int_0^1 \sigma^2_{t+\tau}\, d\tau\right) = \int_0^1 E_t\left(\sigma^2_{t+\tau}\right) d\tau .$$

Of course, with time-varying volatility it is generally the case that $E_{(1),t}(r^2_{(1),t+1}) \neq E_t(r^2_{(1),t+1})$. Thus, any discrete-time daily ARCH forecast is necessarily inefficient in a MSE sense relative to the optimal forecast based on the continuous sample path. Meanwhile, it is evident that the relevant notion of daily volatility in this setting becomes $\int_0^1 \sigma^2_{t+\tau}\, d\tau$. This quantity is also of central importance for the pricing of

[15] Any mean predictability could easily be incorporated into the subsequent analysis, but the assumption of serially uncorrelated mean-zero returns in (6) greatly simplifies the notation. This assumption is also consistent with the empirical evidence for the two exchange rates analyzed throughout.

derivative securities under stochastic volatility; see Hull and White (1987), Melino (1994), Scott (1987), and Wiggins (1987). Equation (7) shows that $r^2_{(1),t+1}$ continues to provide an unbiased, albeit noisy, estimator of the relevant latent volatility factor for daily returns, generalizing the results from the discrete-time setting discussed earlier.

4.1. *Continuous Time Modeling of Daily Volatility.* In our setting, the natural continuous-time model for the volatility process is given by the diffusion limit of the GARCH(1,1) process, as developed in Nelson (1990).[16] It takes the form,

$$d\sigma_t^2 = \theta\left(\omega - \sigma_t^2\right)\cdot dt + (2\lambda\theta)^{1/2}\sigma_t^2 \cdot dW_{\sigma,t}, \tag{8}$$

where $\omega > 0$, $\theta > 0$, $0 < \lambda < 1$, and the Wiener processes, $W_{p,t}$ and $W_{\sigma,t}$, are independent.

With the notable exception of the model in Meddahi and Renault (1997), exact discretization for stochastic volatility models are typically not available in closed form. However, following Drost and Nijman (1993) and Drost and Werker (1996) the discretely sampled returns defined by equations (6) and (8), $r_{(m),t} \equiv p_t - p_{t-1/m}$, satisfy the weak GARCH(1,1) model restrictions,

$$\varsigma^2_{(m),t} = \psi_{(m)} + \alpha_{(m)}\cdot r^2_{(m),t-1/m} + \beta_{(m)}\cdot \varsigma^2_{(m),t-1/m}, \tag{9}$$

where $\varsigma^2_{(m),t}$ refers to the linear projection of $r^2_{(m),t}$ on the Hilbert space spanned by $1, r_{(m),t-1/m}, r_{(m),t-2/m},\ldots$ and $r^2_{(m),t-1/m}, r^2_{(m),t-2/m},\ldots$. Although the formal interpretations differ, the recursions for the weak GARCH model defined by (9) and the conditional variance in (3) obviously result in identical numerical values for $\varsigma^2_{(m),t}$ and $\sigma^2_{(m),t}$. We therefore refer to the one-day-ahead weak GARCH(1,1) projections as $\sigma^2_{(1),t}$, instead of $\varsigma^2_{(1),t}$, in the sequel. However, insofar as $E_{(1),t}(r^2_{(1),t}) \neq \sigma^2_{(1),t}$, the results for the daily GARCH(1,1) forecasts provided only a lower bound on the predictability afforded by higher-order discrete-time ARCH approximations. Nonetheless, given the weak GARCH(1,1) interpretation of the diffusion approximation in (9), more complicated stochastic differential equations should at best result in minor improvements relative to the findings below.

The exact one-to-one relationship between the discrete-time weak GARCH(1,1) parameters and the continuous-time stochastic volatility parameters in equation (8) is conveniently expressed by,

$$\theta = -m\cdot\log\left(\alpha_{(m)} + \beta_{(m)}\right), \tag{10}$$

$$\omega = m\cdot\psi_{(m)}\cdot\left(1 - \alpha_{(m)} - \beta_{(m)}\right)^{-1}, \tag{11}$$

[16] Note, however, that many other properly designed ARCH filters will yield consistent estimates for the same σ_t process as the sampling frequency increases; see Nelson (1996) and Nelson and Foster (1994).

TABLE 2
IMPLIED CONTINUOUS TIME GARCH(1, 1): MODEL ESTIMATES*

	DM-$	¥-$
θ	0.035	0.054
ω	0.636	0.476
λ	0.296	0.480

*Gives the parameters for the continuous-time stochastic volatility model defined by equations (6) and (8), implied by the discrete-time daily GARCH(1, 1) model estimates reported in Table 1. The formal relationship between the continuous- and discrete-time parameters is detailed in equations (10), (11), and (12).

and

(12)

$$\begin{aligned}\lambda = {} & 2\cdot\alpha_{(m)}\cdot\log^2(\alpha_{(m)}+\beta_{(m)})\cdot\left[1-\beta_{(m)}\cdot(\alpha_{(m)}+\beta_{(m)})\right]\\ & \cdot\Big\{\left[1-(\alpha_{(m)}+\beta_{(m)})^2\right]\cdot(1-\beta_{(m)})^2-\alpha_{(m)}\cdot\left[1-\beta_{(m)}\cdot(\alpha_{(m)}+\beta_{(m)})\right]\\ & \cdot\left[6\cdot\log(\alpha_{(m)}+\beta_{(m)})+2\cdot\log^2(\alpha_{(m)}+\beta_{(m)})+4\cdot(1-\alpha_{(m)}-\beta_{(m)})\right]\Big\}^{-1}.\end{aligned}$$

Equation (10) implies that $\lim_{m\to\infty}(\alpha_{(m)}+\beta_{(m)})=1$, so the weak GARCH(1, 1) model converges to the IGARCH(1, 1) case of Engle and Bollerslev (1986) as the sampling frequency increases.

The continuous-time parameters implied by the daily, or $m=1$, GARCH(1, 1) estimates for the exchange rates reported in Table 1 are listed in Table 2. These parameters correspond quite closely to those implied by the daily GARCH(1, 1) estimates reported in Baillie and Bollerslev (1989), as inferred by Drost and Werker (1996). The parameters in Table 2 are also in line with the results reported elsewhere in the literature for other stochastic volatility models; see Andersen (1994), Jacquier et al., (1994), Shephard (1996) and the collection of papers in Rossi (1996). As such, the findings for the diffusion parameterizations in Table 2 may serve as a yardstick for the degree of predictability afforded by daily discrete-time ARCH approximations to the continuous-time specifications employed throughout the theoretical asset pricing literature.[17]

[17] While continuous-time diffusions provide a convenient framework for asset pricing, the specifications in (6) and (8) ignore pertinent market microstructure features. For instance, nonsynchronous trading induces negative serial correlation in individual returns, whereas index returns become positively correlated. Similarly, the bid-ask spread on organized exchanges, as well as the systematic positioning of quotes in dealer markets, cause the observed returns to be negatively serially correlated. Moreover, the return variances differ over trading versus nontrading periods, and there are pronounced intraday volatility patterns in financial markets. Several studies also argue for the simultaneous importance of jumps and time-varying volatility. The specification of richer continuous-time stochastic volatility models that accommodate some or all of the above features would be very interesting, but beyond the scope of the present analysis. See Goodhart and O'Hara (1997) for a recent survey of the relevant empirical literature.

4.2. *Continuous Time Measurement of Volatility.* The relevant gauge for the performance of daily volatility forecasts in the diffusion context is given by $\int_0^1 \sigma_{t+\tau}^2 \, d\tau$. Although the corresponding daily squared returns, $r_{(1),t+1}^2$, constitute an unbiased estimator of this quantity, it is also an extremely noisy estimator. Specifically, for the diffusions in Table 2, the population values of $E[(\int_0^1 \sigma_{t+\tau}^2 \, d\tau - r_{(1),t+1}^2)^2]$ equal 1.138 and 0.842, respectively, while the variances for the one-day-ahead latent volatility factor, $Var(\int_0^1 \sigma_{t+\tau}^2 \, d\tau)$, equal 0.166 and 0.191. The population values for the daily volatility variance are thus orders of magnitude less than the corresponding MSE's for the daily squared returns.[18]

To further illustrate the pitfalls in using the squared daily returns for ex-post volatility forecast evaluation, consider the following decomposition of the ideal one-day-ahead latent volatility forecast error for the GARCH(1,1) model,

(13)

$$E\left[\left(\sigma_{(1),t}^2 - \int_0^1 \sigma_{t-1+\tau}^2 \, d\tau\right)^2\right] = E\left[\left(\sigma_{(1),t}^2 - r_{(1),t}^2\right)^2\right] + E\left[\left(r_{(1),t}^2 - \int_0^1 \sigma_{t-1+\tau}^2 \, d\tau\right)^2\right]$$
$$+ 2 \cdot E\left[\left(\sigma_{(1),t}^2 - r_{(1),t}^2\right) \cdot \left(r_{(1),t}^2 - \int_0^1 \sigma_{t-1+\tau}^2 \, d\tau\right)\right].$$

The prediction error calculated in practice using squared daily returns is given by the first term on the right-hand-side of equation (13). For the diffusions in Table 2, this term equals 1.221 and 0.944, respectively. In contrast, the ideal MSE for each of the daily weak-form GARCH(1,1) models, given by the left-hand-side of equation (13), equal 0.084 and 0.097. This glaring discrepancy reflects the impact of the measurement error, comprised of the second and third term on the right-hand-side of equation (13). Thus, whereas the population $R_{(1)}^2$'s from the daily squared return-volatility regressions in equation (4) suggest that the true GARCH(1,1) model only explains between five and ten per cent of the daily variability, when measured by the more appropriate statistic

$$R_{(1)\infty}^2 \equiv 1 - E\left[\left(\sigma_{(1),t+1}^2 - \int_0^1 \sigma_{t+\tau}^2 \, d\tau\right)^2\right] \cdot \mathrm{Var}\left(\int_0^1 \sigma_{t+\tau}^2 \, d\tau\right)^{-1}, \tag{14}$$

[18] The numbers reported here, and throughout the remainder of the paper, are based on numerical simulations of the continuous-time model in equations (6) and (8) using a standard Euler discretization scheme; i.e., $p_{t+\Delta} = p_t + \sigma_t \cdot \Delta^{1/2} \cdot w_{p,t}$ and $\sigma_{t+\Delta}^2 = \theta \cdot \omega \cdot \Delta + \sigma_t^2 \cdot (1 - \theta \cdot \Delta + [2 \cdot \lambda \cdot \theta \cdot \Delta]^{1/2} \cdot w_{\sigma,t})$, where $w_{p,t}$ and $w_{\sigma,t}$ denote independent standard normal variables. In the actual implementation we took $\Delta = 1/2{,}880$, corresponding to 10 observations per five-minute interval, while the $N(0,1)$ random variables were generated by the RNDNS routine in the GAUSS computer language. The sample size was fixed at 1,000,000 'daily' observations, which along with the use of antithetic variates based on $-w_{p,t}$ and $-w_{\sigma,t}$, was deemed sufficient to reduce the sampling variation beyond the reported decimal points for all relevant summary statistics; see Geweke (1995) for a recent discussion of simulation-based methods in econometrics.

both of the weak GARCH(1,1) models account for close to fifty per cent ($1 - 0.084/0.166 \approx \frac{1}{2}$ and $1 - 0.097/0.191 \approx \frac{1}{2}$) of the variance in the one-day-ahead volatility factors. These findings underscore the importance of proper ex-post evaluation criteria when assessing volatility forecasts.

Of course, the sample path realization for the volatility process is inherently unobservable, rendering the computation of the sample equivalent of the $R^2_{(1),\infty}$ statistic in equation (14) infeasible in practice. However, if the discretely sampled returns are serially uncorrelated, and the sample path for σ_t is continuous, it follows by the theory of quadratic variation (see Karatzas and Shreve 1988), that,

$$\operatorname{plim}_{m\to\infty}\left(\int_0^1 \sigma^2_{t+\tau}\,d\tau - \sum_{j=1,\ldots,m} r^2_{(m),t+j/m}\right) = 0.$$

This result is noteworthy because it shows that the daily volatility factor, in principle, is observable from the sample path realization of the returns process. In reality, because of discontinuities in the price process and a plethora of market microstructure effects, we do not obtain a continuous reading from a diffusion process, so the limiting result cannot apply literally. Nonetheless, it suggests that the cumulative sum of squared intraday returns may greatly improve the ex-post volatility measurement, in turn resulting in more meaningful volatility forecast evaluations.

To illustrate the potential benefits from the use of the high-frequency data, consider again the measurement errors for the two continuous-time diffusions, or $E([\int_0^1 \sigma^2_{t+\tau}\,d\tau - \Sigma_{j=1,\ldots,m} r^2_{(m),t+j/m}]^2)$, reported in Table 3. As previously noted, with daily returns, or $m = 1$, the measurement errors equal 1.138 and 0.842. Increasing the sampling frequency to eight hours, or $m = 3$, lowers the measurement errors to 0.381 and 0.289. Further reducing the return interval to one hour, or $m = 24$, yield 0.048 and 0.036. For five-minute returns, or $m = 288$, the measurement

TABLE 3
IMPLIED CONTINUOUS TIME GARCH(1,1): VOLATILITY MEASUREMENT ERRORS*

m	DM-$	¥-$
1	1.138	0.842
3	0.381	0.289
24	0.048	0.036
288	0.004	0.003

*Reports the measurement errors using the sum of squared intraday returns as a measure for true daily latent volatility; i.e., $E([\int_0^1 \sigma^2_{t+\tau}\,d\tau - \Sigma_{j=1,\ldots,m} r^2_{(m),t+j/m}]^2)$. The returns are generated by the stochastic volatility model in equations (6) and (8) with the parameter values in Table 2. The aggregation frequencies $m = 1, 3, 24, 288$, correspond to daily, 8-hours, hourly, and 5-minute returns, respectively. All figures are based on simulations using antithetic variates and 1,000,000 'daily' observations.

errors of 0.004 and 0.003, are both less than 2.5 per cent of the daily variability in the latent volatility factor.[19,20]

Motivated by these findings, consider the one-day-ahead squared return–volatility regression obtained by replacing the squared daily returns on the left-hand-side of equation (4) with the sum of the corresponding squared intraday returns,

$$\sum_{j=1,\ldots,m} r^2_{(m),t+j/m} = a_{(1)m} + b_{(1)m} \cdot \sigma^2_{(1)1,t+1} + u_{(1)m,t+1}, \tag{15}$$

where $t = 0, 1, \ldots,$ and by definition $a_{(1)1} \equiv a_{(1)}$, $b_{(1)1} \equiv b_{(1)}$, and $u_{(1)1,t+1} \equiv u_{(1),t+1}$. Regardless of the sampling frequency m, if the conditional variance is correctly specified, the population values of $a_{(1)m}$ and $b_{(1)m}$ equal zero and unity, respectively.[21] However, the improved volatility measurement afforded by the intraday returns allows for more meaningful qualitative assessments of the daily forecasts, $\sigma^2_{(1)1,t+1}$, when judged by the resulting coefficient of multiple determination, $R^2_{(1)m}$.

The numerical results for the two continuous-time diffusions are reported in Table 4. The $R^2_{(1)1}$'s confirm that the weak GARCH(1, 1) forecasts explain little of the ex-post variability.[22] Meanwhile, the population $R^2_{(1)m}$'s increase monotonically with sampling frequency towards the much larger $R^2_{(1)\infty}$'s defined in equation (14). For instance, using the cumulative hourly squared returns on the left-hand-side of equation (15), the $R^2_{(1)24}$'s equal 0.383 and 0.419. Going to five-minute returns result

[19] Note that the measurement errors are almost perfectly inversely related to m. Hence, the findings effectively extend the theoretical developments in Merton (1980), which show that the variance of the sample variance of a homoskedastic diffusion is inversely related to the sampling frequency, whereas the accuracy of the estimate for the drift in the logarithmic price process only depends on the span of the data. A similar idea for more efficiently estimating the daily volatility of a homoskedastic diffusion allowing for measurement noise in the observed high-frequency price process has been explored by Zhou (1996). The results in Table 3 may also be seen as a practical guide to the applicability of the continuous-record asymptotics for rolling regressions formally developed by Foster and Nelson (1996).

[20] While high-frequency intradaily data have only recently become readily available, intraday high-low prices (the intraday range) have long been recorded daily for some equity markets. Given the availability of these statistics Garman and Klass (1980), Parkinson (1980), Ball and Torous (1984), and Kunitomo (1992), among others argued for the use of the intraday range in order to develop more accurate daily volatility estimates for homoskedastic diffusions. The properties of extreme value estimators in continuous-time models allowing for jumps are analyzed by Rogers and Satchell (1991) and Maheswaran (1996). Meanwhile, the autocorrelations in Fung and Hsieh (1991) and the time-series models estimated in Hsieh (1993) show that the intraday range is strongly serially correlated. Although the high-low range is not an unbiased estimator for the latent volatility over the day, it follows by numerical simulation that the MSE for the correspondingly scaled unbiased estimator, $E[(\gamma \cdot [\max_{0 \le \tau \le 1} p_{t+\tau} - \min_{0 \le \tau \le 1} p_{t+\tau}] - \int_0^1 \sigma^2_{t+\tau}\, d\tau)^2]$, equal 0.103 and 0.114 for the two stochastic volatility models in Table 2. Thus, compared to the measurement errors reported in Table 3, this puts the accuracy of the high-low estimator around that afforded by the intraday sample variance based on two- or three-hour returns.

[21] The same errors-in-variables problem that plagues the estimation of $b_{(m)}$ in equation (4) will result in a downward bias in the estimate for $b_{(1)m}$ formally given by $-\mathrm{Var}_{(m)}(\nu_{(1),t}) \cdot [\mathrm{Var}_{(m)}(\nu_{(1),t}) + \mathrm{Var}_{(m)}(\sigma^2_{(1),t})]^{-1}$.

[22] Note that the numerical values for the $R^2_{(1)1}$'s from the daily weak GARCH(1, 1) approximations in Table 4 are slightly lower than the implied $R^2_{(1)}$'s from the daily strong GARCH(1, 1) model with the same conditional variance parameters which, by equation (5), equal 0.064 and 0.096, respectively.

TABLE 4
IMPLIED CONTINUOUS TIME GARCH(1, 1): PREDICTIVE R^2'S*

m	DM-\$	¥-\$
1	0.063	0.089
3	0.151	0.198
24	0.383	0.419
288	0.483	0.488
∞	0.495	0.495

*Reports the population R^2 from the squared return–volatility regression in equation (15); $R^2_{(1)m} \equiv 1 - \mathrm{Var}_{(m)}(\Sigma_{j=1,\ldots,m} r^2_{(m),t+j/m} - \sigma^2_{(1)1,t+1}) \cdot \mathrm{Var}_{(m)}(\Sigma_{j=1,\ldots,m} r^2_{(m),t+j/m})^{-1}$. The returns are generated by the stochastic volatility model in equations (6) and (8) with the parameter values in Table 2. The daily GARCH(1, 1) forecasts, $\sigma^2_{(1)1,t+1}$, are based on equation (3) with $m \equiv 1$ and the parameter values in Table 1. The rows $m = 1, 3, 24, 288$ correspond to daily, 8-hours, hourly, and five-minute returns, respectively. The population $R^2_{(1)\infty} \equiv 1 - \mathrm{Var}(\int_0^1 \sigma^2_{t+\tau}\, d\tau - \sigma^2_{(1)1,t+1}) \cdot \mathrm{Var}(\int_0^1 \sigma^2_{t+\tau}\, d\tau)^{-1}$ is reported in the row labelled $m = \infty$. All figures are based on simulations using antithetic variates and 1,000,000 'daily' observations.

in $R^2_{(1)288}$'s of 0.483 and 0.488, both of which are extremely close to the ideal $R^2_{(1)\infty}$'s of 0.495 for each of the rates. These findings highlight the theoretical advantages associated with the use of high-frequency intraday returns in the construction of interdaily volatility forecast evaluation criteria. The next section explores whether these advantages actually manifest themselves in empirical work.

5. INTRADAY RETURNS AND INTERDAILY VOLATILITY FORECAST EVALUATION

The computation of daily return variances from high-frequency intraday returns parallels the use of daily returns in calculating monthly ex-post volatility, as exemplified by Schwert (1989, 1990a) and Schwert and Seguin (1990). This idea has previously been applied by, among others, Hsieh (1991) and Schwert (1990b) in measuring daily equity market volatility from the sample standard deviations of intraday returns, while Fung and Hsieh (1991) analyze daily sample standard deviations for bonds and currencies. The estimation of standard time series models for these ex-post volatility measures tend to confirm the very high degree of intertemporal volatility dependencies documented in the ARCH literature. However, the connection between volatility modeling and forecasting on the one hand and the ex-post volatility measurements on the other has hitherto not been formally explored.

5.1. *Improved Daily Volatility Forecast Evaluation.* Direct interpretation of the low $R^2_{(1)}$'s for the one-day-ahead GARCH(1, 1) DM-\$ and ¥-\$ volatility forecasts suggests that the models perform poorly, explaining less than five per cent of the ex-post variability in either rate. However, increasing the sampling frequency of the ex-post squared returns on the left-hand-side of equation (15) dramatically modifies this conclusion. For instance, with hourly sampling the $R^2_{(1)24}$'s reported in Table 5 equal 0.331 and 0.237. Further increasing the sampling frequency results in still

TABLE 5
DAILY GARCH(1,1) PREDICTIVE R^2'S*

m	DM-$	¥-$
1	0.047	0.025
3	0.133	0.095
24	0.331	0.237
288	0.479	0.392

*Reports the $R^2_{(1)m}$ from the squared–volatility regression in equation (15). The returns are calculated from continuously recorded quotations for the Deutschemark–U.S. Dollar and Japanese Yen–U.S. Dollar spot exchange rates from October 1, 1992, through September 29, 1993. Quotes from Friday 21:00 G.M.T. through Sunday 21:00 G.M.T. are excluded, resulting in a total of 260 weekdays, or 74,880 observations for the five-minute return interval. The daily GARCH(1,1) volatility forecasts, $\sigma^2_{(1)1,t+1}$, are based on equation (3) with $m \equiv 1$ evaluated at the parameters in Table 1. The rows labelled $m = 1, 3, 24, 288$ give the results for daily, 8-hours, hourly, and five-minute sampling frequencies, respectively.

higher correlations, with $R^2_{(1)288}$'s at the five-minute level of 0.479 and 0.392. These latter statistics signify more than a tenfold increase in the explanatory power of the GARCH(1,1) models relative to the conventional $R^2_{(1)1}$'s reported in the existing literature.

The reduced measurement error is also apparent in Figure 2, which graphs the one-step-ahead volatility forecasts for the daily DM-$ rate, $\sigma^2_{(1)1,t+1}$, along with the ex-post volatility measures based on the five-minute squared returns, $\Sigma_{j=1,\ldots,288}\, r^2_{(288),t+j/288}$. Clearly, the cumulative five-minute squared returns correlates much more closely with the daily GARCH(1,1) predictions than do the squared daily returns in Figure 1, and except for a few isolated episodes, the one-day-ahead predictions do a remarkable job of tracking the ex-post volatility measures. Combining these findings with the drastic improvement in the stability of the intraday volatility measure as sampling frequency increases in effect endows the notion of a latent volatility factor with concrete empirical content.

A particularly noteworthy result is the close correspondence between the implied continuous-time GARCH(1,1) predictive $R^2_{(1)m}$'s in Table 4 and the actual empirical results for the DM-$ rate in Table 5. It suggests that the market microstructure rigidities and pronounced intraday volatility patterns not accommodated by the continuous-time process in equations (6) and (8) are annihilated at the daily level. Moreover, it indicates that the simple GARCH(1,1) model does a good job of characterizing the volatility clustering for the DM-$ rate over the ex-post sample period. Meanwhile, the out-of-sample $R^2_{(1)m}$'s for the ¥-$ rate in Table 5 are all slightly below the corresponding theoretical values for the ¥-$ diffusion in Table 4. However, the discrepancy between the empirical and theoretical ¥-$ results is in part attributable to a few pronounced appreciations that occurred during the out-of-sample period.[23] Eliminating the two largest ex-post volatility measures, the value of $R^2_{(1)288}$ for the ¥-$ rate increases from 0.392 to 0.456.

[23] An analysis of the economic determinants behind these large rate changes is beyond the scope of the present paper.

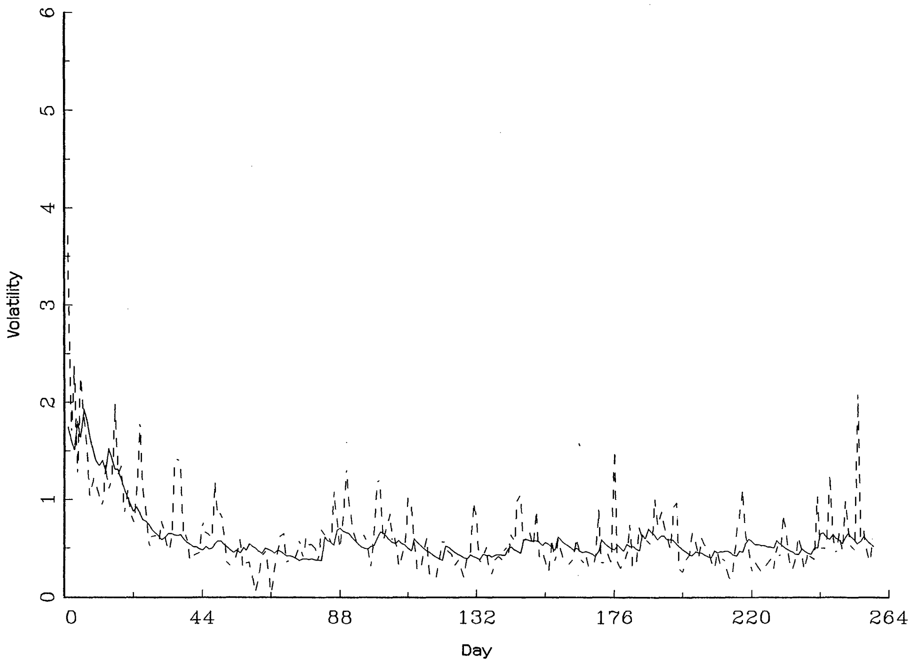

FIGURE 2

DAILY CUMULATIVE 5-MINUTE SQUARED RETURNS AND GARCH(1, 1) FORECASTS. THE SOLID LINE GRAPHS THE DAILY ONE-STEP-AHEAD GARCH(1, 1) VOLATILITY FORECASTS FOR THE DEUTSCHEMARK–U.S. DOLLAR EXCHANGE RATE GIVEN BY EQUATION (3) WITH $m \equiv 1$ AND THE PARAMETER VALUES IN TABLE 1. THE DOTTED LINE REFERS TO THE EX-POST DAILY SAMPLE VARIANCE BASED ON THE CUMULATIVE FIVE-MINUTE SQUARED RETURNS; I.E., $\sum_{j=1,\dots,288} r^2_{(288),t+j/288}$. THE SAMPLE PERIOD EXTENDS FROM OCTOBER 1, 1992 THROUGH SEPTEMBER 29, 1993.

6. CONCLUDING REMARKS

Numerous studies have suggested that ARCH and stochastic volatility models provide poor volatility forecasts. Contrary to this perception, both the theoretical and empirical analysis in this paper demonstrate that, for empirically relevant specifications, the volatility forecasts correlate closely with the future latent volatility factor that is of interest in most practical applications, typically accounting for close to fifty per cent of the variability in ex-post volatility. Yes, ARCH and stochastic volatility models do provide good volatility forecasts!

Several important questions remain. First, it is of interest to further explore the role of model misspecification. The formal conditions developed by Nelson (1992) and Nelson and Foster (1995) pertaining to the use of misspecified ARCH models in forecasting, along with the robustness results in Nelson and Foster (1994), should provide a useful guide for future work along these lines.

It would also be of immediate interest, and of great importance for current issues in financial risk management, to extend the present analysis to assess the forecasting ability of volatility models at alternative forecast horizons. However, when lengthening the forecast horizon beyond one day, issues related to the proper modeling of long-term volatility dependencies become especially important; see Baillie et al., (1996).

Our main results hinge on the effective use of frequently sampled data in constructing more accurate ex-post volatility measurements. A closely related question pertains to the precision of the volatility forecasts as a function of the sampling frequency. Do the additional costs and complications in model construction and data gathering warrant the use of intraday data for volatility forecasting as well?

The volatility forecasts analyzed above are based solely on ad-hoc time-series models. There is a voluminous literature on alternative ways in which to extract information about the latent volatility factor from sources other than, or in addition to, the corresponding squared or absolute returns. They include implied volatilities extracted from options prices, as in the recent work of Canina and Figlewski (1993), Jorion (1995), and Lamoureux and Lastrapes (1993), along with information provided by the joint distribution of return and trading volume, as in the work by Andersen (1996), and Gallant et al., (1992). The evaluation criteria proposed here should allow for more meaningful comparisons of these structural methods for estimating volatility.

Most of the volatility forecast comparisons in the literature rely on some variant of the squared return–volatility regression analyzed here. While such evaluation criteria may be natural when evaluating forecasts for the conditional mean, they are less obvious when evaluating volatility forecasts; see the discussion in Diebold and Christoffersen (1997), Diebold et al., (1998), Engle et al., (1993), West et al., (1993) and Lopez (1995). Our results suggest that further analysis along these lines may similarly benefit from the use of high-frequency data. All of these issues await future research.

REFERENCES

AKGIRAY, V., "Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts," *Journal of Business* 62 (1989), 55–80.

ANDERSEN, T.G., "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," *Mathematical Finance* 4 (1994), 75–102.

———, "Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility," *Journal of Finance* 51 (1996), 169–204.

——— AND T. BOLLERSLEV, "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance* 4 (1997a), 115–158.

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

——— AND ———, "DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer-Run Dependencies," *Journal of Finance* 53 (1998), 219–265.

BAILLIE, R.T. AND T. BOLLERSLEV, "The Message in Daily Exchange Rates: A Conditional Variance Tale," *Journal of Business and Economic Statistics* 7 (1989), 297–305.

———, ———, AND H.O. MIKKELSEN, "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 74 (1996), 3–30.

BALL, C.A. AND W.N. TOROUS, "The Maximum Likelihood Estimation of Security Price Volatility: Theory, Evidence and Applications to Option Pricing," *Journal of Business* 57 (1984), 97–112.

BOLLERSLEV, T., "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31 (1986), 307–327.

———, "A Conditional Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," *Review of Economics and Statistics* 69 (1987), 542–547.

———, R.Y. CHOU, AND K.F. KRONER, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52 (1992), 5–59.

———, R.F. ENGLE, AND D B. NELSON, "ARCH Models," in R.F. Engle and D. McFadden, eds., *Handbook of Econometrics Volume IV* (Amersterdam: North Holland Press, 1994, pp. 2959–3038).

——— AND H.O. MIKKELSEN, "Modeling and Pricing Long-Memory in Stock Market Volatility," *Journal of Econometrics* 73 (1996), 151–184.

——— AND J.M. WOOLDRIDGE, "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances," *Econometric Reviews* 11 (1992), 143–172.

BOUDOUKH, J., M. RICHARDSON, AND R.F. WHITELAW, "Investigation of a Class of Volatility Estimators," *Journal of Derivatives* 4 (1997), 63–71.

BRAILSFORD, T.J. AND R.W. FAFF, "An Evaluation of Volatility Forecasting Techniques," *Journal of Banking and Finance* 20 (1996), 419–438.

CANINA, L. AND S. FIGLEWSKI, "The Informational Content of Implied Volatility," *Review of Financial Studies* 6 (1993), 659–681.

CHOW, G.C., *Econometrics*, 2nd edn. (New York: McGraw-Hill, 1983).

CHRISTENSEN, B.J. AND N.R. PRABHALA, "The Relation Between Implied and Realized Volatility," *Journal of Financial Economics*, (1998), in press.

CHRISTOFFERSEN, P.F., Evaluating Interval Forecasts, *International Economic Review*, this issue, 841–862.

CUMBY, R., S. FIGLEWSKI, AND J. HASBROUCK, "Forecasting Volatility and Correlations with EGARCH Models," *Journal of Derivatives*, Winter (1993), 51–63.

DAY, T.E. AND C.M. LEWIS, "Stock Market Volatility and the Information Content of Stock Index Options," *Journal of Econometrics* 52 (1992), 267–287.

DIEBOLD, F.X. AND R.S. MARIANO, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), 253–263.

——— AND P.F. CHRISTOFFERSEN, "How Relevant is Volatility Forecasting for Financial Risk Management," mimeo, Department of Economics, University of Pennsylvania, 1997.

———, T.A. GUNTHER, AND A.S. TAY, "Evaluating Density Forecasts," *International Economic Review*, this issue, 863–883.

DIMSON, E. AND P. MARSH, "Volatility Forecast without Data-Snooping," *Journal of Banking and Finance* 14 (1990), 399–421.

DROST, F.C. AND T.E. NIJMAN, "Temporal Aggregation of GARCH Processes," *Econometrica* 61 (1993), 909–927.

——— AND B.J.M. WERKER, "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics* 74 (1996), 31–57.

ENGLE, R.F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica* 50 (1982), 987–1008.

——— AND T. BOLLERSLEV, "Modelling the Persistence of Conditional Variances," *Econometric Reviews* 5 (1986), 1–50.

———, C-H. HONG, A. KANE, AND J. NOH, "Arbitrage Valuation of Variance Forecasts with Simulated Options," in D.M. Chance and R.R. Trippi, eds., *Advances in Futures and Options Research* (Greenwich, Connecticut: JAI Press, 1993, pp. 393–415).

FAMA, E., "The Behavior of Stock Market Prices," *Journal of Business* 38 (1965), 34–105.

FIGLEWSKI, S., "Forecasting Volatility," *Financial Markets, Institutions and Instruments* 6 (1997), 1–88.

FOSTER, D.P. AND D.B. NELSON, "Continuous Record Asymptotics for Rolling Sample Variance Estimators," *Econometrica* 64 (1996), 139–174.

FRENNBERG, P. AND B. HANSSON, "An Evaluation of Alternative Models for Predicting Stock Volatility: Evidence from a Small Stock Market," *Journal of International Financial Markets, Institutions and Money* 5 (1996), 117–134.

FUNG, W.K.H. AND D.A. HSIEH, "Empirical Analysis of Implied Volatility: Stocks, Bonds and Currencies," mimeo, Department of Finance, Fuqua School of Business, Duke University, 1991.

GALLANT, A.R., P.E. ROSSI, AND G.E. TAUCHEN, "Stock Prices and Volume," *Review of Financial Studies* 5 (1992), 199–242.

GARMAN, M.B. AND M.J. KLASS, "On the Estimation of Security Price Volatilities from Historical Data," *Journal of Business* 53 (1980), 67–78.

GEWEKE, J., "Monte Carlo Simulation and Numerical Integration," Staff Report No. 192, Federal Reserve Bank of Minneapolis, 1995.

GHYSELS, E., A. HARVEY, AND E. RENAULT, "Stochastic Volatility," in G.S. Maddala and C.R. Rao, eds., *Handbook of Statistics Vol.* 14, *Statistical Methods in Finance*, (Amsterdam: North Holland, 1996, pp. 119–191).

GOODHART, C.A.E. AND M. O'HARA, "High Frequency Data in Financial Markets: Issues and Applications," *Journal of Empirical Finance* 4 (1997), 73–114.

HEYNEN, R.C. AND H.M. KAT, "Volatility Prediction: A Comparison of the Stochastic Volatility, GARCH(1,1), and EGARCH(1,1) Models," *Journal of Derivatives* 2 (1994), 50–65.

HSIEH, D.A., "Modeling Heteroskedasticity in Daily Foreign Exchange Rates," *Journal of Business and Economic Statistics* 7 (1989), 307–317.

———, "Chaos and Nonlinear Dynamics: Application to Financial Markets," *Journal of Finance* 46 (1991), 1839–1877.

———, "Implications of Nonlinear Dynamics for Financial Risk Management," *Journal of Financial and Quantitative Analysis* 28 (1993), 41–64.

HULL, J. AND A. WHITE, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42 (1987), 381–400.

JACQUIER, E., N.G. POLSON, AND P.E. ROSSI, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics* 12 (1994), 371–417.

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

———, "Risk and Turnover in the Foreign Exchange Market," in J.A. Frankel, G. Galli, and A. Giovannini, eds., *The Microstructure of Foreign Exchange Markets* (Chicago: The University of Chicago Press, 1996, pp. 19–37).

KARATZAS, I. AND S.E. SHREVE, *Brownian Motion and Stochastic Calculus* (New York: Springer Verlag, 1988).

KUNITOMO, N., "Improving the Parkinson Method of Estimating Security Price Volatilities," *Journal of Business* 65 (1992), 295–302.

LAMOUREUX, C.G. AND W.D. LASTRAPES, "Forecasting Stock Returns Variances: Towards Understanding Stochastic Implied Volatility," *Review of Financial Studies* 6 (1993), 293–326.

LOPEZ, J.A., "Evaluating the Predictive Accuracy of Volatility Models," mimeo, Research and Market Analysis Group, Federal Reserve Bank of New York, 1995.

MAHESWARAN, S., "Volatility Estimation Using Extreme Values of Asset Prices," mimeo, Olin School of Business, Washington University, 1996.

MANDELBROT, B.B., "The Variation of Certain Speculative Prices," *Journal of Business* 36 (1963), 394–419.

MEDDAHI, N. AND E. RENAULT, "Aggregation and Marginalization of GARCH and Stochastic Volatility Models," mimeo, GREMAQ, University of Toulouse, 1997.

MELINO, A., "The Estimation of Continuous-Time Models in Finance," in C.A. Sims, ed., *Advances in Econometrics, Sixth World Congress Volume II* (Cambridge: Cambridge University Press, 1994, pp. 313–351).

MERTON, R.C., "On Estimating the Expected Return on the Market," *Journal of Financial Economics* 8 (1980), 323–361.

MINCER, J. AND V. ZARNOWITZ, "The Evaluation of Economic Forecasts," in J. Mincer, ed., *Economic Forecasts and Expectations* (New York: National Bureau of Economic Research, 1969).

NELSON, D.B., "ARCH Models as Diffusion Approximations," *Journal of Econometrics* 45 (1990), 7–38.

———, "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics* 52 (1992), 61–90.

———, "Asymptotic Filtering Theory for Multivariate ARCH Models," *Journal of Econometrics* 71 (1996), 1–47.

——— AND D.P. FOSTER, "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica* 62 (1994), 1–41.

——— AND D.P. FOSTER, "Filtering and Forecasting with Misspecified ARCH Models II: Making the Right Forecast with the Wrong Model," *Journal of Econometrics* 67 (1995), 303–335.

PAGAN, A.R. AND H. SABAU, "Consistency Tests for Heteroskedasticity and Risk Models," *Estudios Económicos* 7 (1992), 3–30.

——— AND G.W. SCHWERT, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics* 45 (1990), 267–290.

——— AND A. ULLAH, "The Econometric Analysis of Models with Risk Terms," *Journal of Applied Econometrics* 3 (1988), 87–105.

PARKINSON, M., "Extreme Value Method for Estimating the Variance of the Rate of Return," *Journal of Business* 53 (1980), 61–65.

ROGERS, L.C.G. AND S.E. SATCHELL, "Estimating Variance from High, Low and Closing Prices," *Annals of Applied Probability* 1 (1991), 504–512.

ROLL, R., "R^2," *Journal of Finance* 43 (1988), 541–566.

ROSSI, P., *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time* (San Diego: Academic Press, 1996).

SCHWERT, G.W., "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance* 44 (1989), 1115–1153.

———, "Stock Volatility and the Crash of '87," *Review of Financial Studies* 3 (1990a), 77–102.

———, "Stock Market Volatility," *Financial Analysts Journal* 46 (1990b), 23–34.

——— AND P.J. SEGUIN, "Heteroskedasticity in Stock Returns," *Journal of Finance* 45 (1990), 1129–1155.

SCOTT, L., "Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application," *Journal of Financial and Quantitative Analysis* 22 (1987), 419–436.

SHEPHARD, N., "Statistical Aspects of ARCH and Stochastic Volatility," in D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen, eds., *Likelihood, Time Series with Econometric and Other Applications* (London: Chapman and Hall, 1996, pp. 1–67).

WEST, K.D. AND D. CHO, "The Predictive Ability of Several Models of Exchange Rate Volatility," *Journal of Econometrics* 69 (1995), 367–391.

———, H.J. EDISON, AND D. CHO, "A Utility Based Comparison of Some Models of Exchange Rate Volatility," *Journal of International Economics* 35 (1993), 23–45.

WIGGINS, M., "Option Values under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics* 19 (1987), 351–372.

ZHOU, B., "High-Frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business and Economic Statistics* 14 (1996), 45–52.