HIGH-FREQUENCY DATA, FREQUENCY DOMAIN INFERENCE, AND VOLATILITY FORECASTING

Tim Bollerslev and Jonathan H. Wright*

Abstract—Although it is clear that the volatility of asset returns is serially correlated, there is no general agreement as to the most appropriate parametric model for characterizing this temporal dependence. In this paper, we propose a simple way of modeling financial market volatility using high-frequency data. The method avoids using a tight parametric model by instead simply fitting a long autoregression to log-squared, squared, or absolute high-frequency returns. This can either be estimated by the usual time domain method, or alternatively the autoregressive coefficients can be backed out from the smoothed periodogram estimate of the spectrum of log-squared, squared, or absolute returns. We show how this approach can be used to construct volatility forecasts, which compare favorably with some leading alternatives in an out-of-sample forecasting exercise.

I. Introduction

ALTHOUGH asset returns generally appear to be close to a martingale difference sequence, there is overwhelming evidence that asset returns are not independently distributed over time because of volatility clustering. Serial correlation in the volatility of asset returns has been documented in an enormous number of papers, going back to Mandelbrot (1963) and Fama (1965), and, more recently, the ARCH literature pioneered by Engle (1982). Recent survey articles include Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994), and Diebold and Lopez (1995). Many parametric models have been proposed for modeling this persistence of volatility in asset returns. These include the ARCH and GARCH models (Engle, 1982; Bollerslev, 1986), the EGARCH model (Nelson, 1991) and stochastic volatility models (Taylor, 1986; Andersen, 1994). Some researchers have proposed models with long memory in the volatility process (including Baillie, Bollerslev, and Mikkelsen (1996), Breidt, Crato, and de Lima (1998), Harvey (1998), and Robinson (1991)). Other authors, such as Engle and Lee (1999) have proposed models in which the volatility process has two components: one of which is nearly nonstationary whereas the other is much less persistent. A number of papers have considered nonparametric approaches to representing time-varying conditional heteroskedasticity (Pagan & Schwert, 1990; Gallant, Rossi, & Tauchen, 1992, 1993). In these models, squared asset returns are modeled as a nonparametric function of lagged returns. However, in practice it is necessary to choose a relatively small number of lags, because of the well-known problems in applying nonparametric methods to high-dimensional models. When working with high-frequency data, there are also important intradaily patterns in volatility (Andersen & Bollerslev, 1997). So, although it is clear that the volatility of asset returns is highly persistent, there is no consensus as to the best model for representing these volatility dynamics.

A different approach to modeling volatility dynamics is proposed in this paper, explicitly utilizing the additional information in high-frequency data. The idea is to model volatility dynamics by fitting a long AR representation to log-squared, squared, or absolute high-frequency asset returns. This can be implemented by first estimating the spectrum of log-squared, squared, or absolute returns and then using a numerical method, sometimes known as the Wiener-Kolmogorov filter, to solve for the coefficients in the corresponding AR representation. Alternatively, the AR representation may simply be estimated by the usual time domain method. These approaches are not fully nonparametric, but they may allow for flexible dynamics. We emphasize that they are appropriate only in the presence of a very large sample of high-frequency data, so that fitting very long autoregressions is both feasible and necessary to reasonably approximate the observed patterns of volatility clustering. Indeed, these methods fare poorly with a sample of daily returns, as shall be demonstrated below. In a related context, Andersen and Bollerslev (1998) showed that intradaily data was vitally important in the meaningful ex post evaluation of daily volatility forecasts. Moreover, Andersen, Bollerslev, and Lange (1999) have recently shown that, although the gains in forecast error accuracy from correctly specified high-frequency GARCH volatility models can be very large from a theoretical perspective, the standard models tend to perform very poorly when applied directly to high-frequency data. This paper shows how the high-frequency data may easily be used to construct superior daily volatility forecasts.

The plan for the remainder of the paper is as follows. The proposed method for modeling volatility is introduced in section II. In section III, it is used to predict future values of the volatility of the Deutschemark–U.S. Dollar spot exchange rate based on a ten-year sample of five-minute returns. Section IV concludes.

II. Estimation of Volatility Dynamics

A. *The Assumed Model*

We begin by making a high-level primitive assumption that an appropriate proxy for the time series of volatilities, such as the log-squared returns, has an AR representation. Specifically, we assume that $\{y_t\}_{t=1}^{T}$ is a martingale difference sequence of asset returns such that

Received for publication December 8, 1998. Revision accepted for publication October 23, 2000.

* Duke University and National Bureau of Economic Research, and Board of Governors of the Federal Reserve System, respectively.

The first author would like to acknowledge the financial support by a grant from the NSF to the NBER. We are grateful to Jim Stock and two anonymous referees for helpful comments on an earlier draft, and to Olsen and Associates for making the high-frequency exchange rates available.

Assumption A1: $a(L)(\log(y_t^2) - \mu_1) = \epsilon_t$.

Here and throughout, $a(L) = 1 - a_1L - a_2L^2 \ldots$ denotes a (possibly infinite-order) lag polynomial such that $\sum_{j=0}^{\infty} |a_j| j^{1/2} < \infty$,[1] L is the lag operator, and ϵ_t is a serially uncorrelated white noise sequence with mean zero, variance σ^2, and finite fourth moment. As alternatives to assumption A1, we could instead specify that the squared or absolute returns have an AR representation, modified so that the squared or absolute returns cannot be negative.[2] To accommodate this, we introduce alternative assumptions A1′ and A1″:

Assumption A1′: $a(L)(z_t - \mu_2) = \epsilon_t$ and $y_t^2 = \max(z_t, 0)$.

Assumption A1″: $a(L)(z_t - \mu_3) = \epsilon_t$ and $|y_t| = \max(z_t, 0)$.

Assumption A1 nests some standard models of stochastic volatility as special cases. For example, the standard autoregressive stochastic volatility (ARSV) model specifies that

$$y_t = \exp(h_t/2)\sigma_u u_t$$

$$(1 - \phi L)h_t = \sigma_v v_t$$

where u_t and v_t are i.i.d. $N(0, 1)$ and $|\phi| < 1$. This implies that $\log(y_t^2) = h_t + \log(\sigma_u^2) + \log(u_t^2)$ which has a representation as an ARMA(1, 1) reduced form (see, for example, Harvey, Ruiz, and Shephard (1994)), and assumption A1 is satisfied.

One approach to estimating $a(L)$ is simply to fit an $AR(p)$ to the observed sample, where the order of the estimated autoregression, p, goes to infinity but at a rate slower than $T^{1/3}$, as the sample size T goes to infinity.[3] Let the resulting estimates for $a_1, a_2, \ldots$ be denoted by $\tilde{a}_1, \tilde{a}_2, \ldots$. The Wiener-Kolmogorov filter provides an alternative approach to estimating these coefficients, which may potentially work better when $a(L)$ is not in fact a small-order autoregressive polynomial.

B. The Wiener-Kolmogorov Filter

Suppose that assumption A1 holds and that the spectrum of $\log(y_t^2) - \mu_1$ is known. Call this spectrum $f(\lambda)$. Of course, the spectrum of any time series contains, in principle, the same information as its AR representation, and going from the AR representation to the spectrum is numerically straightforward, as $f(\lambda) = \frac{\sigma^2}{2\pi}|a(e^{i\lambda})|^{-2}$. Inverting this transformation is harder, but there is a standard result (see, for example, Brillinger (1981, p. 79)) that provides a closed-form representation for the autoregressive coefficients in terms of the spectrum of a univariate time series. Specifically,

$$a_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} B(\lambda)^{-1}\exp(ij\lambda)d\lambda \tag{1}$$

where

$$B(\lambda) = \exp\left(\frac{1}{2}c(0) + \sum_{v=1}^{\infty} c(v)\exp(-iv\lambda)\right) \tag{2}$$

and

$$c(v) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(f(\lambda))\exp(iv\lambda)d\lambda. \tag{3}$$

This algorithm is sometimes referred to as the Wiener-Kolmogorov filter (Bhansali & Karavellas, 1983). Naturally, in empirical applications, no researcher ever knows the true spectrum of the log-squared returns. However, the spectrum may be estimated by smoothing the periodogram, and the estimated spectrum may then be substituted into the Wiener-Kolmogorov filter. In particular, define the periodogram of the log-squared data as

$$I(\lambda) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T}(\log(y_t^2) - \hat{\mu}_1)e^{it\lambda}\right|^2, \tag{4}$$

where $\hat{\mu}_1$ denotes the sample average of $\log(y_t^2)$. The associated smoothed estimate of the spectrum is then given by

$$\hat{f}(\lambda) = \frac{1}{Th}\sum_{k=-n}^{n} K\left(\frac{\lambda - \lambda_k}{h}\right)I(\lambda_k), \tag{5}$$

where $n = [T/2]$,

$\lambda_k = \dfrac{2\pi k}{T}$,

$K(\cdot)$ is a kernel function, and

h is a bandwidth parameter which converges to zero but at a rate slower than $O(T^{-1})$.

In this paper, we use the Epanechnikov kernel which sets

$$K(\omega) = 0.75(1 - w^2)1(|\omega| \le 1). \tag{6}$$

Under the given conditions, $\hat{f}(\lambda)$ is consistent for $f(\lambda)$ uniformly on $[-\pi, \pi]$ (Brillinger, 1981).[4]

Let the resulting estimates for $a_1, a_2, \ldots$ obtained by substituting this estimated spectrum into the Wiener-Kol-

[1] $\frac{1}{2}$-summability of these coefficients is required to enable the spectrum of the log-squared data to be uniformly consistently estimable.

[2] It is of course an advantage of assumption A1 that no such modification is required.

[3] See Berk (1974).

[4] Other kernels will, of course, also guarantee the uniform consistency of $\hat{f}(\lambda)$, as discussed by Brillinger (1981).

mogorov filter in equations (1), (2), and (3) be denoted by $\hat{a}_1$, $\hat{a}_2$, ... Because each of these coefficients is a continuous function of $f(\lambda)$, it follows that, if $\hat{f}(\lambda)$ is uniformly consistent for $f(\lambda)$, then each $\hat{a}_j$ is consistent for a_j. Under assumption A1′ or A1″, exactly the same approach may be used to solve for the AR coefficients of squared returns or absolute returns from the estimated spectrum of squared or absolute returns, respectively. The approach of estimating the spectrum by smoothing the periodogram and then backing out the implied AR coefficients has previously been used by Bhansali (1973, 1974, 1977), among others. More recently, a similar idea has also been applied by Wright (1999) in the context of impulse response analysis. The present paper shows how this frequency domain approach may be applied to modeling the volatility of asset returns.[5]

C. *The Modified Log-Squared Transformation*

Assumption A1 applies to log-squared asset returns. An inlier problem often arises when dealing with the log-squared transformation; if the asset return is very close to zero, then the log-squared transformation yields a large negative number. Such an observation can then greatly affect the results of subsequent data analysis. In the extreme case, if the asset return is equal to zero, then the log-squared transformation is not even defined. Consequently, Fuller (1996) proposes a slight modification of the log-squared transformation, which does not converge to minus infinity as the argument converges to zero. This specifies that the transformed series of asset returns is

$$y_t^* = \log\,(y_t^2 + \tau s^2) - \frac{\tau s^2}{y_t^2 + \tau s^2}$$

where s^2 is the sample variance of y_t and τ is a small constant. In all empirical work in this paper, we use y_t^* with $\tau = 0.02$, instead of the log-squared returns.[6] However, for convenience, we adopt the shorthand of referring to y_t^* as the log-squared returns.

[5] Long-memory in the volatility process is not strictly speaking allowed for, as we have assumed that the coefficients in $a(L)$ are $\frac{1}{2}$-summable. Concretely, if $f(\lambda)$ is not bounded away from zero at the origin, then $\hat{f}(\lambda)$ is not uniformly consistent. For example, the fractionally integrated stochastic volatility model of Breidt, Crato, and de Lima (1998) yields an AR representation for log-squared returns, but the $\frac{1}{2}$-summability requirement is not satisfied. Nevertheless, from a practical empirical perspective, the estimated $a(L)$ coefficients may get arbitrarily close to a true long-memory process. Moreover, in a recent paper, Hidalgo and Yajima (1999) have shown how to adapt the Wiener-Kolmogorov filter to admit long memory; this involves using an alternative spectral density estimator. We have reworked the empirical results in this paper using their spectral density estimator and have found that it makes very little difference in practice.

[6] The choice of τ follows Fuller (1996) and Breidt and Carriquiry (1996).

III. Forecasting Integrated Volatility

A. *Integrated Volatility and Alternative Volatility Measures*

A leading motivation for studying models of time-varying conditional heteroskedasticity is to be able to forecast future volatility. One common measure of the quality of a forecast of an arbitrary variable, x_t, is the R^2 in a regression of the ex post realized values of x_t on its forecast values (and a constant). We refer to this procedure as the Mincer-Zarnowitz method (Mincer and Zarnowitz, 1969). The R^2 in the Mincer-Zarnowitz regression indicates that GARCH and other standard volatility models provide poor forecasts of future squared returns. (See, for example, Jorion (1995) and Andersen and Bollerslev (1998).)

However, the squared one-period return is generally a very noisy measure of the true latent volatility and is, as such, virtually unforecastable. Meanwhile, if the researcher is interested in the volatility of the return of an asset over any fixed time period from t_0 to t_1 (such as a day, or the life of an option), and if the researcher has access to high-frequency intradaily data on the asset returns, then the squared high-frequency returns, summed over the period from t_0 to t_1, constitute a much more accurate estimate of the true ex post volatility of the returns over that fixed time period.[7] We refer to this measure as the *integrated volatility.* In a related context, Andersen and Bollerslev (1998) point out that standard GARCH models provide good forecasts of future values of integrated volatility. In particular, they show that the R^2 in the Mincer-Zarnowitz regression is quite high for some standard data sets of asset returns. In addition to allowing for more meaningful ex post volatility forecast evaluation, this integrated volatility measure also corresponds directly to the theoretical notion of volatility entertained in the diffusion models proposed by Barndorff-Nielsen and Shephard (2001). This same measure also figures prominently in the stochastic volatility option pricing literature (such as Hull and White (1987)) and its formal estimation has recently been explored by Gallant, Hsu, and Tauchen (1999).

In this section, we consequently focus on forecasting integrated volatility. For concreteness, and to tie in with the data set analyzed later, let y_t denote a five-minute return series. With 24-hour markets, there are 288 five-minute observations in a day. It is also convenient to use the notation $y_{s,n}$ to refer to the nth five-minute return on day s, $s = 1, \ldots S$, $n = 1, \ldots 288$. Clearly, $y_t = y_{s,n}$ with $t = 288(s - 1) + n$. The integrated volatility over the day s is then defined as $V_I(s) = \frac{1}{288} \sum_{n=1}^{288} y_{s,n}^2$.

[7] As discussed more formally in Andersen et al. (2001), if the returns follow a special semimartingale, the quadratic variation of the process constitutes a natural measure of the ex post realized volatility.

B. The Data

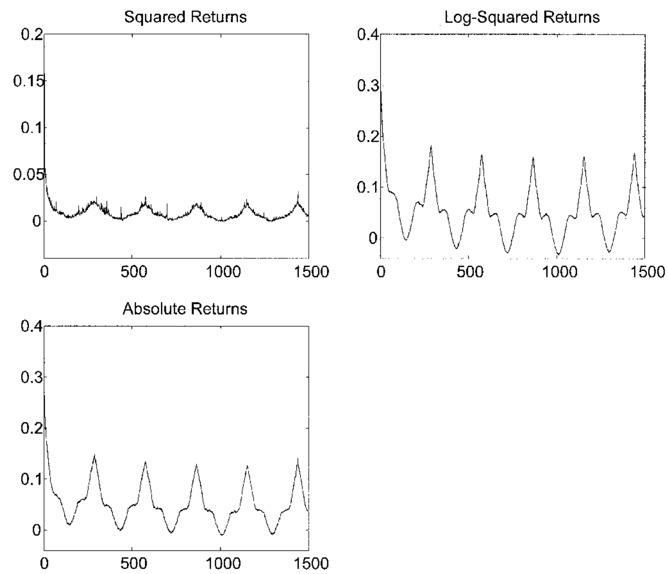
The spot Deutschemark–U.S. Dollar exchange rate data was collected and provided by Olsen and Associates in Zürich, Switzerland. The full sample spans the period from December 1, 1986 through December 1, 1996. The returns are calculated as the difference between the linearly interpolated average of the midpoint of the logarithmic bid and ask for the two nearest quotes, resulting in a total of 288 five-minute return observations per day.[8] Although the foreign exchange market is officially open 24 hours a day and 365 days a year, the trading activity slows decidedly during the weekend period. To avoid confounding the evidence by such weekend patterns, we simply excluded all returns from Friday 21:00 Greenwich Mean Time (GMT) through Sunday 21:00 GMT; a similar weekend no-trade convention was adopted by Andersen and Bollerslev (1997). Furthermore, the market slows decidedly over certain holiday periods. Excluding the most important of these days[9] leaves us with a sample of 2,445 complete days, for a total of $T = 2{,}445 \times 288 = 704{,}160$ five-minute y_t return observations.

Consistent with the notion of efficient markets, the five-minute returns are approximately mean zero and serially uncorrelated. At the same time, the evidence for volatility clustering is overwhelming. For instance, the lag-1 sample autocorrelation coefficient for the squared five-minute returns equals 0.195, which is overwhelmingly significant at any level. The autocorrelograms of the squared, log-squared, and absolute returns, in figure 1, all show a rapid initial decay but then decay only slowly. Additionally, these autocorrelograms have a distinct seasonal pattern. Similar periodic autocorrelograms for other speculative returns and time periods have previously been reported in the literature by Dacorogna et al. (1993) and Andersen and Bollerslev (1997), among others, who attribute the periodicities to the existence of strong intradaily volatility patterns associated with the opening and closing of the various financial centers around the world.

C. Volatility Forecasts with Daily Data

Arguably, the most common approach to volatility forecasting is based on the estimation of daily GARCH models. Specifically, let $y_s^{(D)} = \sum_{n=1}^{288} y_{s,n}$ denote the daily return for day s. The GARCH(p, q) model then specifies that

FIGURE 1.—AUTOCORRELOGRAM OF VOLATILITY MEASURES FOR DM-DOLLAR RETURNS



This figure gives the autocorrelogram of the squared, log-squared, and absolute five-minute demeaned DM–$ returns, as described in the text.

$$y_s^{(D)} = \sigma_s \eta_s$$

$$\sigma_s^2 = \omega + \sum_{i=1}^{p} \alpha_i \sigma_{s-i}^2 \eta_{s-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{s-j}^2,$$

where η_s is serially uncorrelated with mean zero and variance one,

$\omega > 0$,

$\alpha_i \geq 0$,

$\beta_j \geq 0$, and

the parameters satisfy the conditions in Nelson and Cao (1992) for σ_s^2 to be positive (almost surely).

The quasi-maximum likelihood estimates of these parameters, obtained under the auxiliary assumption of conditional normality, may be calculated, and the estimated value of σ_s^2 can be viewed as a forecast of $V_I(s)$.[10] This forecast can then be evaluated in terms of its mean bias and mean square prediction error, or in terms of the R^2, coefficient estimates and their standard errors in the Mincer-Zarnowitz regression of $V_I(s)$ on a constant and σ_s^2.

Table 1 shows the resulting quasi-maximum likelihood estimates of the parameters of GARCH(1, 1), GARCH(1, 2), GARCH(2, 1), and GARCH(2, 2) models, estimated using the full sample of 2,445 days.[11] Table 2 shows the out-of-sample forecast evaluation criteria for these specifi-

[8] See Dacorogna et al. (1993) and Müller et al. (1990) for a more detailed description of the activity patterns in the foreign exchange market and the method of data capturing and filtering that underlie the return calculations.

[9] For further discussion of the specific exclusions, we refer to Andersen et al. (2001), in which the same data is analyzed from a different perspective.

[10] As discussed in Andersen et al. (2001), if the returns follow a special semimartingale, then $\sigma_s^2 = E_{s-1}(y_s^{(D)2}) = E_{s-1}(V_I(s))$, but $y_s^{(D)2}$ is a much more noisy measure of this expectation than $V_I(s)$; see also Andersen and Bollerslev (1998).

[11] Among all GARCH(p, q) models for daily data with p, $q \leq 3$, the Akaike and Schwartz information criteria both selected the GARCH(2, 1) specification.

TABLE 1.—QUASI-MAXIMUM LIKELIHOOD GARCH PARAMETER ESTIMATES WITH DAILY DATA (STANDARD ERRORS IN PARENTHESIS)

Parameter	GARCH(1, 1)	GARCH(1, 2)	GARCH(2, 1)	GARCH(2, 2)
ω	0.0069	0.0049	0.0072	0.0125
	(0.0017)	(0.0021)	(0.0017)	(0.0034)
α_1	0.0449	0.0313	0.0313	0.0472
	(0.0054)	(0.0119)	(0.0158)	(0.0053)
α_2			0.0151	0.0424
			(0.0165)	(0.0054)
β_1	0.9414	1.2819	0.9395	−0.0564
	(0.0075)	(0.2709)	(0.0077)	(0.0075)
β_2		−0.3229		0.9423
		(0.2560)		(0.0072)
AIC	2.0579	2.0584	2.0585	2.0561
SIC	2.0650	2.0679	2.0680	2.0679

This table reports the quasi-maximum likelihood estimates of the GARCH parameters using the full 2,445 days of daily data, along with the associated standard errors. The Akaike and Schwartz information criteria (AIC and SIC) are also included.

TABLE 2.—OUT-OF-SAMPLE PROPERTIES OF ALTERNATIVE ESTIMATES OF $V_I(s)$ USING DAILY DATA

Forecasting Method	$\hat{b}_0$	$\hat{b}_1$	Bias	MSPE	R^2
GARCH(1, 1)	−0.322	1.758	0.050	0.180	39.5
	(0.033)	(0.062)			
GARCH(1, 2)	−0.145	1.379	0.044	0.197†	27.4
	(0.034)	(0.064)			
GARCH(2, 1)	−0.249	1.606	0.050	0.189†	33.3
	(0.034)	(0.065)			
GARCH(2, 2)	−0.013	1.125	0.048	0.188	29.1
	(0.028)	(0.050)			
EGARCH(1, 1)	−0.219	1.571	0.058	0.190†	32.7
	(0.033)	(0.064)			
AR(10) fitted to $y_t^{(D)2}$	−0.305	1.691	0.041	0.218†	20.7
	(0.049)	(0.095)			

The table reports the estimated intercept and slope coefficients from the Mincer-Zarnowitz regression, $\hat{b}_0$ and $\hat{b}_1$, respectively, along with their estimated standard errors (in parentheses) and the percentage R^2 from this regression, for all of the alternative forecasting procedures. The table also reports the bias (sample mean of the forecast minus the ex post realized value) and mean square prediction error (MSPE) for each of these forecasts. The superscript * indicates that the mean square prediction error is significantly lower than that of the GARCH(1, 1) forecast, and the superscript † indicates that it is significantly higher (at the 5% level, using the test of Diebold and Mariano (1995)). The details of the construction of these alternative volatility forecasts are provided in the text. The GARCH and EGARCH models are all estimated using the daily data $y_t^{(D)}$. All parameters are estimated using the first 1,222 days of data, and the forecasts are evaluated using the remaining 1,223 days in the sample.

cations, estimating the model parameters with the first half of the data and evaluating the forecasts over the remainder of the sample. In table 2, we also report the forecast evaluation criteria for an EGARCH(1, 1) model and for forecasts constructed by simply fitting an AR(10) to daily squared returns. For each forecast in table 2, we tested the significance of the difference in the mean square prediction error between that forecast and the forecast obtained from the GARCH(1, 1) model using the procedure described by Diebold and Mariano (1995).

In table 2, we can see that the GARCH(1, 1) model provides the forecast with the lowest mean square prediction error and the highest R^2 in the Mincer-Zarnowitz regression.[12] Most of the forecasts in table 2 do significantly worse than the GARCH(1, 1) model, in terms of mean square error. Simply fitting an AR(10) to the daily squared returns produces forecasts with quite low bias but very high variance, so that these forecasts do poorly in terms of mean square error. (The forecasts using autoregressions of other orders do worse again.) Overall, we conclude that the simple GARCH(1, 1) model forecasts best, when working with daily data.

D. *Volatility Forecasts using the Intradaily Data*

We now turn to volatility forecasts, explicitly constructed from the five-minute returns. One forecasting strategy is to start with assumption A1′, estimate the spectrum of $v_t = y_t^2 - \hat{\mu}_2$ (where $\hat{\mu}_2$ denotes the sample mean of y_t^2), and then use the Wiener-Kolmogorov filter to calculate $\{\hat{a}_j\}$, the associated estimates of $\{a_j\}$. Let $\hat{v}_{t+k|t}$ denote the resulting forecast of v_{t+k} given v_t and lagged values, as given by the recursions

$$\hat{v}_{t+1|t} = \sum_{j=1}^{\infty} \hat{a}_j v_{t-j} \tag{7}$$

and

$$\hat{v}_{t+k|t} = \sum_{j=1}^{k-1} \hat{a}_j \hat{v}_{t+k-j|t} + \sum_{j=k}^{\infty} \hat{a}_j v_{t+k-j}. \tag{8}$$

The integrated volatility at date s, $V_I(s)$, conditional on returns from day $s - 1$ and earlier may then be forecast[13] as

$$\hat{V}_I(s) = \frac{1}{288} \sum_{n=1}^{288} \max\,(\hat{\mu}_2 + \hat{v}_{288(s-1)+n|288(s-1)}, 0). \tag{9}$$

Forecasts of $V_I(s)$ may equally be formed using $\{\tilde{a}_j\}$, the estimates of the AR coefficients obtained from the usual time domain autoregression. Let this forecast be denoted $\tilde{V}_I(s)$.

Table 3 shows the forecast evaluation criteria for $\hat{V}_I(s)$ and $\tilde{V}_I(s)$ in the out-of-sample forecasting exercise. (As in table 2, the models were estimated using the first half of the data, whereas the forecasts were evaluated using the second half of the data.)[14] In constructing these forecasts, the spectrum of the squared returns was estimated using a bandwidth $h = 0.0009$, and the fitted time domain autoregression was of order 2,050. These parameters were chosen so as to minimize the out-of-sample mean square prediction error. This leaves open the question of how a researcher should select these parameters in practice, but it ensures that we can make a fair comparison between $\hat{V}_I(s)$ and $\tilde{V}_I(s)$.

[12] The GARCH(1, 1) forecasts of integrated volatility have an R^2 in the Mincer-Zarnowitz regression similar to that found by Andersen and Bollerslev (1998).

[13] In the empirical work, we truncated the infinite sum in equation (7) after 10,000 terms, although the results are virtually identical using just 5,000 terms.

[14] Counterparts of tables 2 and 3 for in-sample forecasts and for the Japanese Yen–U.S. Dollar exchange rate are similar, and are available on request.

TABLE 3.—OUT-OF-SAMPLE PROPERTIES OF ALTERNATIVE ESTIMATES OF $V_I(s)$ USING FIVE-MINUTE DATA

Forecasting Method	$\hat{b}_0$	$\hat{b}_1$	Bias	MSPE	R^2
$\tilde{V}_I(s)$	−0.131	1.257	0.007	0.145*	46.5
(time domain)	(0.023)	(0.039)			
$\tilde{V}_I^{LOG-SQ}(s)$	−0.158	0.954	−0.192	0.179	45.7
(time domain)	(0.024)	(0.030)			
$\tilde{V}_I^{ABS}(s)$	−0.042	0.900	−0.107	0.148*	48.1
(time domain)	(0.020)	(0.027)			
$\hat{V}_I(s)$	−0.182	1.356	0.008	0.146*	47.1
(frequency domain)	(0.024)	(0.041)			
$\hat{V}_I^{LOG-SQ}(s)$	−0.216	1.059	−0.174	0.171	46.3
(frequency domain)	(0.026)	(0.033)			
$\hat{V}_I^{ABS}(s)$	−0.066	0.955	−0.095	0.144*	48.5
(frequency domain)	(0.021)	(0.028)			
AR(10) fitted to $V_I(s)$	−0.122	1.241	0.007	0.156*	45.9
	(0.023)	(0.039)			

See notes for table 2.

Assumption A1′ is the natural starting point for forecasting integrated volatility as it relates directly to the squared returns. Alternatively, assumption A1 or A1″ can be used to justify the corresponding forecasts of the future log-squared or absolute five-minute returns. These forecasts may then in turn be transformed into ad hoc forecasts of the future squared five-minute returns by exponentiating or squaring the forecasts, followed by multiplication by a scaling factor.[15] In the case of assumption A1, the scaling factor is the ratio of the mean of the squared returns to the exponent of the mean of the log-squared returns. In the case of assumption A1″, it is the ratio of the mean of the squared returns to the square of the mean of the absolute returns. The forecasts of future squared five-minute returns can be summed up to obtain forecasts of future integrated volatility, based on assumptions A1 or A1″. These forecasts are ad hoc, but their practical usefulness is an empirical question. Of course, these forecasts could alternatively be based on fitting a long autoregression to the log-squared or the absolute returns in the time domain. Table 3 also shows the results for these forecasts, which we refer to (in obvious notation) as $\hat{V}_I^{LOG-SQ}(s)$, $\tilde{V}_I^{LOG-SQ}(s)$, $\hat{V}_I^{ABS}(s)$, and $\tilde{V}_I^{ABS}(s)$.

Table 3 also shows the results for the forecasts of daily integrated volatility obtained by fitting an AR(10) model directly to the daily integrated volatility, $V_I(s)$. This is still a forecast based on high-frequency data in the sense that it cannot be constructed by a researcher who has access to only daily data. As a final comparison, we also used GARCH models fitted directly to the high-frequency, five-minute returns to construct the daily volatility forecasts. However, to conserve space, we omit these results, as they yielded very unreasonable forecasts. Consistent with the earlier findings in Andersen, Bollerslev, and Lange (1999), a small-order GARCH model is grossly underparameterized as a model of intradaily asset returns.[16] For each forecast in table 3, we tested the significance of the difference in the mean square prediction error between that forecast and the forecast obtained from the commonly employed daily GARCH(1, 1) model using the procedure described by Diebold and Mariano (1995), as we did in table 2.

The forecast evaluation criteria for $\hat{V}_I(s)$ and $\tilde{V}_I(s)$ shown in table 3 are very similar, although the frequency domain forecasts give a slightly higher R^2 in the Mincer-Zarnowitz regressions. Meanwhile, both work much better than any of the forecasts in table 2, based on daily data alone (in terms of bias, mean square error, and the fit of the Mincer-Zarnowitz regression). The out-of-sample mean square error of $\hat{V}_I(s)$ is 19% below that of the standard GARCH(1, 1) forecast; this improvement in forecasting performance is highly significant. Both $\hat{V}_I(s)$ and $\tilde{V}_I(s)$ are virtually unbiased forecasts, unlike any of the forecasts in table 2.

The forecasts based on fitting autoregressions to log-squared and absolute high-frequency returns have considerable bias, which is not surprising in view of their ad hoc justification. Among these forecasts, the frequency domain forecasts consistently have a slight edge over the time domain forecasts (in terms of bias, mean square error, and the fit of the Mincer-Zarnowitz regression). Both $\hat{V}_I^{LOG-SQ}(s)$ and $\tilde{V}_I^{LOG-SQ}(s)$ have higher mean square error than $\hat{V}_I(s)$ and $\tilde{V}_I(s)$, whereas $\hat{V}_I^{ABS}(s)$ and $\tilde{V}_I^{ABS}(s)$ both have lower mean square error than $\hat{V}_I(s)$ and $\tilde{V}_I(s)$. The mean square error of $\hat{V}_I^{ABS}(s)$, the best of all the forecasts, is 20% below that of the standard daily GARCH(1, 1) forecast.[17] Simply fitting an autoregression to $V_I(s)$ produces forecasts that are clearly superior to any of the forecasts in table 2, but yields a higher mean square error and a lower R^2 in the Mincer-Zarnowitz regression than most of the other forecasts using intradaily data in table 3. It is crucial to volatility forecasting to have a good estimate of current conditional volatility, obtained from high-frequency data, but the additional flexibility of a full high-frequency model can be of some extra help.

IV. Concluding Remarks

In this paper, we have proposed modeling volatility dynamics with high-frequency data by simply fitting an autoregression to log-squared, squared, or absolute returns. This autoregression can be estimated in the usual way, or it can be backed out from a nonparametric smoothed periodogram estimate of the spectrum. We conclude that, when working with high-frequency intradaily data, these simple autoregressions tend to work better in forecasting future volatility than standard GARCH and EGARCH models,

[15] The scaling factor is a simple attempt to correct for the fact that the expectation of a nonlinear function of a random variable is not equal to the nonlinear function of the expectation of that random variable.

[16] For instance, fitting a GARCH(1, 1) model to five-minute returns, the sum of the GARCH coefficients was 1.04, and the in-sample mean square prediction error of the associated forecasts was 2.4×10^6!

[17] We attribute the good performance of the forecasts based on five-minute absolute returns to the fact that absolute returns are relatively outlier resistant.

fitted either to daily or intradaily data. Overall, this general conclusion is not very sensitive to whether the autoregressions are estimated in the time domain or in the frequency domain, although the latter procedure results in the lowest mean square prediction error. In sum, Andersen and Bollerslev (1998) showed that intradaily data were vitally important in the meaningful ex post evaluation of daily volatility forecasts; this paper shows how the high-frequency data may easily be used to construct superior daily volatility forecasts. Meanwhile, the approach to modeling volatility dynamics advocated here can of course be used in applications other than volatility forecasting; for example, it could be used to construct bootstrap distributions for test statistics. It will be interesting to further explore these issues in future research.

REFERENCES

Andersen, T. G., "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling," *Mathematical Finance* 4:2 (1999), 75–102.

Andersen, T. G., and T. Bollerslev, "Intraday Periodicity and Volatility Persistence in Financial Markets," *Journal of Empirical Finance* 4:2 (1997), 115–158.

——— "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review* 39:4 (1998), 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association* (2001 (forthcoming)).

Andersen, T. G., T. Bollerslev, and S. Lange, "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance* 6 (December 1999), 457–477.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen, "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 74:1 (1996), 3–30.

Barndorff-Nielsen, O. E., and N. Shephard, "Non-Gaussian Ornstein-Uhlenbeck Models and their Uses in Financial Economics," *Journal of the Royal Statistical Society, Series B* (2001 (forthcoming)).

Berk, K. N., "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2:3 (1974), 489–502.

Bhansali, R. J., "A Monte Carlo Comparison of the Regression Method and the Spectral Methods of Prediction," *Journal of the American Statistical Association* 68:343 (1973), 621–625.

——— "Asymptotic Properties of the Wiener-Kolmogorov Predictor I," *Journal of the Royal Statistical Society,* Series B 36 (1974), 61–73.

——— "Asymptotic Properties of the Wiener-Kolmogorov Predictor II," *Journal of the Royal Statistical Society,* Series B 39 (1977), 66–72.

Bhansali, R. J., and D. Karavellas, "Wiener Filtering (With Emphasis on Frequency-Domain Approaches)" in D. R. Brillinger and P. R. Krishnaiah (Eds.), *Handbook of Statistics,* vol. 3 (Amsterdam: Elsevier, 1983).

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

Bollerslev, T., R. Y. Chou, and K. F. Kroner, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52:1 (1992), 5–59.

Bollerslev, T., R. F. Engle, and D. B. Nelson, "ARCH Models" (pp. 2959–3038) in R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics,* vol. 4 (Amsterdam: Elsevier, 1994).

Breidt, F. J., N. Crato, and P. de Lima, "The Detection and Estimation of Long-Memory in Stochastic Volatility," *Journal of Econometrics* 83:1/2 (1998), 325–348.

Breidt, F. J., and A. L. Carriquiry, "Improved Quasi-Maximum Likelihood Estimators for Stochastic Volatility Models" in A. Zellner and J. S. Lee (Eds.), *Modeling and Prediction: Honoring Seymour Geisel* (New York: Springer-Verlag, 1996).

Brillinger, D. R., *Time Series: Data Analysis and Theory* (San Francisco: Holden-Day, 1981).

Dacorogna, M. M., U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, "A Geographical Model for the Daily and Weekly Seasonal Volatility in the Foreign Exchange Market," *Journal of International Money and Finance* 12:4 (1993), 413–438.

Diebold, F. X., and J. A. Lopez, "Modeling Volatility Dynamics" (pp. 427–472), in K. Hoover (Ed.), *Macroeconometrics: Developments, Tensions and Prospects* (Boston: Kluwer Academic Press, 1995).

Diebold, F. X., and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13:3 (1995), 253–263.

Engle, R. F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica* 50:4 (1982), 987–1008.

Engle, R. F., and G. Lee, "A Permanent and Transitory Component Model of Stock Return Volatility" (pp. 475–497), in R. F. Engle and H. White (Eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honor of Clive W. Granger* (Oxford: Oxford University Press, 1999).

Fama, E., "The Behavior of Stock Prices," *Journal of Business* 38:1 (1965), 34–105.

Fuller, W. A., *Introduction to Statistical Time Series* (New York: Wiley, 1996).

Gallant, A. R., P. E. Rossi, and G. E. Tauchen, "Stock Prices and Volume," *Review of Financial Studies* 5:2 (1992), 199–242.

——— "Nonlinear Dynamic Structures," *Econometrica* 61:4 (1993), 871–907.

Gallant, A. R., C. T. Hsu, and G. E. Tauchen, "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," this REVIEW 81:4 (1999), 617–631.

Harvey, A. C., "Long-Memory in Stochastic Volatility" (pp. 307–320), in J. Knight and S. Satchell (Eds.), *Forecasting Volatility in the Financial Markets* (London: Butterworth-Heineman, 1998).

Harvey, A. C., E. Ruiz, and N. G. Shephard, "Multivariate Stochastic Variance Models," *Review of Economic Studies* 61 (1994), 247–264.

Hidalgo, J., and Y. Yajima, "Prediction of Strongly Dependent Processes in the Frequency Domain with Application to Signal Extraction," University of London manuscript (1999).

Hull, J., and A. White, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42:2 (1987), 381–400.

Jorion, P., "Predicting Volatility in the Foreign Exchange Market," *Journal of Finance* 50:2 (1995), 507–528.

Mandelbrot, B. B., "The Variation of Certain Speculative Prices," *Journal of Business* 36:4 (1963), 394–419.

Mincer, J., and V. Zarnowitz, "The Evaluation of Economic Forecasts" (pp. 3–46), in J. Mincer (Ed.), *Economic Forecasts and Expectations* (Cambridge: National Bureau of Economic Research, 1969).

Müller, U. A., M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, and C. Morgenegg, "Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis," *Journal of Banking and Finance* 14 (1990), 1189–1208.

Nelson, D. B., "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59 (1991), 347–370.

Nelson, D. B., and C. Q. Cao, "Inequality Constraints in the Univariate GARCH Model," *Journal of Business and Economic Statistics* 10:2 (1992), 229–235.

Pagan, A. R., and G. W. Schwert, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics* 45 (July 1990), 267–290.

Robinson, P. M., "Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression," *Journal of Econometrics* 47 (1991), 67–84.

Taylor, S. J., *Modeling Financial Time Series* (Chichester: Wiley, 1986).

Wright, J. H., "Frequency Domain Inference for Univariate Impulse Responses," *Economics Letters* 63:3 (1999), 269–277.