INTERNATIONAL ECONOMIC REVIEW
Vol. 45, No. 4, November 2004

ANALYTICAL EVALUATION OF VOLATILITY FORECASTS*

BY TORBEN G. ANDERSEN, TIM BOLLERSLEV, AND NOUR MEDDAHI[1]

Northwestern University and National Bureau of Economic Research; Duke University and National Bureau of Economic Research; Université de Montréal (CIRANO, CIREQ), Canada

Estimation and forecasting for realistic continuous-time stochastic volatility models is hampered by the lack of closed-form expressions for the likelihood. In response, Andersen, Bollerslev, Diebold, and Labys (*Econometrica*, 71 (2003), 579–625) advocate forecasting integrated volatility via reduced-form models for the realized volatility, constructed by summing high-frequency squared returns. Building on the eigenfunction stochastic volatility models, we present analytical expressions for the forecast efficiency associated with this reduced-form approach as a function of sampling frequency. For popular models like GARCH, multi-factor affine, and lognormal diffusions, the reduced form procedures perform remarkably well relative to the optimal (infeasible) forecasts.

1. INTRODUCTION

Continuous-time stochastic volatility models figure prominently in modern asset-pricing theories. At the same time, empirical analysis of such models are generally complicated by intractable expressions for the likelihood of the observed discrete-time returns. Even though a burst of research activity has brought important advances (see, e.g., the surveys in Aït-Sahalia et al., 2005; Gallant and Tauchen, 2005; and Johannes and Polson, 2005), the discrete-time (G)ARCH class of models remains the workhorse for modeling and forecasting time-varying volatility in situations of practical import. However, recent developments in econometric methodology and the increasing availability of richer data sources hold the promise of a paradigm shift.

* Manuscript received October 2002; revised June 2003.

[1] This work was supported by a grant from the National Science Foundation to the NBER (Andersen and Bollerslev), and from FQRSC, IFM2, MITACS, NSERC, SSHRC, and Jean-Marie Dufour's Econometrics Chair of Canada (Meddahi). Discussions with Francis X. Diebold on closely related ideas have been very helpful. We would also like to thank seminar audiences at Université de Montréal, University of Pennsylvania, the Federal Reserve Board, the NSF/NBER Time Series Conference 2002, Philladelphia, the CIRANO-CIREQ Conference on Extremal Events in Finance, the 2003 Western Finance Association Meeting, Rob Engle, Rossen Valkanov, and especially Neil Shephard for helpful comments on the first version of the article. Comments from the editor and two anonymous referees also greatly improved the exposition. The third author thanks the Bendheim Center for Finance, Princeton University, for its hospitality during his visit and the revision of the article. Please address correspondence to: Torben G. Andersen, Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2001. Phone: 847-467-1285. Fax: 847-491-5719. E-mail: *t-andersen@kellogg.northwestern.edu*.

In particular, from the perspectives of asset pricing and risk management, interest typically centers on (forecasts for) the *integrated volatility* as opposed to the point-in-time (spot) volatility, which often serves as a (latent) state variable in the formulation of continuous-time models. Moreover, from a statistical perspective, the integrated volatility provides a direct measure of the discrete-time return variability appropriately defined (see, e.g., the discussion in Andersen et al., 2005 and Andersen, Bollerslev, Diebold, and Labys, 2003, henceforth ABDL). These observations, along with the increased availability of continuously recorded intraday prices (ultra-high-frequency data in the terminology of Engle, 2000), have spurred much recent research into the measurement, modeling, and forecasting of integrated volatility based on discretely sampled *realized volatilities* constructed from the summation of finely sampled squared high-frequency returns (e.g., Andersen and Bollerslev, 1998; Comte and Renault, 1998; ABDL, 2001, 2003; Barndorff-Nielsen and Shephard, 2001, 2002a,b; Meddahi, 2002).

Specifically, the empirical results of ABDL (2003) suggest that relatively simple discrete-time ARMA-based forecasts for realized volatility compare admirably to forecasts based on standard volatility models employed within the academic literature and among practitioners. Of course, from a statistical perspective these easy-to-compute reduced-form time series forecast for the observed realized volatilities invariably entail a loss in efficiency relative to the optimal, but generally infeasible, forecasts for the latent integrated volatilities based on the true underlying continuous-time model.

In response to this, the present article provides explicit analytical expressions for the expected future integrated volatility for the most popular stochastic volatility diffusion models employed in the literature, including GARCH, multifactor affine, and lognormal diffusions. We obtain these results by exploiting the general findings for the eigenfunction stochastic volatility model class introduced by Meddahi (2001). By conditioning the expectations on the full sample path realization of the latent volatility process,[2] as well as the coarser information set consisting of only the lagged realized volatilities constructed from the high-frequency returns over fixed-length time intervals, our results allow for a direct assessment of the trade-off between modeling complexity, sampling frequency, and forecast accuracy. As such, it also directly quantifies the loss associated with simple feasible procedures relative to the optimal, but infeasible ones.[3]

Following Andersen and Bollerslev (1998) and ABDL (2003) among others, we focus our forecast comparisons on the value of the coefficient of multiple correlation in the ex post regressions of the (latent) integrated volatility of interests on the forecasts obtained from the different volatility modeling procedures.[4] On

[2] Throughout the article, we identify the path realization of the volatility process and the path realization of the state variable driving the volatility process. We will be more specific when the difference between these two paths is important for forecasting purposes.

[3] This type of analysis parallels previous studies related to the predictability of mean asset returns; see, e.g., the discussion in Campbell et al. (2004).

[4] No universally acceptable loss function exists for the evaluation of nonlinear model forecasts; see, e.g., the discussion in Andersen et al. (1999) and Christoffersen and Diebold (2000). The particular loss function used here is directly inspired by the earlier contributions of Mincer and Zarnowitz (1969),

numerically quantifying this loss for empirically realistic sampling frequencies for several specific models recently reported in the literature, we find that the simple discrete-time autoregressive models for the realized volatilities perform remarkably well compared to the fully efficient (nonfeasible) continuous-time model forecasts conditional on the full sample-path realization of the latent volatility process. Hence, our results lend additional theoretical support to the use of simple empirical reduced-form modeling and forecasting procedures based on the observable realized volatilities in situations of practical import.

The plan for the rest of the article is as follows. The next section formally defines the notions of integrated and realized volatility within the class of continuous-time stochastic volatility models. We also briefly review the arguments for focusing on forecasts of integrated volatility based on projections involving realized volatility. Section 3 introduces the eigenfunction stochastic volatility class of models underlying our theoretical derivations, and also sets out the three specific models that form the basis for our numerical calculations. Section 4 presents analytical expressions for the optimal (nonfeasible) one- and multi-step-ahead forecasts for the integrated volatility conditional on the full sample path realization of the latent spot volatility process, along with the less efficient (still nonfeasible) forecasts conditional on the coarser information set consisting of "only" the lagged integrated volatilities. Section 5 in turn presents the (feasible) forecasts for the future integrated volatilities conditional on the past observable realized volatilities. This section also quantifies the corresponding loss in efficiency for each of the three illustrative candidate models as a function of the sampling frequency of the returns used in the construction of the realized volatilities. Section 6 concludes. All proofs are relegated to the Appendix.

2. INTEGRATED AND REALIZED VOLATILITY

We focus on a single asset traded in a liquid financial market. Assuming the sample path of the corresponding price process, $\{S_t, 0 \le t\}$, to be continuous,[5] the class of continuous-time stochastic volatility models traditionally employed in the finance literature is conveniently expressed in terms of the following generic stochastic differential equation (SDE):

$$d\log(S_t) = \mu_t dt + \sigma_t dW_t \tag{1}$$

where W_t denotes a standard Brownian motion, the volatility term σ_t is predictable, and the drift term μ_t is predictable and of finite variation.[6] The

---

and we will refer to the corresponding regression as such; see also the discussion in Chong and Hendry (1986).

[5] The discussion in this section explicitly rules out discontinuities in the price process. However, our new theoretical results based on the eigenfunction stochastic volatility class of models could fairly easily be extended to allow for jumps. We briefly allude to this possibility in Section 3.

[6] The drift, μ_t, may generally depend explicitly on both S_t and σ_t. However, we suppressed all the arguments for notational simplicity.

point-in-time, or spot, volatility process $\{\sigma_t, 0 \leq t\}$ measures the instantaneous strength of the price variability expressed per unit of time.

Following standard practice we assume that the sample path of the σ_t process is also continuous. Generally, σ_t and W_t may be contemporaneously correlated, so that a so-called leverage style effect is allowed. However, the (asymptotic) distributional result discussed in this section is only known to be true under the assumption that $d\sigma_t$ and dW_t are uncorrelated (no leverage effect). Likewise, the results concerning realized volatility in Section 5 preclude leverage effects. In contrast, the new theoretical results in Sections 3 and 4 explicitly allow for a nonzero (instantaneous) correlation, and we conjecture that the key results on realized volatility in Section 5 will be (approximately) valid in this case as well. Likewise, to facilitate the exposition, we explicitly exclude jump processes although many of the results remain valid for empirically relevant jump specifications.

The SDE in Equation (1) greatly facilitates arbitrage-based pricing arguments. However, as emphasized by Andersen et al. (2005), practical return calculations and volatility measurements are invariably restricted to discrete time intervals. In particular, focusing on the unit time interval, the one-period continuously compounded return corresponding to (1) is formally given by

$$r_t \equiv \log(S_t) - \log(S_{t-1}) = \int_{t-1}^{t} \mu_u \, du + \int_{t-1}^{t} \sigma_u \, dW_u \tag{2}$$

Hence, with no leverage effect and conditional on the sample-path realizations of the drift and volatility processes, $\{\mu_u, t-1 \leq u \leq t\}$ and $\{\sigma_u, t-1 \leq u \leq t\}$, the one-period returns will be Gaussian with conditional mean equal to the first integral on the right-hand side of Equation (2), whereas the conditional variance equals the *integrated volatility*

$$IV_t \equiv \int_{t-1}^{t} \sigma_u^2 \, du \tag{3}$$

The integrated volatility, therefore, affords a natural measure of the (ex post) return variability, as recently highlighted in independent work by Andersen and Bollerslev (1998), Comte and Renault (1998), and Barndorff-Nielsen and Shephard (2001). The integrated volatility also plays a key role in the stochastic volatility option pricing literature. In particular, ignoring the variation in the conditional mean, Hull and White (1987) show that option prices are uniquely determined by the expected future integrated volatility (see also Garcia et al. 2001).

Of course, integrated volatility is not directly observable. This has spurred the development of several new statistical procedures for modeling and forecasting the (latent) integrated volatility based on specific parametric models within the general diffusion class of models in Equation (1) (see, e.g., Gallant et al., 1999; Barndorff-Nielsen and Shephard, 2001; and the references therein). Although these procedures allow for the construction of asymptotically optimal forecasts under appropriate conditions, they are generally not robust to misspecifications of the underlying continuous-time model, and also quite complicated to implement.

Alternatively, consider the so-called *realized volatility* defined by the summation of intraperiod squared returns

$$RV_t(h) \equiv \sum_{i=1}^{1/h} r_{t-1+ih}^{(h)2} \tag{4}$$

where the h-period return is given by $r_t^{(h)} = \log(S_t) - \log(S_{t-h})$ and $1/h$ is a positive integer. By the theory of quadratic variation, $RV_t(h)$ converges uniformly in probability to IV_t as $h \to 0$, thus allowing for increasingly more accurate nonparametric measurements of integrated volatility as the sampling frequency of the underlying intraperiod returns increases.

If we further assume that the realized and integrated volatility measures are square integrable, the asymptotic unbiasedness of $RV_t(h)$ for IV_t, implies that forecasts for IV_{t+j}, $j \geq 1$, based on the projection of $RV_{t+j}(h)$ on any time t information set, will also be (asymptotically) unbiased and optimal in a mean square error (MSE) sense relative to that particular information set.[7] In particular, restricting the information set to the lagged realized volatilities only, as proposed by ABDL (2003), conveniently circumvents the complications associated with the use of latent variable procedures in the construction of the integrated volatility forecasts. Of course, doing so also entails a loss in forecast efficiency relative to the optimal (nonfeasible) forecasts for IV_{t+j} conditional on the full sample path realization of the instantaneous price and (latent) spot volatility processes. However, as we show below, this loss in efficiency is typically fairly small. We next turn to a discussion of the eigenfunction stochastic volatility class of models used in our formal derivation of this important practical result.

3. EIGENFUNCTION STOCHASTIC VOLATILITY MODELS

This section reviews the main properties of the eigenfunction stochastic volatility (ESV) models introduced in Meddahi (2001) and provides a discussion of the specific parametric models considered in our numerical calculations. The ESV class of models includes most continuous-time stochastic volatility models analyzed in the existing literature. Meanwhile, the formulation in terms of orthogonal eigenfunctions provides a particular convenient and elegant framework for the derivation of explicit analytical expressions for volatility forecasts.

3.1. *General Theory.* The generic stochastic volatility model in Equation (1) is only restricted by the requirement that the point-in-time, or spot, volatility

[7] This result holds generally subject to a uniform integrability condition ensuring convergence in expectation of the uniformly consistent realized volatility measure. As such, it includes cases in which the continuous sample path assumption for spot volatility is violated; see, e.g., the discussion in Andersen et al. (2005) and Barndorff-Nielsen and Shephard (2002b). Hoffman-Jørgensen (1994, Sections 3.22–3.25) provides a formal discussion of the necessary uniform integrability conditions on the underlying price process to ensure convergence in expectation. See also Billingsley (1986, Exercise 21.21) for the identical result.

process, σ_t, be nonnegative (and continuous). Most popular stochastic volatility models in the existing literature are based on the additional assumptions that the volatility process is driven by a single (latent) state variable. In the context of the ESV class of models, the corresponding one-factor representation takes the form

$$d\log(S_t) = \mu_t dt + \sigma_t\big[\sqrt{1-\rho^2}\,dW_t^{(1)} + \rho\, dW_t^{(2)}\big] \tag{5}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ denote two independent standard Brownian motions, and the instantaneous volatility is related to the latent state variable

$$df_t = m(f_t)\,dt + \sqrt{v(f_t)}\,dW_t^{(2)} \tag{6}$$

by the functional relationship

$$\sigma_t^2 = \sum_{i=0}^{p} a_i P_i(f_t) \tag{7}$$

where the integer p may be infinite, the a_i coefficients are real numbers, the $P_i(f_t)$s denote the eigenfunctions of the infinitesimal generator associated with f_t, with corresponding eigenvalues $(-\lambda_i)$, and the normalizations $P_0(f_t) = 1$ and $\mathrm{Var}[P_i(f_t)] = 1$ for $i \neq 0$ are imposed for notational simplicity.[8]

The expression for σ_t^2 in Equation (7) may appear somewhat arbitrary. Importantly, however, any square-integrable function $g(f_t)$ can be written as a linear combination of the eigenfunctions associated with f_t, i.e.,

$$g(f_t) = \sum_{i=0}^{\infty} a_i P_i(f_t) \tag{8}$$

where $a_i = E[g(f_t)P_i(f_t)]$ and $\sum_{i=0}^{\infty} a_i^2 = E[g(f_t)^2] < \infty$, so that $g(f_t)$ is the limit in mean square of $\sum_{i=0}^{p} a_i P_i(f_t)$ for p going to infinity. As such, the ESV structure encompasses the popular GARCH diffusion model (Nelson, 1990), as well as the lognormal model (Hull and White, 1987; Wiggins, 1987) and the square-root model of Heston (1993); these examples are considered in the following subsection.

The power of the ESV representations of these and other continuous-time stochastic volatility models essentially stems from the following two properties. First, the eigenfunctions associated with different eigenvalues are orthogonal and any nonconstant eigenfunction is centered at zero (for $i, j > 0$ and $i \neq j$):

$$E[P_i(f_t)P_j(f_t)] = 0 \quad \text{and} \quad E[P_i(f_t)] = 0 \tag{9}$$

[8] For a more detailed discussion of the properties of infinitesimal generators see, e.g., Hansen and Scheinkman (1995) and Aït-Sahalia et al. (2005).

These features, of course, underlie the result noted after Equation (8) that $\sum_{i=0}^{\infty} a_i^2 = E[g(f_t)^2]$. Second, the eigenfunctions are first-order autoregressive processes (in general heteroskedastic):

$$\forall l > 0,\ E[P_i(f_{t+l}) \mid f_\tau, \tau \le t] = \exp(-\lambda_i l) P_i(f_t) \tag{10}$$

Given the structure of the ESV model and the Markovian nature of the joint process $(S_t,\ f_t)$, conditional expectations of any transformation of this variable, including the variance, therefore, only depend on the expectations of the eigenfunctions. The orthogonality of the eigenfunctions coupled with the simple first-order autoregressive dynamics in turn render such calculations straightforward.

The ESV representations discussed above are based on a single (latent) state variable. Meanwhile, several recent studies, including Alizadeh et al. (2002), Bollerslev and Zhou (2002), Engle and Lee (1999), Gallant et al. (1999), and Harvey et al. (1994), among others, have argued for the empirical relevance of allowing for multiple volatility factors. Fortunately, the ESV approach easily allows for multiple factors, while maintaining the validity of the formulas in (8), (9), and (10); see Meddahi (2001) for further details.[9]

3.2. *Specific Illustrations.* The numerical analysis presented in subsequent sections is based on three specific models, namely, a GARCH diffusion model, a two-factor affine model, and a lognormal diffusion. Meddahi (2001) shows how each of these models may be represented as an ESV model by explicitly solving for the corresponding eigenfunctions. The following subsections provide a brief summary of these results along with the actual parameter values used in the numerical calculations.

3.2.1. *Model M1—GARCH diffusion.* The instantaneous volatility in the GARCH diffusion model is defined by the process

$$d\sigma_t^2 = \kappa\left(\theta - \sigma_t^2\right) dt + \sigma \sigma_t^2 dW_t^{(2)}$$

This model was first introduced by Wong (1964) and later popularized by Nelson (1990). The model is readily expressed as an ESV model by defining the state variable

$$df_t = \kappa(\theta - f_t)\, dt + \sigma f_t dW_t^{(2)}$$

and the function $g(x) = x$. Assuming that the variance of σ_t^2 is finite, it is possible to show that

$$\sigma_t^2 = a_0 + a_1 P_1(f_t)$$

[9] See also Chen et al. (2000) for a general approach to eigenfunction modeling for multivariate Markov processes.

where $a_0 = \theta$, $a_1 = \theta\sqrt{\psi/(1-\psi)}$, $\psi = \sigma^2/2\kappa$, and the first eigenfunction for f_t is affine

$$P_1(x) = \frac{\sqrt{1-\psi}}{\theta\sqrt{\psi}}(x-\theta)$$

with corresponding eigenvalue $\lambda_1 = \kappa$. As discussed above, this representation of the process greatly facilitates any expected variance and/or volatility forecast calculations. Note also that the second moment of the variance σ_t^2 is assured to be finite for ψ less than 1. In the numerical calculations reported here we rely on the parameters from Andersen and Bollerslev (1998) as implied from the (weak) daily GARCH(1,1) model estimates for the DM/dollar from 1987 through 1992 using the temporal aggregation results of Drost and Nijman (1993) and Drost and Werker (1996). In particular, $\kappa = 0.035$, $\theta = 0.636$, and $\psi = 0.296$. These parameter values were also used in the studies by Andersen et al. (1999) and Andreou and Ghysels (2002).

3.2.2. *Model M2—two-factor affine.* The instantaneous variance in the two-factor affine model is given by

$$\sigma_t^2 = \sigma_{1,t}^2 + \sigma_{2,t}^2, \quad d\sigma_{j,t}^2 = \kappa_j\big(\theta_j - \sigma_{j,t}^2\big)\,dt + \eta_j\sigma_{j,t}\,dW_t^{(j+1)}, \quad j = 1, 2$$

Following Meddahi (2001), this model may be cast in the form of an ESV model by defining the state variables,

$$df_{j,t} = \kappa_j(\alpha_j + 1 - f_{j,t})\,dt + \sqrt{2\kappa_j}\sqrt{f_{j,t}}\,dW_t^{(j+1)}, \quad j = 1, 2$$

where $\alpha_j = (2\kappa_j\,\theta_j/\eta_j^2) - 1$, and the $f_{j,t}$s are related to the $\sigma_{j,t}$s by the functional relationship

$$f_{j,t} = \frac{2\kappa_j}{\eta_j^2}\sigma_{j,t}^2, \quad j = 1, 2$$

In particular, it is possible to show that the eigenfunctions associated with $f_{j,t}$ are given by the Laguerre polynomials [10] $L_i^{(\alpha_j)}(f_{j,t})$, $i = 0, 1, \ldots$, with corresponding eigenvalues $\lambda_{j,i} = \kappa_j i$. Moreover,

$$\sigma_{j,t}^2 = \tilde{a}_{j,0} + \tilde{a}_{j,1}L_1^{(\alpha_j)}(f_{j,t})$$

with $\tilde{a}_{j,0} = \theta_j$ and $\tilde{a}_{j,1} = -\sqrt{\theta_j}\eta_j/\sqrt{2\kappa_j}$, so that

$$\sigma_t^2 = a_{0,0} + a_{1,0}L_1^{(\alpha_1)}(f_{1,t}) + a_{0,1}L_1^{(\alpha_2)}(f_{2,t})$$

[10] For a reference on Laguerre Polynomials, see, for instance, Szego (1975, p. 100).

where $a_{0,0} = \tilde{a}_{1,0} + \tilde{a}_{2,0}$, $a_{1,0} = \tilde{a}_{1,1}$, and $a_{0,1} = \tilde{a}_{2,1}$. The actual numerical results for the two-factor model are based on the parameter estimates reported in Bollerslev and Zhou (2002) obtained by matching the sample moments of the daily realized volatilities constructed from high-frequency 5-minute DM/dollar returns spanning 1986 through 1996 to the corresponding population moments for the integrated volatility. The resulting values are $\kappa_1 = 0.5708$, $\theta_1 = 0.3257$, $\eta_1 = 0.2286$, $\kappa_2 = 0.0757$, $\theta_2 = 0.1786$, and $\eta_2 = 0.1096$, implying the existence of a very volatile first factor, along with a much more slowly mean-reverting second factor.

3.2.3. *Model M3—lognormal diffusion.* Our last numerical example is based on the lognormal diffusion model

$$d \log\left(\sigma_t^2\right) = \kappa\left[\theta - \log\left(\sigma_t^2\right)\right] dt + \sigma\, dW_t^{(2)}$$

Again, following Meddahi (2001), this model may be expressed in the form of an ESV model by defining the state variable

$$df_t = -\kappa f_t dt + \sqrt{2\kappa}\, dW_t^{(2)}$$

related to σ_t by the functional relationship

$$f_t = \frac{\sqrt{2\kappa}}{\sigma}\left(\log \sigma_t^2 - \theta\right)$$

The eigenfunctions associated with this Ornstein–Uhlenbeck process for f_t are given by the usual Hermite polynomials $H_i(f_t)$, $i = 0, 1, \ldots,$ with corresponding eigenvalues $\lambda_i = \kappa i$. Hence

$$\sigma_t^2 = \sum_{i=0}^{\infty} a_i H_i(f_t)$$

where the a_i coefficients take the form

$$a_i = \exp\left(\theta + \frac{\sigma^2}{4\kappa}\right) \frac{(\sigma/\sqrt{2\kappa})^i}{\sqrt{i!}}$$

Our numerical illustrations for this model rely on the EMM-based parameter estimates for the daily S&P500 returns spanning 1953 through 1996 reported in Andersen et al. (2002), where we restrict the estimated correlation between $dW_t^{(1)}$ and $dW_t^{(2)}$ related to the leverage effect to be zero. In particular, $\kappa = 0.0136$, $\theta = -0.8382$, and $\sigma = 0.1148$. Finally, the summation in (7) is truncated at $p = 100$.

4. IDEAL INTEGRATED VOLATILITY FORECASTS

This section provides analytic expressions for the basic dependency structure of integrated volatility across the full range of ESV diffusion models. Building on these findings, we go on to characterize the extent of the predictability of integrated volatility for any member of this important class of asset price processes. The predictability is obviously dependent on the assumed information set. We present results ranging from the ideal case of knowing the current (latent) volatility state, through knowing the spot volatility, to observing the past sequence of integrated volatility only. All such information sets are unattainable in practice, as they contain variables that are not observed, but rather must be estimated or extracted from discrete return data. Nonetheless, they serve as important benchmarks that establish the maximal predictability and reveal, step by step, how much forecast power is lost as we condition on successively less informative, but empirically more readily approximated, variables. The results based on conditioning the forecasts on past integrated volatility set the stage for the analysis of the feasible integrated volatility forecasts based on realized volatility measures extracted directly from observed high-frequency, intraday data in the following section.

Our first proposition characterizes the basic first and second moment properties of the spot and integrated volatility processes within the ESV diffusion class (the proofs are given in the Appendix). The results are all new at this level of generality, but some expressions are clearly related to the abstract characterization of the second moment properties of integrated volatility in Barndorff-Nielsen and Shephard (2002a). Moreover, these authors also provide some concrete results for special cases of the ESV model and the non-Gaussian Ornstein–Uhlenbeck model.

The natural "realized" benchmark for the multi-step-ahead forecasts at horizon n is given by the corresponding integrated volatility

$$IV_{t+1:t+n} = \sum_{i=1}^{n} IV_{t+i} \tag{11}$$

PROPOSITION 1. *For any ESV diffusion model, as defined in Section 3.1 (with potential nonzero drift or leverage effects), and any integers* $n \geq 1, l \geq 0$, *we have*

$$E\big[\sigma_t^2\big] = a_0, \quad E[IV_{t+1:t+n}] = na_0 \tag{12}$$

$$E\big[\sigma_{t+n}^2 \,\big|\, S_\tau,\ f_\tau, \tau \leq t\big] = a_0 + \sum_{i=1}^{p} a_i \exp(-\lambda_i n) P_i(f_t) \tag{13}$$

$$E[IV_{t+1:t+n} \,|\, S_\tau,\ f_\tau, \tau \leq t] = na_0 + \sum_{i=1}^{p} a_i \frac{[1 - \exp(-\lambda_i n)]}{\lambda_i} P_i(f_t) \tag{14}$$

$$\mathrm{Var}\big[\sigma_t^2\big] = \sum_{i=1}^{p} a_i^2, \quad \mathrm{Var}[IV_{t+1:t+n}] = 2\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}[\exp(-\lambda_i n) + \lambda_i n - 1] \tag{15}$$

$$\mathrm{Cov}\big(IV_{t+1:t+n}, \sigma^2_{t-l}\big) = \sum_{i=1}^{p} a_i^2 \frac{[1-\exp(-\lambda_i n)]}{\lambda_i} \exp(-\lambda_i l) \tag{16}$$

$$\mathrm{Cov}(IV_{t+1:t+n}, IV_{t-l}) = \sum_{i=1}^{p} a_i^2 \frac{[1-\exp(-\lambda_i)][1-\exp(-\lambda_i n)]}{\lambda_i^2} \exp(-\lambda_i l) \tag{17}$$

$$\mathrm{Cov}\big(\sigma^2_{t+n}, \sigma^2_t\big) = \sum_{i=1}^{p} a_i^2 \exp(-\lambda_i n) \tag{18}$$

One set of novel implications is given by the following corollary to Proposition 1. It provides a rather intuitive, yet useful, ranking of the second moment expressions for the different latent volatility notions.

PROPOSITION 2. *Under the assumptions of Proposition 1, and* $\forall n \geq 1$, *we have*

$$\mathrm{Cov}\big(\sigma^2_{t+n}, \sigma^2_t\big) \leq \mathrm{Cov}(IV_{t+n}, IV_t) \leq \mathrm{Cov}\big(IV_{t+n}, \sigma^2_t\big) \leq \mathrm{Var}[IV_t] \leq \mathrm{Var}\big[\sigma^2_t\big] \tag{19}$$

The inequalities are most readily understood by recalling the timing between spot and integrated volatility and recognizing that integrated volatility is a smoothed version of the spot volatility process. The first inequality, for example, reflects the fact that the spot volatilities are separated by n periods, whereas the gap between the intervals over which the integrated volatilities, IV_{t+n} and IV_t, are measured is only $n-1$ periods. The second inequality suggests that the spot volatility at the interval end is more informative about future integrated volatility than the smoothed (average) spot volatility over the corresponding interval. Such a conclusion appears natural in the single eigenfunction case, given the Markov structure of ESV models, but is less obvious for the multiple eigenfunction case where affine functions of spot volatility cannot provide a sufficient statistic for the volatility state vector, f_t. The final inequalities are hardly surprising, but have important implications. The variability of either volatility measure exceeds the covariability measures and the smoothed integrated volatility measure is less variable than the spot volatility. The latter finding implies that the comparatively high covariance between spot volatility and future integrated volatility may be due to (excess) variability of spot volatility rather than superior correlation between spot and future (integrated) volatility. Hence, as discussed below, it is not clear in general whether conditioning on the spot or the integrated volatility will allow for the more efficient forecast.

4.1. *Optimal One-Period-Ahead Forecasts.* We are now in position to assess the optimal integrated volatility forecasts generated by different information sets. We adopt the standard expected quadratic loss function, implying that optimal forecasts equal the conditional expectation of integrated volatility given the available information. The universally best forecast is based on the full history of the log-price and volatility path, denoted by the sigma algebra, $\sigma(S_\tau, f_\tau, \tau \leq t)$. As discussed above, given the Markov structure of the ESV models, knowledge of the state vector, f_t, is sufficient for the construction of the optimal predictor. The coefficient of multiple correlation, or R^2, from the Mincer–Zarnowitz type regression of future integrated volatility on the corresponding (conditional expectation) forecasts (and a constant) serves as a popular and useful summary measure of forecast performance. Moreover, it is consistent with the emphasis on quadratic loss. For notational convenience, we focus our analysis on the one-period-ahead integrated volatility, IV_{t+1}, but similar results are readily available for the n-period-ahead measure, IV_{t+n}.

Proposition 1 and the orthogonality of the eigenfunctions, indicated in (9), imply that the "explained variation" from regressing IV_{t+1} onto $E[IV_{t+1} \mid S_\tau, f_\tau, \tau \leq t]$ (and a constant), denoted $R^2(IV_{t+1}, Best)$, is given by

$$R^2(IV_{t+1}, Best) = \frac{1}{\mathrm{Var}[IV_t]} \sum_{i=1}^{p} a_i^2 \frac{[1 - \exp(-\lambda_i)]^2}{\lambda_i^2} \tag{20}$$

with $\mathrm{Var}[IV_t]$ determined by (15) (for $n = 1$).

Alternatively, consider forecasts based on the current spot volatility. In one-factor ESV models, there is no difference between the conditional expectation of integrated volatility given either spot volatility or the volatility state vector, so the two forecasts coincide. In multifactor models, however, the spot volatility is not a sufficient statistic for the volatility state vector, with the latter being more informative. Moreover, the process (S_t, σ_t) is not Markovian, and, hence, the associated optimal forecast—the conditional expectation of IV_{t+1} given $\sigma(S_\tau, \sigma_\tau^2, \tau \leq t)$—depends in general on the entire path of σ_t^2 and is not available in closed form.

Consequently, we next consider a simple forecast that depends linearly on σ_t^2. Obviously, the best affine forecast of IV_{t+1} is given by the corresponding (population) regression coefficients on σ_t^2 and a constant. Using (16), it follows readily that the R^2 of the corresponding Mincer–Zarnowitz regression, denoted $R^2(IV_{t+1}, \sigma_t^2)$, may be expressed as

$$R^2\big(IV_{t+1}, \sigma_t^2\big) = \frac{1}{\mathrm{Var}[IV_t]\mathrm{Var}\big[\sigma_t^2\big]} \left(\sum_{i=1}^{p} a_i^2 \frac{[1 - \exp(-\lambda_i)]}{\lambda_i} \right)^2 \tag{21}$$

where $\mathrm{Var}[IV_t]$ and $\mathrm{Var}[\sigma_t^2]$ are given by (15). In the univariate factor case, if the variance is solely a function of a single (nonconstant) eigenfunction, this forecast still coincides with the "best." However, when the variance depends on multiple eigenfunctions, the forecast will differ from the "best," even in the case of a single factor. In the specific examples discussed in Section 3.2, we have one case (the GARCH diffusion) where the forecasts coincide (single factor, one eigenfunction), one where they differ in a univariate factor model (the lognormal diffusion, where the variance depends on multiple eigenfunctions), and one where the spot volatility forecast is necessarily inferior to the "best"(the two-factors affine diffusion).

Forecasts based on integrated volatility are of particular interest as they, in practice, may be approximated by the corresponding realized volatility obtained from high-frequency data. Current (and past) integrated volatility will generally not provide a sufficient statistic for the volatility state vector, so the optimal forecasts are not available in closed form and we again restrict attention to forecasts generated by affine functions of the integrated volatility. In particular, on using (17) it follows that the R^2 from the regression of IV_{t+1} on IV_t and a constant, denoted $R^2(IV_{t+1}, IV_t)$, takes the form

$$R^2(IV_{t+1}, IV_t) = \left(\sum_{i=1}^{p} a_i^2[1 - \exp(-\lambda_i)]^2/\lambda_i^2\right)^2 \Big/ \mathrm{Var}[IV_t]^2 \tag{22}$$

Obviously, the optimal forecast for the one-period integrated volatility will, by construction, dominate in terms of the population R^2 from the associated Mincer–Zarnowitz regressions. More interesting is the relative performance of the forecasts in (21) and (22). Assuming a unique eigenfunction in (7), it follows that[11]

$$R^2(IV_{t+1}, IV_t) \leq R^2\big(IV_{t+1}, \sigma_t^2\big)$$

However, no general ranking of $R^2(IV_{t+1}, IV_t)$ relative to $R^2(IV_{t+1}, \sigma_t^2)$ is feasible for the case of multiple eigenfunctions.[12] Intuitively, low persistence of the eigenfunctions hurts the relative performance of integrated volatility based forecasts, whereas the main culprit behind poor performance of spot volatility-based predictors is a large discrepancy in persistence across the eigenfunctions (and associated high variability of the spot volatility process). Interestingly, in spite of the higher covariance between the spot volatility and the future (integrated) volatility, it is generally not the case that the corresponding forecasts dominate those generated by conditioning on the integrated volatility. In fact, this is only assured when there is a single eigenfunction—in which case the spot volatility-based predictor coincides with the optimal one.[13] In this situation, the integrated volatility forecasts should ideally be based on (an estimate of) current spot volatility or (an estimate of) a current integrated volatility measure covering as short an intraday interval as possible. However, practical difficulties in obtaining precise intraday-horizon volatility estimates limit the applicability of this insight.

Indeed, from a practical perspective the more relevant question is how *much* predictive power is lost for the different forecasts and how one may alleviate the loss in forecast accuracy if only historical integrated volatility-based predictors are feasible. We return to the first question in our numerical comparisons

[11] By Equations (21), (22), and (15), the ratio $R^2(IV_{t+1}, IV_t)/R^2(IV_{t+1}, \sigma_t^2)$ equals $(1 - \exp(-\lambda_1))^2/2(\exp(-\lambda_1) + \lambda_1 - 1))$, which necessarily is smaller than 1 by Equation (A.12) in the Appendix.

[12] For instance, consider the two numerical examples: $p = 2$, $a_1 = \sqrt{0.9}$, $a_2 = \sqrt{0.1}$, $\lambda_1 = 0.002$, and $\lambda_2 = 4$ in the first, and $\lambda_2 = 1$ in the second. The ratio $R^2(IV_{t+1}, IV_t)/R^2(IV_{t+1}, \sigma_t^2)$ then equals 1.023 and 0.977, respectively.

[13] As noted above, the GARCH diffusion model falls in this category.

in Section 4.3. One approach for addressing the second question involves the inclusion of additional (lagged) explanatory variables, as discussed formally in the following section.

4.2. *Multiple Explanatory Variables and Multiperiod Forecasts.* Since the optimal forecast of IV_{t+1} conditional on the history of spot volatility generally depends on the entire volatility path, it is natural to extend the expression for the best affine forecast of IV_{t+1} given σ_t^2 to include a fixed, finite number of lagged spot volatilities (and a constant) as regressors. Similarly, it is relevant to consider the best affine forecast of IV_{t+1} given IV_t and its lagged values.

For this purpose, it is convenient first to introduce some notation. For a covariance-stationary random variable (y_τ, z_t) and an integer l, we let $C(y_\tau, z_t, l)$ denote the $(l+1)$ vector defined by

$$C(y_\tau, z_t, l) = (\mathrm{Cov}[y_\tau, z_t], \mathrm{Cov}[y_\tau, z_{t-1}], \dots, \mathrm{Cov}[y_\tau, z_{t-l}])^\top \tag{23}$$

Moreover, let $M(z_t, l)$ denote the $(l+1)\times(l+1)$ matrix whose (i, j)th component is given by

$$M(z_t, l)[i, j] = \mathrm{Cov}[z_t, z_{t+i-j}] \tag{24}$$

The R^2 from the regression of IV_{t+1} onto a constant and $(\sigma_t^2, \sigma_{t-1}^2, \dots, \sigma_{t-l}^2)$, $l \geq 0$, denoted $R^2(IV_{t+1}, \sigma_t^2, l)$, may then be succinctly expressed as[14]

$$R^2\big(IV_{t+1}, \sigma_t^2, l\big) = C\big(IV_{t+1}, \sigma_t^2, l\big)^\top \big(M(\sigma_t^2, l)\big)^{-1} C\big(IV_{t+1}, \sigma_t^2, l\big)/\mathrm{Var}[IV_t] \tag{25}$$

The R^2 from the regression of IV_{t+1} onto a constant and $(IV_t, IV_{t-1}, \dots, IV_{t-l})$, $l \geq 0$, denoted $R^2(IV_{t+1}, IV_t, l)$, takes a similar form, as detailed in the Appendix.

The final type of forecasts we consider is based on ARMA-type representations of integrated volatility. Barndorff-Nielsen and Shephard (2002a) note that autoregressive specifications for spot volatility induce an ARMA structure for the integrated volatility process. Meddahi (2003) shows that integrated volatility is an ARMA(1,1) process (respectively, ARMA(2,2)) if spot volatility depends on a single eigenfunction (respectively, two eigenfunctions), and also provides closed-form expressions for all the ARMA parameters. Utilizing these results, the corresponding (population) R^2 based on the ARMA representation for the integrated volatility

$$R^2(IV_{t+1}, \mathrm{ARMA}) = 1 - \mathrm{Var}[\mathrm{Innovation}]/\mathrm{Var}[IV_t] \tag{26}$$

may be evaluated numerically on the basis of the exact expression for the innovation variance, Var[Innovation], given in Meddahi (2003).

[14] Recall that the R^2 from a regression with multiple regressors, i.e., $y_t = c + x_t\beta + \eta_t$ where x_t denotes a vector of explanatory variables, is given by $R^2 = \mathrm{Cov}(y, x)\,(\mathrm{Var}[x])^{-1}\,\mathrm{Cov}(x, y)/\mathrm{Var}[y]$.

This detailed characterization of the properties of the various one-period-ahead volatility forecasts readily extends to the corresponding multiperiod forecasts. In particular, in analogy to the results for the one-period forecasts, it follows from Proposition 1 that

$$(27)\quad R^2(IV_{t+1:t+n}, Best) = \frac{1}{\mathrm{Var}[IV_{t+1:t+n}]} \sum_{i=1}^{p} a_i^2 \frac{[1-\exp(-\lambda_i n)]^2}{\lambda_i^2}$$

$$(28)\quad R^2\big(IV_{t+1:t+n}, \sigma_t^2\big) = \frac{1}{\mathrm{Var}[IV_{t+1:t+n}]\mathrm{Var}\big[\sigma_t^2\big]} \left(\sum_{i=1}^{p} a_i^2 \frac{[1-\exp(-\lambda_i n)]}{\lambda_i}\right)^2$$

$$(29)\quad R^2(IV_{t+1:t+n}, IV_t) = \frac{\left(\sum_{i=1}^{p} a_i^2[1-\exp(-\lambda_i)][1-\exp(-\lambda_i n)]/\lambda_i^2\right)^2}{\mathrm{Var}[IV_{t+1:t+n}]\mathrm{Var}[IV_t]}$$

where $\mathrm{Var}[IV_{t+1:t+n}]$, $\mathrm{Var}[IV_t]$, and $\mathrm{Var}[\sigma_t^2]$ are given in (15). Similar expressions for R^2s from the regressions of IV_{t+1} onto a constant and lagged values of σ_t^2 and IV_t are given in the Appendix.

4.3. *Quantifying the Forecast Performance.* The population R^2s of the Mincer–Zarnowitz regressions for the forecast of the one-period-ahead integrated volatilities for the three specific models detailed in Section 3.2 are given in Table 1. First, it is noteworthy that when only one volatility factor is employed in the ESV diffusion, then the factor tends to be strongly serially correlated, leading to a high degree of predictability for the integrated volatility (models M1 and M3), irrespective of the volatility measure in the information set. Second, if another factor is brought into the model, it will typically allow for a more volatile and less persistent factor in the volatility dynamics (model M2). This lowers the fundamental persistence and predictability of the volatility process and renders the integrated volatility measures more noisy indicators of current (spot) volatility. Hence, the corresponding forecasts based on IV_{t-j}, $j \geq 0$, become less accurate relative to the

TABLE 1

IDEAL ONE-PERIOD-AHEAD FORECASTS OF INTEGRATED VOLATILITY

Model	M1	M2	M3
$R^2(IV_{t+1}, Best)$	0.977	0.830	0.989
$R^2(IV_{t+1}, \sigma_t^2)$	0.977	0.819	0.989
$R^2(IV_{t+1}, \sigma_t^2, 1)$	0.977	0.820	0.989
$R^2(IV_{t+1}, \sigma_t^2, 4)$	0.977	0.821	0.989
$R^2(IV_{t+1}, IV_t)$	0.955	0.689	0.977
$R^2(IV_{t+1}, IV_t, 1)$	0.957	0.694	0.979
$R^2(IV_{t+1}, IV_t, 4)$	0.957	0.698	0.979
$R^2(IV_{t+1}, ARMA)$	0.957	0.699	–

TABLE 2

IDEAL MULTI-PERIOD-AHEAD FORECASTS OF INTEGRATED VOLATILITY

Model	M1			M2			M3		
Horizon	5	10	20	5	10	20	5	10	20
$R^2(IV_{t+1:t+n}, Best)$	0.891	0.797	0.645	0.586	0.479	0.338	0.945	0.895	0.807
$R^2(IV_{t+1:t+n}, \sigma_t^2)$	0.891	0.797	0.645	0.492	0.349	0.222	0.945	0.894	0.804
$R^2(IV_{t+1:t+n}, \sigma_t^2, 1)$	0.891	0.797	0.645	0.499	0.359	0.231	0.945	0.894	0.804
$R^2(IV_{t+1:t+n}, \sigma_t^2, 4)$	0.891	0.797	0.645	0.508	0.371	0.242	0.945	0.894	0.804
$R^2(IV_{t+1:t+n}, IV_t)$	0.871	0.779	0.630	0.445	0.320	0.214	0.934	0.885	0.796
$R^2(IV_{t+1:t+n}, IV_t, 1)$	0.873	0.781	0.632	0.445	0.330	0.216	0.936	0.886	0.796
$R^2(IV_{t+1:t+n}, IV_t, 4)$	0.874	0.781	0.632	0.446	0.343	0.227	0.936	0.886	0.797
$R^2(IV_{t+1:t+n}, ARMA)$	0.874	0.781	0.632	0.460	0.347	0.231	–	–	–

forecasts that exploit more current information. Third, there is little evidence that the addition of lagged variables to the information set has any practical impact on forecast performance.

The R^2 values for the multi-step-ahead forecasts for the same three models are reported in Table 2. The results are reported for forecasts covering 1 week ($n = 5$), 2 weeks ($n = 10$), and about 1 month ($n = 20$).

For each of the three models, a large fraction of the variability of the (integrated) volatility process remains predictable, even at the monthly horizon, although the proportion now varies substantially across the models. For models M1 and M3, the loss of explanatory power associated with the construction of forecasts from spot or integrated volatility rather than the true volatility state is still limited. Of course, for model M1 the use of spot volatility is equivalent to the use of the true volatility state. As a result, there is limited scope for improvement through the addition of lagged variables in the information set, or the use of the theoretically warranted ARMA(1,1) structure for model M1. For model M2, however, there is now an appreciable deterioration in performance as we move from full information to spot volatility, and then further on to integrated volatility. The use of additional lagged variables and the theoretically motivated ARMA(2,2) model for integrated volatility now also produces a small, but nonnegligible improvement.

Overall, our investigation suggests that models with a single persistent volatility factor are relatively insensitive to the choice of variables in the information set. All natural forecast procedures do well and capture a large degree of the theoretical predictability. In contrast, there is clearly some loss in predictive power when the model contains a second, less persistent volatility factor. Moreover, for such models one may obtain nontrivial gains by expanding the information set to include several lags of the integrated volatility through a simple AR model or, better, a theoretically motivated ARMA structure.

5. VOLATILITY FORECASTS BASED ON REALIZED VOLATILITY

None of the forecasts discussed in the previous section are actually feasible as the true volatility state vector, the spot volatility, and the integrated volatility are all latent, when only discretely sampled price data are available. The variable among these that may most readily be approximated with reasonably good precision from observed data is the integrated volatility. In particular, as discussed in Section 2, the observable realized volatility consistently approximates the (latent) integrated volatility for increasingly finer sampled returns. Of course, any practical application necessarily relies on realized volatility constructed from finitely sampled asset prices, and as such inevitably embodies a measurement error vis-à-vis the corresponding integrated volatility. It is consequently important to assess the magnitude of this measurement error and the associated loss in forecast efficiency. This section addresses these issues analytically.

5.1. *Theoretical Relationships.* In order to assess the loss of precision in the forecast evaluation regressions, we explore the relation between integrated and realized volatility in more detail. Throughout this section, we preclude drift and leverage effects. In this setting, as shown in Barndorff-Nielsen and Shephard (2002a), and also emphasized by Andersen et al. (2005) and Meddahi (2002), the measurement error, $U_t(h) \equiv RV_t(h) - IV_t$, is mean-zero, serially uncorrelated, and orthogonal to the IV_t process (i.e., $\mathrm{Cov}(U_t(h), IV_{t-i}) = 0$ for all $i \in \mathbf{Z}$). As a consequence,

$$\mathrm{Var}[RV_t(h)] = \mathrm{Var}[IV_t] + \mathrm{Var}[U_t(h)] \tag{30}$$

$$\mathrm{Var}[RV_{t+1:t+n}(h)] = \mathrm{Var}[IV_{t+1:t+n}] + n\mathrm{Var}[U_t(h)] \tag{31}$$

$$\mathrm{Cov}[RV_t(h), RV_{t-i}(h)] = \mathrm{Cov}[IV_t, IV_{t-i}] = \mathrm{Cov}[RV_t(h), IV_{t-i}] \tag{32}$$

where $i \neq 0$, and $\mathrm{Cov}[IV_t, IV_{t-i}]$ is given by (17) for $n = 1$ and $l = i - 1$.

Moreover, within the context of the ESV class of models, it follows from the results in Meddahi (2002) that[15]

$$\mathrm{Var}[U_t(h)] = \frac{4}{h}\left(\frac{a_0^2 h^2}{2} + \sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}[\exp(-\lambda_i h) - 1 + \lambda_i h]\right) \tag{33}$$

The R^2s for the Mincer–Zarnowitz regressions involving the realized volatility are now readily derived from the corresponding R^2s for the integrated volatility. The next proposition collects the general results.

[15] The same formula has previously been derived by Barndorff-Nielsen and Shephard (2002a) under the more restrictive assumption that the spot variance is a finite linear combination of autoregressive and independent processes (corresponding to the CEV and positive Ornstein–Uhlenbeck processes). The result derived here coincides with this earlier formula in the case of a unique eigenfunction in (7), but otherwise is more general. Similarly, expressions corresponding to the formulas in (12), (15), (17), and (18) have previously been established by Barndorff-Nielsen and Shephard (2002a) in their more restrictive setting, whereas Barndorff-Nielsen and Shephard (2004, Chapter 7) give the second equality in (32).

PROPOSITION 3. *For any ESV diffusion model without drift and leverage effects, and with the realized volatility as the only regressor (apart from a constant), we have*

$$R^2(\cdot, RV_t(h)) = R^2(\cdot, IV_t)\frac{\mathrm{Var}[IV_t]}{\mathrm{Var}[RV_t(h)]} = R^2(\cdot, RV_t(h))\frac{\mathrm{Var}[IV_t]}{\mathrm{Var}[IV_t] + \mathrm{Var}[U_t(h)]} \tag{34}$$

With realized volatility as the dependent variable, any set of regressors, and for any integer $n \geq 1$,

$$R^2(RV_{t+1:t+n}(h), \cdot) = R^2(IV_{t+1:t+n}, \cdot)\frac{\mathrm{Var}[IV_{t+1:t+n}]}{\mathrm{Var}[RV_{t+1:t+n}(h)]} = R^2(IV_{t+1:t+n}, \cdot)\frac{\mathrm{Var}[IV_{t+1:t+n}]}{\mathrm{Var}[IV_{t+1:t+n}] + n\mathrm{Var}[U_t(h)]} \tag{35}$$

As a simple implication of this proposition, it follows that the R^2s associated with the one-period realized volatility forecast evaluation regressions are always lower than the infeasible ones determined in Section 4:

$$R^2(RV_{t+1}, RV_t) \leq R^2(IV_{t+1}, RV_t) = R^2(RV_{t+1}, IV_t) \leq R^2(IV_{t+1}, IV_t) \tag{36}$$

This documents the intuitive result that the use of a (feasible) realized volatility proxy in place of the (latent) integrated volatility systematically lowers the predictive power, irrespective of whether the proxy is inserted as a regressor, regressand, or both. The main issue is, of course, how serious the loss in forecast efficiency will be in empirically realistic situations.

5.2. *Illustrations Based on Specific Models*

5.2.1. *Forecasts from past realized volatility.* To quantify the efficiency loss that is likely to occur in practice, Tables 3 and 4 report the population R^2s for the regressions of future integrated volatility on lagged values of realized volatility for the three specific ESV models considered previously. The sampling frequency used for the realized volatility measures correspond to 5-minute returns for a 24-hour trading day ($1/h = 288$), an 8-hour trading day ($1/h = 96$), and 30-minute returns for a 24-hour trading day ($1/h = 48$).

Table 3 provides an indication of the feasible predictability of one-period-ahead integrated volatility. There is a noticeable drop compared to Table 1, but a very large proportion of integrated volatility remains predictable. The predictability also improves markedly as we move from realized volatility measures constructed using 48 to 288 intraday observations. Moreover, it is evident that the use of additional lagged variables in the information set is helpful only for the more imprecisely realized volatility measures based on 48 intraday price observations.

TABLE 3

ONE-PERIOD-AHEAD FORECASTS OF INTEGRATED VOLATILITY BASED ON REALIZED VOLATILITY

Model	M1			M2			M3		
1/h	48	96	288	48	96	288	48	96	288
$R^2(IV_{t+1}, RV_t(h))$	0.836	0.891	0.932	0.476	0.563	0.641	0.881	0.927	0.960
$R^2(IV_{t+1}, RV_t(h), 1)$	0.873	0.906	0.934	0.507	0.574	0.642	0.917	0.943	0.962
$R^2(IV_{t+1}, RV_t(h), 4)$	0.883	0.908	0.934	0.519	0.580	0.642	0.929	0.946	0.963
$R^2(IV_{t+1}, RV_t(h), ARMA)$	0.883	0.908	0.934	0.522	0.582	0.646	–	–	–

TABLE 4

MULTI-PERIOD-AHEAD FORECASTS OF INTEGRATED VOLATILITY BASED ON REALIZED VOLATILITY

Horizon	5			10			20		
1/h	48	96	288	48	96	288	48	96	288
Model M1									
$R^2(IV_{t+1:t+n}, RV_t(h))$	0.762	0.813	0.851	0.682	0.727	0.761	0.551	0.588	0.615
$R^2(IV_{t+1:t+n}, RV_t(h), 1)$	0.797	0.827	0.852	0.713	0.740	0.762	0.576	0.598	0.616
$R^2(IV_{t+1:t+n}, RV_t(h), 4)$	0.805	0.829	0.852	0.720	0.741	0.762	0.582	0.599	0.616
$R^2(IV_{t+1:t+n}, RV_t(h), ARMA)$	0.806	0.829	0.852	0.721	0.741	0.762	0.582	0.599	0.616
Model M2									
$R^2(IV_{t+1:t+n}, RV_t(h))$	0.307	0.364	0.414	0.226	0.268	0.305	0.148	0.175	0.199
$R^2(IV_{t+1:t+n}, RV_t(h), 1)$	0.339	0.381	0.419	0.255	0.285	0.312	0.169	0.188	0.205
$R^2(IV_{t+1:t+n}, RV_t(h), 4)$	0.360	0.395	0.429	0.277	0.302	0.325	0.186	0.202	0.216
$R^2(IV_{t+1:t+n}, RV_t(h), ARMA)$	0.368	0.400	0.434	0.286	0.309	0.330	0.194	0.208	0.221
Model M3									
$R^2(IV_{t+1:t+n}, RV_t(h))$	0.843	0.886	0.918	0.797	0.839	0.869	0.717	0.754	0.781
$R^2(IV_{t+1:t+n}, RV_t(h), 1)$	0.877	0.901	0.920	0.830	0.853	0.871	0.747	0.768	0.783
$R^2(IV_{t+1:t+n}, RV_t(h), 4)$	0.889	0.904	0.920	0.841	0.856	0.871	0.757	0.770	0.784

Table 4 considers the predictability of integrated volatility over longer horizons. For models M1 and M3, the conclusions mirror those for Table 3 discussed above. For model M2, it is increasingly obvious that more frequent sampling of the intraday returns is beneficial. Likewise, for the scenarios with lower predictability—long horizons and relatively infrequent sampling of intraday returns—it is more important to include additional explanatory variables in the formation of the volatility forecasts.

Of course, the R^2s reported above cannot be mimicked by actual data, since the left-hand-side variable of interest—integrated volatility—is not observable. Feasible regressions must rely on, e.g., ex post realized volatility measures as a proxy for the realization of future integrated volatility. Since the use of such a proxy will bias the observable predictability downward, it is important to recognize the size of the potential bias. The relevant population R^2 from such feasible regressions may be derived from (35). Moreover, combining the above result with (34) allows for the derivation of the fully feasible regression R^2s based on realized volatility

TABLE 5

ONE-PERIOD-AHEAD FORECASTS OF REALIZED VOLATILITY BASED ON REALIZED VOLATILITY

Model	M1			M2			M3		
1/h	48	96	288	48	96	288	48	96	288
$R^2(RV_{t+1}(h), RV_t(h))$	0.731	0.832	0.911	0.328	0.460	0.597	0.795	0.879	0.943
$R^2(RV_{t+1}(h), RV_t(h), 1)$	0.765	0.846	0.912	0.350	0.469	0.597	0.827	0.894	0.945
$R^2(RV_{t+1}(h), RV_t(h), 4)$	0.773	0.848	0.912	0.358	0.474	0.600	0.838	0.897	0.945
$R^2(RV_{t+1}(h), RV_t(h), ARMA)$	0.773	0.848	0.912	0.360	0.475	0.601	–	–	–

TABLE 6

MULTI-PERIOD-AHEAD FORECASTS OF REALIZED VOLATILITY BASED ON REALIZED VOLATILITY

Horizon	5			10			20		
1/h	48	96	288	48	96	288	48	96	288
Model M1									
$R^2(RV_{t+1:t+n}(h), RV_t(h))$	0.740	0.801	0.847	0.671	0.722	0.759	0.546	0.585	0.614
$R^2(RV_{t+1:t+n}(h), RV_t(h), 1)$	0.774	0.815	0.848	0.702	0.734	0.760	0.571	0.595	0.615
$R^2(RV_{t+1:t+n}(h), RV_t(h), 4)$	0.782	0.816	0.848	0.709	0.735	0.760	0.577	0.597	0.615
$R^2(RV_{t+1:t+n}(h), RV_t(h), ARMA)$	0.782	0.816	0.848	0.709	0.735	0.760	0.577	0.597	0.615
Model M2									
$R^2(RV_{t+1:t+n}(h), RV_t(h))$	0.274	0.343	0.406	0.210	0.258	0.302	0.140	0.170	0.197
$R^2(RV_{t+1:t+n}(h), RV_t(h), 1)$	0.303	0.359	0.410	0.237	0.275	0.308	0.160	0.184	0.203
$R^2(RV_{t+1:t+n}(h), RV_t(h), 4)$	0.321	0.372	0.421	0.258	0.291	0.321	0.177	0.197	0.214
$R^2(RV_{t+1:t+n}(h), RV_t(h), ARMA)$	0.328	0.378	0.425	0.266	0.297	0.326	0.184	0.202	0.219
Model M3									
$R^2(RV_{t+1:t+n}(h), RV_t(h))$	0.824	0.876	0.914	0.788	0.834	0.867	0.713	0.752	0.781
$R^2(RV_{t+1:t+n}(h), RV_t(h), 1)$	0.858	0.891	0.917	0.821	0.848	0.869	0.742	0.765	0.783
$R^2(RV_{t+1:t+n}(h), RV_t(h), 4)$	0.869	0.895	0.917	0.832	0.851	0.869	0.752	0.768	0.783

proxies for integrated volatility, both as regressor and regressand. Tables 5 and 6 report on the amount of predictability associated with such feasible regressions.

Compared to Table 3, Table 5 reveals a significant loss of predictive power, even for models M1 and M3. Of course, this is purely artificial as it is induced solely by the measurement error in the integrated volatility proxy. Nonetheless, the R^2s still reveal a large amount of verifiable predictability in the integrated (realized) volatility process. As before, we also find that the impact of measurement errors is mitigated when more frequent sampling is undertaken.[16]

[16] Barndorff-Nielsen and Shephard (2002a) provide complementary one-period-ahead forecast of integrated volatility based on realized volatilities. Their (model-based) approach uses the state-space representation of the integrated volatility (combined with the Kalman filter) when the spot variance depends on autoregressive and independent factors, as in models M1 and M2. Indeed, for these two models, their one-period-ahead forecasts correspond exactly to our forecasts based on the ARMA representation of the realized volatility provided in Table 5.

Table 6 extends the results in Table 5 to longer forecast horizons. The general conclusions are reinforced, but one interesting difference from Tables 3 and 4 becomes apparent. It is now possible for the five-period-ahead forecasts to display a higher degree of predictability than the one-period-ahead forecasts, as evidenced by the results for models M1 and M3 with a sampling frequency of $1/h = 48$. This occurs because the realization of integrated volatility is approximated more accurately over five periods as opposed to just one period, relative to the true variability in the latent integrated volatility. Thus, the decline in fundamental predictability associated with a longer horizon is more than offset by the relatively smaller measurement error in the dependent variable.[17] This provides a vivid illustration of the importance of recognizing the downward bias in the true predictability induced by the need to rely on an observable proxy for the ex post integrated volatility realizations. At the same time, it is clear that the downward bias is almost nonexistent for the longer 20-period horizon, where the R^2 figures in Table 6 are very close to those in Table 4 across all models and sampling frequencies.

5.2.2. *Forecasts from past daily squared returns.* To highlight the improved "signal-to-noise ratio" achieved by employing the realized volatility measures based on high-frequency intraday data rather than the traditional approach that only relies on daily data, we finally consider the results related to the forecasts of integrated and realized volatility based on past daily squared returns, i.e., $h = 1$.

The first panel of Table 7 demonstrates that it is critical to employ long lags of daily squared returns in order to predict future (integrated) volatility with any accuracy. Consistent with the general weak GARCH principle of Drost and Werker (1996) and Meddahi (2003), it is evident how effective a simple (recursive) GARCH structure is in parsimoniously capturing the information in the lagged squared daily returns. However, even in the best of circumstances, a comparison of the upper panel in Table 7 with Tables 3 and 4 reveals that the forecast efficiency is severely curtailed by restricting the information set to the history of daily squared returns rather than the past realized volatilities constructed from the high-frequency data.

The lower panels of Table 7 provide corresponding evidence for feasible volatility forecast regressions where the integrated volatility regressor is replaced by (feasible) realized volatility approximations computed from sampling frequencies ranging from 288 intraday observations (5-minute returns covering 24 hours) to daily data. The use of daily squared returns as a one-step-ahead volatility proxy is representative of much of the empirically oriented volatility literature over the last decade. The use of cumulative squared daily returns as a volatility measure over longer weekly, monthly, and quarterly horizons has been emphasized by French et al. (1987) and Schwert (1989, 1990), among others.

For all scenarios in the lower panels of Table 7, we inevitably find an even lower degree of predictability than implied by the corresponding integrated volatility

[17] Formally, in (35) the last equation will have the fundamental predictability, $R^2(IV_{t+1:t+n}, \cdot)$, decline with n, but the ratio $\mathrm{Var}[IV_{t+1:t+n}]/\mathrm{Var}[RV_{t+1:t+n}(h)]$ will increase with n, and for lower values of $1/h$ this may actually raise the observed R^2.

TABLE 7

FORECASTS OF INTEGRATED AND REALIZED VOLATILITIES FROM PAST DAILY SQUARED RETURNS

	Model	M1				M2				M3			
	Horizon	1	5	10	20	1	5	10	20	1	5	10	20
IV	lag												
	0	0.122	0.111	0.100	0.081	0.031	0.020	0.015	0.010	0.157	0.150	0.142	0.128
	1	0.210	0.191	0.171	0.138	0.048	0.033	0.025	0.017	0.266	0.255	0.241	0.217
	4	0.360	0.329	0.294	0.238	0.072	0.054	0.043	0.029	0.452	0.432	0.409	0.369
	19	0.493	0.450	0.402	0.325	0.092	0.074	0.061	0.043	0.639	0.611	0.580	0.523
	39	0.498	0.454	0.406	0.328	0.093	0.075	0.062	0.044	0.653	0.625	0.593	0.535
	GARCH	0.498	0.454	0.406	0.328	0.093	0.075	0.062	0.044	–	–	–	–
1/h	lag												
288	0	0.119	0.110	0.099	0.080	0.029	0.019	0.015	0.009	0.154	0.149	0.142	0.128
	1	0.205	0.190	0.171	0.138	0.044	0.032	0.024	0.016	0.261	0.254	0.241	0.217
	4	0.352	0.327	0.293	0.238	0.066	0.052	0.042	0.029	0.444	0.430	0.408	0.369
	19	0.482	0.448	0.401	0.325	0.085	0.072	0.060	0.043	0.628	0.609	0.579	0.522
	39	0.486	0.452	0.405	0.328	0.086	0.074	0.061	0.043	0.641	0.623	0.592	0.534
	GARCH	0.486	0.452	0.405	0.328	0.086	0.074	0.061	0.043	–	–	–	–
96	0	0.114	0.109	0.099	0.080	0.025	0.019	0.014	0.009	0.149	0.148	0.141	0.128
	1	0.196	0.188	0.170	0.137	0.039	0.031	0.024	0.016	0.252	0.252	0.240	0.216
	4	0.336	0.324	0.292	0.237	0.058	0.050	0.040	0.028	0.429	0.427	0.407	0.368
	19	0.460	0.443	0.399	0.324	0.075	0.069	0.058	0.042	0.606	0.604	0.577	0.521
	39	0.465	0.447	0.403	0.327	0.076	0.070	0.059	0.043	0.619	0.618	0.590	0.533
	GARCH	0.465	0.447	0.403	0.327	0.076	0.070	0.059	0.043	–	–	–	–
48	0	0.107	0.108	0.098	0.080	0.021	0.018	0.014	0.009	0.142	0.147	0.140	0.127
	1	0.184	0.185	0.168	0.137	0.033	0.029	0.023	0.016	0.240	0.249	0.238	0.216
	4	0.315	0.319	0.289	0.236	0.049	0.048	0.039	0.029	0.408	0.423	0.404	0.367
	19	0.432	0.437	0.396	0.322	0.063	0.066	0.056	0.042	0.576	0.598	0.573	0.520
	39	0.436	0.441	0.400	0.325	0.064	0.067	0.057	0.043	0.589	0.611	0.586	0.532
	GARCH	0.436	0.441	0.400	0.325	0.064	0.067	0.057	0.043	–	–	–	–
1	0	0.016	0.046	0.057	0.057	0.001	0.003	0.003	0.003	0.026	0.073	0.092	0.099
	1	0.027	0.079	0.098	0.098	0.002	0.005	0.005	0.005	0.043	0.123	0.156	0.168
	4	0.046	0.136	0.168	0.167	0.003	0.008	0.009	0.008	0.073	0.209	0.264	0.286
	19	0.063	0.185	0.229	0.229	0.004	0.011	0.013	0.012	0.103	0.295	0.374	0.405
	39	0.064	0.187	0.232	0.231	0.004	0.012	0.014	0.013	0.105	0.302	0.383	0.415
	GARCH	0.064	0.187	0.232	0.231	0.004	0.012	0.014	0.013	–	–	–	–

regressions. This is an immediate consequence of Equations (34) and (35). Nonetheless, for $h = 1/288$ or $1/96$ we have sufficiently good approximations to integrated volatility that the qualitative results are very similar to those in the upper panel, and even for $h = 1/48$ the loss in observed forecast power is limited. As such, this reinforces the conclusion from above: The (observable) forecast efficiency is severely curtailed if one uses only the past daily returns in forecasting the future volatility.

The bottom panel of Table 7 further documents the extraordinarily poor coherence between future daily squared returns and the (near) optimal forecasts constructed from past daily squared returns. Again, for the case where the realized volatility proxy works very poorly, the R^2s actually increase with the forecast horizon—now often dramatically so—illustrating the significance of the measurement error in a single day's squared return relative to the corresponding volatility. It is evident across all forecast horizons that the true predictability of integrated volatility is wildly underestimated by the R^2 from these feasible regressions based on volatility measures constructed from daily data. This is consistent with the findings from the large literature trying to evaluate the performance of alternative volatility forecasts by studying the R^2 from the associated Mincer–Zarnowitz regressions using daily squared returns as the ex post volatility measure. Such studies invariably find the relative performance to be unstable and to differ across both asset classes and time periods. This is what one should expect if the dependent variable of interest—realized ex post (integrated) volatility—is measured with a large degree of imprecision. Even for long daily samples, the findings are largely random. Only by moving toward more meaningful ex post realized volatility measures for integrated volatility does it become possible to assess the forecast performance with any degree of reliability. This is, of course, exactly the point elaborated in Andersen and Bollerslev (1998). In fact, the current exposition for predictability of volatility based on daily squared returns may be seen as an analytic extension to a much broader range of models of the simulation-based investigation of the continuous-time GARCH model in Andersen and Bollerslev (1998) and Andersen et al. (1999).[18]

6. CONCLUSION

This article develops a direct analytic approach to the construction and assessment of volatility forecasts for continuous-time diffusion models within the broad ESV class of models. This class incorporates the most popular volatility diffusion models in current use, and may be calibrated to account well for the major empirical features of asset return volatility. The results provide theoretical upper bounds for the degree of predictability based on optimal (infeasible) forecasts along with direct measures of the loss in forecast efficiency associated with less precise, but more practical (feasible) reduced-form realized volatility-based procedures. As such, our results provide theoretical foundation and inspiration for the further development of new and improved easy-to-implement empirical volatility forecasting procedures guided by proper (optimal) benchmark comparisons. The insights obtained from empirical comparisons of options implied volatilities may, likewise, be improved by properly accounting for the volatility error. The techniques developed here could also be used in more effectively calibrating the type of continuous-time models routinely employed in modern asset-pricing theories. We leave further theoretical and empirical work along these lines for future research.

[18] Specifically, the entry in Table 7, Panel 1, M1, row GARCH, corresponds directly to the entry $m = \infty$ (infinitely frequent sampling) for DM-dollar in Table 4 of Andersen and Bollerslev (1998). The minor discrepancy between the numbers is due to the presence of small simulation errors in Andersen and Bollerslev (1998).

APPENDIX

We start out with the proofs of two useful lemmas.

LEMMA 1. *For any ESV model*,

$$E\left[\int_0^h P_i(f_u)\,du \int_0^h P_j(f_u)\,du\right] = \delta_{ij}\frac{2}{\lambda_i^2}[\exp(-\lambda_i h) + \lambda_i h - 1] \tag{A.1}$$

where $\delta_{ij} = 1$ *if* $i = j$ *and* $\delta_{ij} = 0$ *if* $i \neq j$. *Moreover, for* $n \geq 1$,

$$\begin{aligned} E\left[\int_0^h P_i(f_u)\,du \int_{nh}^{(n+1)h} P_j(f_u)\,du\right] \\ = \delta_{ij}\exp(-\lambda_j(n-1)h)\frac{[1-\exp(-\lambda_i h)]^2}{\lambda_i^2} \end{aligned} \tag{A.2}$$

PROOF OF LEMMA 1. Fubini's Theorem implies that

$$\int_0^h P_i(f_u)\,du \int_0^h P_j(f_u)\,du = \int_0^h \int_0^h P_i(f_u)P_j(f_s)\,du\,ds,$$

for a given path of the process f_u, $u \in [0, h]$. In addition, we have

$$\begin{aligned} \int_0^h \int_0^h P_i(f_u)P_j(f_s)\,du\,ds &= \int_0^h P_i(f_u)\left(\int_0^u P_j(f_s)\,ds\right)du \\ &\quad + \int_0^h P_j(f_u)\left(\int_0^u P_i(f_s)\,ds\right)du \end{aligned}$$

Hence,

$$\begin{aligned} &E\left[\int_0^h P_i(f_u)\,du \int_0^h P_j(f_u)\,du\right] \\ &= \int_0^h \left(\int_0^u E[P_i(f_u)P_j(f_s)]\,ds\right)du + \int_0^h \left(\int_0^u E[P_j(f_u)P_i(f_s)]\,ds\right)du \\ &= \int_0^h \left(\int_0^u \exp(-\lambda_i(u-s))E[P_i(f_s)P_j(f_s)]\,ds\right)du \\ &\quad + \int_0^h \left(\int_0^u \exp(-\lambda_j(u-s))E[P_j(f_s)P_i(f_s)]\,ds\right)du, \\ &= 2\delta_{ij}\int_0^h \left(\int_0^u \exp(-\lambda_i(u-s))\,ds\right)du = \delta_{ij}\frac{2}{\lambda_i^2}[\exp(-\lambda_i h) + \lambda_i h - 1] \end{aligned}$$

where the second and third equalities follow from (10) and (9), respectively. Equation (A.2) follows from

$$
\begin{aligned}
& E\left[\int_0^h P_i(f_u)\,du \int_{nh}^{(n+1)h} P_j(f_u)\,du\right] \\
&\quad = E\left[\int_0^h P_i(f_u)\,du \int_{nh}^{(n+1)h} E[P_j(f_u) \mid f_\tau, \tau \le h]\,du\right] \\
&\quad = E\left[\int_0^h P_i(f_u)\,du \int_{nh}^{(n+1)h} \exp(-\lambda_j(u-h))P_j(f_h)\,du\right] \\
&\quad = \exp(-\lambda_j(n-1)h)\frac{1-\exp(-\lambda_i h)}{\lambda_i} \int_0^h E[P_i(f_u)P_j(f_h)]\,du \\
&\quad = \delta_{ij}\exp(-\lambda_j(n-1)h)\frac{[1-\exp(-\lambda_i h)]^2}{\lambda_i^2}. \qquad \blacksquare
\end{aligned}
$$

LEMMA 2. *Let the ARMA(2,2) model for y_t be denoted by*

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \eta_t + \theta_1 \eta_{t-1} + \theta_2 \eta_{t-2}$$

where η_t is a weak white noise. Then,

$$R^2(y_{t+1:t+n}, \text{ARMA}) = 1 - \frac{\text{Var}[y_{t+1:t+n} - BP[y_{t+1:t+n} \mid H_t(y)]]}{\text{Var}[y_{t+1:t+n}]} \tag{A.3}$$

with

$$\text{Var}[y_{t+1:t+n} - BP[y_{t+1:t+n} \mid H_t(y)]] = \sum_{i=0}^{n-1}\left(\sum_{s=0}^{i}\psi_s\right)^2 \text{Var}[\eta_t] \tag{A.4}$$

where $\psi_i = A_1^\top \Phi^i A_2$, with $A_1 = (1,0,0,0)^\top$, $A_2 = (1,0,1,0)^\top$, Φ is a 4×4 matrix with $\Phi[1,1] = \phi_1$, $\Phi[1,2] = \phi_2$, $\Phi[1,3] = \theta_1$, $\Phi[1,4] = \theta_2$, $\Phi[2,1] = 1$, $\Phi[4,3] = 1$, and 0 otherwise, $H_t(z)$ denotes the Hilbert-space generated by $\{1, z_\tau, \tau \le t\}$, whereas for a second-order stationary variable w, $BP[w \mid H_t(z)]$ denotes the best linear predictor of w given $H_t(z)$; i.e., $w = BP[w \mid H_t(z)] + \varepsilon$, with $\text{Cov}(\varepsilon, x) = 0$, $\forall x \in H_t(z)$.

PROOF OF LEMMA 2. Following Baillie and Bollerslev (1992),

$$y_{t+n} - BP[y_{t+n} \mid H_t(y)] = \sum_{i=0}^{n-1} \psi_i \eta_{t+n-i}$$

for any $n > 0$. Therefore, one obtains the following two equalities:

$$y_{t+1:t+n} - BP[y_{t+1:t+n} \mid H_t(y)] = \sum_{i=0}^{n-1} \left(\sum_{s=0}^{i} \psi_s \right) \eta_{t+n-i}$$

$$\mathrm{Var}[y_{t+1:t+n} - BP[y_{t+1:t+n} \mid H_t(y)]] = \sum_{i=0}^{n-1} \left(\sum_{s=0}^{i} \psi_s \right)^2 \mathrm{Var}[\eta_t]$$

i.e., (A.4). Note that although the results of Baillie and Bollerslev (1992) are derived under the assumption that η_t is a martingale difference sequence and refers to $E[y_{t+n} \mid y_\tau, \tau \leq t]$ rather than $BP[y_{t+n} \mid H_t(y)]$, the final result remains valid when η_t is a weak white noise process (as for the integrated and realized volatilities); see also Meddahi and Renault (2004) and Meddahi (2003) for further discussion along these lines. Finally, observe that when $\phi_2 = \theta_2 = 0$, i.e., y_t is an ARMA(1,1), (A.3) becomes

$$R^2(y_{t+1:t+n}, \mathrm{ARMA}) = \frac{(1-\phi_1^n)^2}{(1-\phi_1)^2} \frac{\mathrm{Var}[y_t]}{\mathrm{Var}[y_{t+1:t+n}]} R^2(y_{t+1}, \mathrm{ARMA}) \tag{A.5}$$

Proof of Proposition 1. Given that the unconditional mean of any eigenfunction $P_i(\cdot)$, with $i \geq 1$, is zero, one gets (12). Let s be a positive real; then we have

$$E\big[\sigma^2_{t+s} \bigm| S_\tau, f_\tau, \tau \leq t\big] = \sum_{i=0}^{p} a_i E[P_i(f_{t+s}) \mid f_t] = a_0 + \sum_{i=1}^{p} a_i \exp(-\lambda_i s) P_i(f_t) \tag{A.6}$$

which corresponds to (13) for $s = n$. By using (A.6), one can show easily

$$E[IV_{t+n} \mid S_\tau, f_\tau, \tau \leq t] = a_0 + \sum_{i=1}^{p} a_i \exp(-\lambda_i(n-1)) \frac{(1-\exp(-\lambda_i))}{\lambda_i} P_i(f_t) \tag{A.7}$$

Therefore,

$$\begin{aligned}
&E[IV_{t+1:t+n} \mid S_\tau, f_\tau, \tau \leq t] \\
&\quad = na_0 + \sum_{s=1}^{n} \sum_{i=1}^{p} a_i \exp(-\lambda_i(s-1)) \frac{[1-\exp(-\lambda_i)]}{\lambda_i} P_i(f_t) \\
&\quad = na_0 + \sum_{i=1}^{p} a_i \left(\sum_{s=1}^{n} \exp(-\lambda_i(s-1)) \right) \frac{[1-\exp(-\lambda_i)]}{\lambda_i} P_i(f_t) \\
&\quad = na_0 + \sum_{i=1}^{p} a_i \frac{[1-\exp(-\lambda_i n)]}{\lambda_i} P_i(f_t)
\end{aligned}$$

i.e., (14). The first result in (15) follows from the orthonormality of the eigenfunction, indicated in (9). In addition, for any real $s \geq 0$

$$\begin{aligned}
\operatorname{Var}\left[\int_{t-1}^{t-1+s} \sigma_u^2 du\right] &= E\left[\left(\sum_{i=1}^{p} a_i \int_{t-1}^{t-1+s} P_i(f_u)\, du\right)^2\right] \\
&= \sum_{1\leq i,j\leq p} a_i a_i\, E\left[\int_{t-1}^{t-1+s} P_i(f_u)\, du \int_{t-1}^{t-1+s} P_j(f_u)\, du\right] \\
&= \sum_{1\leq i,j\leq p} a_i a_i \delta_{ij} \frac{2}{\lambda_i^2}[\exp(-\lambda_i s) + \lambda_i s - 1]
\end{aligned}$$

where the last equality follows from (A.1). Thus,

$$\operatorname{Var}\left[\int_{t-1}^{t-1+s} \sigma_u^2 du\right] = 2\sum_{i=1}^{p} \frac{a_i^2}{\lambda_i^2}[\exp(-\lambda_i s) + \lambda_i s - 1] \tag{A.8}$$

which corresponds to the second result in (15) for $s = n$. By using the orthonormality of the eigenfunctions, Equations (A.7), and the equality, $\operatorname{Cov}(IV_{t+n}, \sigma_t^2) = \operatorname{Cov}(E[IV_{t+n} \mid S_\tau,\ f_\tau, \tau \leq t], \sigma_t^2)$, it follows that

$$\operatorname{Cov}\big(IV_{t+n}, \sigma_t^2\big) = \sum_{i=1}^{p} a_i^2 \exp(-\lambda_i(n-1)) \frac{[1-\exp(-\lambda_i)]}{\lambda_i} \tag{A.9}$$

Hence,

$$\begin{aligned}
\operatorname{Cov}\big(IV_{t+1:t+n}, \sigma_{t-l}^2\big) &= \sum_{s=1}^{n}\sum_{i=1}^{p} a_i^2 \exp(-\lambda_i(s+l-1)) \frac{[1-\exp(-\lambda_i)]}{\lambda_i} \\
&= \sum_{i=1}^{p} a_i^2 \exp(-\lambda_i l) \left(\sum_{s=1}^{n} \exp(-\lambda_i(s-1))\right) \frac{[1-\exp(-\lambda_i)]}{\lambda_i} \\
&= \sum_{i=1}^{p} a_i^2 \frac{[1-\exp(-\lambda_i n)]}{\lambda_i} \exp(-\lambda_i l)
\end{aligned}$$

i.e., (16). Equation (18) is obtained by a similar argument. By using (A.2), one gets

$$\begin{aligned}
\text{(A.10)}\qquad \mathrm{Cov}(IV_t, IV_{t+n}) &= \mathrm{Cov}(IV_1, IV_{1+n})\\
&= E\left[\left(\sum_{i=0}^{p} a_i \int_0^1 P_i(f_u)\,du\right)\left(\sum_{i=1}^{p} a_i \int_n^{n+1} P_i(f_u)\,du\right)\right]\\
&= \sum_{1\le i,j\le p} a_i a_j E\left[\int_0^1 P_i(f_u)\,du \int_1^{n+1} P_j(f_u)\,du\right]\\
&= \sum_{1\le i,j\le p} a_i a_j \delta_{ij} \exp(-\lambda_i(n-1))\frac{[1-\exp(-\lambda_i)]^2}{\lambda_i^2}\\
&= \sum_{i=1}^{p} a_i^2 \exp(-\lambda_i(n-1))\frac{[1-\exp(-\lambda_i)]^2}{\lambda_i^2}
\end{aligned}$$

Utilizing (A.10) in a similar set of arguments, it is possible to show (17). ■

PROOF OF PROPOSITION 2. By straightforward calculations, it is easy to show

$$\text{(A.11)}\qquad \forall\lambda > 0,\ \frac{[1-\exp(-\lambda)]^2}{\exp(-\lambda)} \ge \lambda^2 \ge 2[\exp(-\lambda)+\lambda-1]$$

$$\text{(A.12)}\qquad \frac{2[\exp(-\lambda)+\lambda-1]}{[1-\exp(-\lambda)]} \ge \lambda \ge [1-\exp(-\lambda)]$$

Thus, by using the first inequality in (A.11), it follows that for any $n \ge 1$,

$$\exp(-\lambda n) \le \exp(-\lambda(n-1))[1-\exp(-\lambda)]^2\lambda^{-2}$$

so that $\mathrm{Cov}(\sigma_{t+n}^2, \sigma_t^2) \le \mathrm{Cov}(IV_{t+n}, IV_t)$. Also, by using the second inequality in (A.12), one gets $[1-\exp(-\lambda)]^2\lambda^{-2} \le [1-\exp(-\lambda)]\lambda^{-1} \le 1$. Therefore,

$$\begin{aligned}
\forall n \ge 1,\ &\exp(-\lambda(n-1))[1-\exp(-\lambda)]^2\lambda^{-2}\\
&\le \exp(-\lambda(n-1))[1-\exp(-\lambda)]\lambda^{-1} \le 1,
\end{aligned}$$

and, hence, $\mathrm{Cov}(IV_{t+n}, IV_t) \le \mathrm{Cov}(IV_{t+n}, \sigma_t^2)$. The first inequality in (A.12) implies

$$\begin{aligned}
\forall n \ge 1,\ &\exp(-\lambda(n-1))[1-\exp(-\lambda)]\lambda^{-1}\\
&\le [1-\exp(-\lambda)]\lambda^{-1} \le 2[\exp(-\lambda)+\lambda-1]\lambda^{-2},
\end{aligned}$$

and, hence, $\mathrm{Cov}(IV_{t+n}, \sigma_t^2) \le \mathrm{Var}[IV_t]$. Finally, by the second inequality in (A.11), $\mathrm{Var}[IV_t] \le \mathrm{Var}[\sigma_t^2]$, which completes the proof of Proposition 2. ■

PROOF OF (20). $R^2(IV_{t+1}, Best)$ equals $\mathrm{Var}[E[IV_{t+1} \mid S_\tau, f_\tau, \tau \leq t]]/\mathrm{Var}[IV_{t+1}]$. By using (A.7) for $n = 1$ along with the orthonormality of the eigenfunctions, i.e., (9), we get $\mathrm{Var}[E[IV_{t+1} \mid S_\tau, f_\tau, \tau \leq t]] = \sum_{i=1}^{p} a_i^2[1 - \exp(-\lambda_i)]^2\lambda_i^{-2}$, which combines to show (20). ■

PROOF OF (21) AND (22). By definition, we have

$$R^2(IV_{t+1}, \sigma_t^2) = \mathrm{Cov}(IV_{t+1}, \sigma_t^2)^2/\mathrm{Var}[IV_{t+1}]\mathrm{Var}[\sigma_t^2]$$

Thus, by using (16) for $n = 1$, we get (21). A similar argument leads to (22). ■

Formulas for R^2s Based on Multiple Explanatory Variables.

$$R^2(IV_{t+1}, IV_t, l) = C(IV_{t+1}, IV_t, l)^\top (M(IV_t, l))^{-1} C(IV_{t+1}, IV_t, l)/\mathrm{Var}[IV_t]$$

$$R^2(IV_{t+1:t+n}, \sigma_t^2, l) = \frac{C(IV_{t+1:t+n}, \sigma_t^2, l)^\top M(\sigma_t^2, l)^{-1} C(IV_{t+1:t+n}, \sigma_t^2, l)}{\mathrm{Var}[IV_{t+1:t+n}]}$$

$$R^2(IV_{t+1:t+n}, IV_t, l) = \frac{C(IV_{t+1:t+n}, IV_t, l)^\top M(IV_t, l)^{-1} C(IV_{t+1:t+n}, IV_t, l)}{\mathrm{Var}[IV_{t+1:t+n}]}$$

Formulas for R^2s Based on Multi-Step-Ahead Integrated Volatility. These follow by a simple application of Lemma A.2. The ARMA representation of integrated volatility is given in Meddahi (2003). ■

PROOF OF PROPOSITION 3. Let y be a second-order stationary variable such that $\mathrm{Cov}(y, U_t(h)) = 0$. Then

$$R^2(y, RV_t(h)) = \frac{\mathrm{Cov}(y, RV_t(h))^2}{\mathrm{Var}[y]\mathrm{Var}[RV_t(h)]} = \frac{\mathrm{Cov}(y, IV_t + U_t(h))^2}{\mathrm{Var}[y]\mathrm{Var}[RV_t(h)]}$$

$$= \frac{\mathrm{Cov}(y, IV_t)^2}{\mathrm{Var}[y]\mathrm{Var}[IV_t]} \frac{\mathrm{Var}[IV_t]}{\mathrm{Var}[RV_t(h)]}$$

i.e., (34). Furthermore, for any $n \geq 1$ and $h > 0$, $BP[RV_{t+1:t+n}(h) \mid H_t(RV(h))]$ equals $BP[IV_{t+1:t+n} \mid H_t(RV(h))]$ given that $RV_{t+i}(h) = IV_{t+i} + U_{t+i}(h)$, whereas $U_{t+i}(h), i \geq 1$, is uncorrelated with any variable in $H_t(RV(h))$. Note that the same results hold if one considers the best predictor (of $RV_{t+1:t+n}(h)$ or $IV_{t+1:t+n}$) given lags of $RV_t(h)$ (and a constant). Hence, $BP[RV_{t+1:t+n}(h) \mid .] = BP[IV_{t+1:t+n} \mid .]$, and

$$
\begin{aligned}
R^2(RV_{t+1:t+n}(h), .) &= \frac{\mathrm{Var}[BP[RV_{t+1:t+n}(h) \mid .]]}{\mathrm{Var}[RV_{t+1:t+n}(h)]} = \frac{\mathrm{Var}[BP[IV_{t+1:t+n} \mid .]]}{\mathrm{Var}[RV_{t+1:t+n}(h)]} \\
&= \frac{\mathrm{Var}[BP[IV_{t+1:t+n} \mid .]]}{\mathrm{Var}[IV_{t+1:t+n}]} \frac{\mathrm{Var}[IV_{t+1:t+n}]}{\mathrm{Var}[RV_{t+1:t+n}(h)]} \\
&= R^2(IV_{t+1:t+n}, .) \frac{\mathrm{Var}[IV_{t+1:t+n}]}{\mathrm{Var}[IV_{t+1:t+n}] + n\mathrm{Var}[U_t(h)]},
\end{aligned}
$$

i.e., (35). ■

REFERENCES

Aït-Sahalia, Y., L. P. Hansen, and J. Scheinkman, "Operator Methods for Continuous-Time Markov Models," in Y. Ait-Sahalia and L. P. Hansen, eds., *Handbook of Financial Econometrics* (Amsterdam: North Holland Press, 2005), forthcoming.

Alizadeh, S., M. Brandt, and F. Diebold, "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance* 57 (2002), 1047–91.

Andersen, T. G., and T. Bollerslev, "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review* 39 (1998), 885–905.

——, L. Benzoni, and J. Lund, "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance* 57 (2002), 1239–84.

——, T. Bollerslev, and F. X. Diebold, "Parametric and Nonparametric Measurements of Volatility," in Y. Aït-Sahalia and L. P. Hansen, eds., *Handbook of Financial Econometrics* (Amsterdam: North Holland Press, 2005), forthcoming.

——, ——, and S. Lange, "Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon," *Journal of Empirical Finance* 6 (1999), 457–77.

——, ——, F. X. Diebold, and P. Labys, "Modeling and Forecasting Realized Volatility," *Econometrica* 71 (2003), 579–625.

——, ——, ——, and ——, "The Distribution of Exchange Rate Volatility," *Journal of the American Statistical Association* 96 (2001), 42–55.

Andreou, E., and E. Ghysels, "Rolling-Sampling Volatility Estimators: Some New Theoretical, Simulation and Empirical Results," *Journal of Business and Economic Statistics* 20 (2002), 363–76.

Baillie, R. T., and T. Bollerslev, "Prediction in Dynamic Models with Time-Dependent Conditional Variances," *Journal of Econometrics* 52 (1992), 91–113.

Barndorff-Nielsen, O. E., and N. Shephard, "Non-Gaussian OU Based Models and Some of Their Uses in Financial Economics" (with discussion), *Journal of the Royal Statistical Society B* 63 (2001), 167–241.

——, and ——, "Econometric Analysis of Realised Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society B* 64 (2002a), 253–80.

——, and ——, "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics* 5 (2002b), 457–77.

——, and ——, *Financial Volatility and Levy Based Models*, Nuffield College, Oxford University, 2004, forthcoming.

Billingsley, P., *Probability and Measure* (New York: John Wiley & Sons, 1986).

Bollerslev, T., and H. Zhou, "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility," *Journal of Econometrics* 109 (2002), 33–65.

Campbell, B., N. Meddahi, and E. Sentana, "Understanding Long-Horizon Predictability of Asset Returns," Université of Montréal Working Paper, 2004.

CHEN, X., L. P. HANSEN, AND J. SCHEINKMAN, "Principal Components and the Long Run," University of Chicago Working Paper, 2000.

CHONG, Y. Y., AND D. HENDRY, "Econometric Evaluation of Linear Macro-Economic Models," *Review of Economic Studies* 53 (1986), 671–90.

CHRISTOFFERSEN, P. F., AND F. X. DIEBOLD, "How Relevant is Volatility Forecasting for Financial Risk Management," *Review of Economics and Statistics* 81 (2000), 12–22.

COMTE, F., AND E. RENAULT, "Long Memory in Continuous Time Stochastic Volatility Models," *Mathematical Finance* 8 (1998), 291–323.

DROST, F. C., AND T. E. NIJMAN, "Temporal Aggregation of GARCH Processes," *Econometrica* 61 (1993), 909–27.

——, AND B. J. M. WERKER, "Closing the GARCH Gap: Continuous Time GARCH Modeling," *Journal of Econometrics* 74 (1996), 31–58.

ENGLE, R. F., "The Econometrics of Ultra-High Frequency Data," *Econometrica* 68 (2000), 1–22.

——, AND G. G. J. LEE, "A Permanent and Transitory Component Model for Stock Return Volatility," in R. F. Engle and H. White, eds., *Cointegration, Causality and Forecasting: a Festschrift in Honour of Clive W.J. Granger* (New York: Oxford University Press, 1999), 475–97.

FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH, "Expected Stock Returns and Volatility," *Journal of Financial Economics* 19 (1987), 3–29.

GALLANT, R., AND G. TAUCHEN, "Simulated Score Methods and Indirect Inference for Continuous-Time Models," in Y. Aït-Sahalia and L. P. Hansen, eds., *Handbook of Financial Econometrics* (Amsterdam: North Holland Press, 2005), forthcoming.

——, C. HSU, AND G. TAUCHEN, "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *Review of Economics and Statistics* 81 (1999), 617–31.

GARCIA, R., M. A. LEWIS, AND E. RENAULT, "Estimation of Objective and Risk-Neutral Distributions Based on Moments of Integrated Volatility," Working Paper, University of Montréal, 2001.

HANSEN, L. P., AND J. SCHEINKMAN, "Back to the Future: Generating Moment Implications for Continuous Time Markov Processes," *Econometrica* 63 (1995), 767–804.

HARVEY, A. C., E. RUIZ, AND N. SHEPHARD, "Multivariate Stochastic Variance Models," *Review of Economic Studies* 61 (1994), 247–64.

HESTON, S. L., "A Closed Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies* 6 (1993), 327–44.

HOFFMAN-JØRGENSEN, J., *Probability with a View Toward Statistics*, Volume 1 (New York: Chapman & Hall, 1994).

HULL, J., AND A. WHITE, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42 (1987), 281–300.

JOHANNES, M., AND N. POLSON, "Numerical Bayesian Methods for Estimating Continuous-Time Models," in Y. Aït-Sahalia and L. P. Hansen, eds., *Handbook of Financial Econometrics* (Amsterdam: North Holland Press, 2005), forthcoming.

MEDDAHI, N., "An Eigenfunction Approach for Volatility Modeling," CIRANO Working Paper 2001s-70, 2001.

——, "A Theoretical Comparison Between Integrated and Realized Volatility," *Journal of Applied Econometrics* 17 (2002), 479–508.

——, "ARMA Representation of Integrated and Realized Variances," *The Econometrics Journal* 6 (2003), 334–55.

——, AND E. RENAULT, "Temporal Aggregation of Volatility Models," *Journal of Econometrics* 119 (2004), 355–79.

MINCER, J., AND V. ZARNOWITZ, "The Evaluation of Economic Forecasts," in J. Mincer, ed., *Economic Forecasts and Expectation* (New York: National Bureau of Research, 1969), 3–46.

NELSON, D. B., "ARCH Models as Diffusion Approximations," *Journal of Econometrics* 45 (1990) 7–39.

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

——, "Why Does Stock Market Volatility Change Over Time?" *Journal of Finance* 44 (1989), 1115–54.

SZEGO, G., *Orthogonal Polynomials*, 4th edition (Providence, RI: American Mathematical Society, 1975).

WIGGINS, J. B., "Options Values under Stochastic Volatility: Theory and Empirical Estimates," *Journal of Financial Economics* 19 (1987), 351–72.

WONG, E., "The Construction of a Class of Stationary Markoff Processes," in R. Belleman, ed., *Sixteenth Symposium in Applied Mathematics, Stochastic Processes in Mathematical Physics and Engineering* (Providence, RI: American Mathematical Society, 1964), 264–76.