Integrated Variance Forecasting: Model-Based vs. Reduced-Form

Natalia Sizova*

Abstract

This paper compares model-based and reduced-form forecasts of financial volatility when high-frequency return data are available. We derived exact formulas for the forecast errors and analyzed the contribution of the “wrong” data modeling and errors in forecast inputs. The comparison is made for “feasible” forecasts, i.e., we assumed that the true data generating process, latent states and parameters are unknown. As an illustration, the same comparison is carried out empirically for spot 5-minute returns of DM/USD exchange rates.

It is shown that the comparison between feasible reduced-form and model-based forecasts is not always in favor of the latter in contrast to their infeasible versions. The reduced-form approach is generally better for long-horizon forecasting and for short-horizon forecasting in the presence of microstructure noise.

Keywords: volatility forecasting, high-frequency data, reduced-form methods, model misspecification.

JEL classification: C22 C53
1 Introduction

Traditionally, researchers who wanted to extract and forecast financial volatility had to rely on data recorded at only moderate intervals: daily, for instance, or even monthly. But recently, data at much more frequent intervals – high-frequency data – have become increasingly available.

The increasing availability of high-frequency data allows researchers to improve on the techniques used to forecast volatility. Two types of these techniques are model-based, which construct efficient volatility forecasts that rely on the model for returns, and reduced-form, which construct simple projections of volatility on past volatility measures. Both types had been developed before the high-frequency data became available. Initially, however, only model-based techniques were considered sufficiently reliable, as they performed more accurately for daily data. For instance, two classes of volatility models – ARCH models \(^1\) by Bollerslev and Engle (1986) and stochastic volatility models \(^2\) – were conventionally preferred for extracting the volatility series. These model-based approaches gave more accurate estimates for latent volatility and provided better ways to form forecasts of the future volatility. Reduced-form techniques, on the other hand, relied on excessively noisy proxies of the volatility, e.g., daily squared returns.

Now that high-frequency data are available, is it still the case that model-based forecasts are better than reduced-form ones? Is it possible that when using high-frequency data, reduced-form forecasts are just as good as, or perhaps better than, model-based forecasts? Employing high-frequency data increases the efficiency of extracting latent spot variances by model-based techniques; hence, model-based techniques can perform better. On the other hand, for reduced-form techniques, high-frequency data allow researchers to define better proxies for the daily volatility. For example, one nonparametric proxy – the realized variance – is a sum of squared returns. It is observable, and hence it admits a direct modeling, as in Corsi (2004), and Andersen et al. (2003).

Apparently, the availability of high-frequency data starts a new chapter in the contest between reduced-form and model-based approaches as the most efficient for forecasting volatility. In this paper, we compare these two approaches for observed data and simulated data. Furthermore, we investigate this comparison analytically.

The object of interest is a forecast of the integrated variance. Integrated variance is a natural descriptor of the volatility of daily returns. It is an analog of the variance of a daily return in a discrete-time model, e.g., extracted variance in the GARCH model. Integrated variance can be used in VaR models of risk management, as an input to option pricing models and for variance hedging in trading; see Andersen et al. (2006)

The primary goal of the current study is to compare analytically feasible model-based and reduced-form forecasts of integrated variance. The defining word here is “feasible”, i.e., the comparison is carried out under “realistic” conditions, assuming that the true data generating process is unknown, and forecast inputs are estimated with errors.

To guarantee results that have the greatest generality, we formulate this comparison analytically within the framework of Meddahi’s ESV models. The groundbreaking result of Meddahi is that any square-integrable variance process can be decomposed into the sum of simple processes. This decomposition allows us to write down many of our results in analytical form. Therefore, rather than resorting to time consuming simulations to compare forecasts, we can plug parameters into a formula and immediately evaluate the comparative performance.

The forecasts to be compared are briefly defined as follows. The first one is model-based. To implement this forecast, we model returns by a stochastic volatility (SV) model. Daily data can be used to estimate the parameters of the model. We then use this model to predict the future

---

\(^1\)ARCH and GARCH models are reviewed by Andersen, Bollerslev, Christoffersen and Diebold (2005).

\(^2\)Reviewed by Tauchen (2004).
integrated variance. For example, SV-models that can be used to form the forecast include affine models by Heston (1993) and by Duffie, Pan and Singleton (2000), CEV-models, and exponential models as in Chernov et al. (2003).

The second forecast is reduced-form. In this case, the predictor of the future integrated variance is a linear function of its historic values. This form of the forecast is based on theory developed by Meddahi (2003).

The current study is logically connected to the paper by Andersen, Bollerslev and Meddahi (2004). They compared the performance of “infeasible” model-based forecasts to the performance of reduced-form forecasts. Their interesting finding is that the reduced-form forecast performs remarkably well even though the “infeasible” model-based forecast minimizes forecast error by definition. In their study, the authors assumed that a true model is known, and all the inputs to the forecasts are observed. The major difference between their work and the current study is that we take things a step further and consider the case in which the model and parameters are unknown.

Another paper that relates to the current study is the paper by Barndorff-Nielsen and Shephard (2002). Among other things, the authors looked at the problem of extracting the integrated variance using prior knowledge about a model. They compared this model-based estimate of the integrated variance to the reduced-form estimate – realized variance. They found that the results of the comparison are affected by sampling frequency and volatility-of-volatility.

In keeping with the aforementioned work we show that the sampling frequency and the volatility-of-volatility parameters are major factors in comparison between model-based and reduced-form forecasts of integrated variance. Furthermore, we present theoretical reasons for supposing that these parameters necessarily affect the comparison.

The paper is organized as follows. In Section 2, the model framework is set up and the forecasts to be compared are defined. In the same section, we introduce definitions for error components. These definitions will be used throughout the paper. In Section 3, we make analytical comparisons of feasible forecasts. The analysis includes the cases of long-horizon forecasting and forecasting using high-frequency data contaminated by microstructure noise. In Section 4 the theoretical findings from the previous part will be evaluated for observed data. Section 5 concludes the paper.

2 One Model and Two Forecasts

2.1 Model

Throughout the paper we assume the following dynamics for the logarithm of prices $s_t$. Denote $x_t$ as an $n$-dimensional vector of independent states. The $W_{i,t}, i = 1,...,n$, are independent standard Brownian motions. $W^s_t$ is a standard Brownian motion which may correlate with $W_{i,t}, i = 1,...,n$, i.e., $\text{corr}(dW^s_t,dW_{i,t}) = \rho_i(x_t)$.

The dynamics of states $x_t$ and log-prices $s_t$ are described by the following system of equations:

$$
\begin{bmatrix}
    ds_t \\
    dx_{1,t} \\
    ... \\
    dx_{n,t}
\end{bmatrix} =
\begin{bmatrix}
    \mu(x_t) \\
    \kappa_1(x_{1,t}) \\
    ... \\
    \kappa_n(x_{n,t})
\end{bmatrix} dt +
\begin{bmatrix}
    \sigma(x_t)dW^s_t \\
    \Lambda_1(x_{1,t})dW_{1,t} \\
    ... \\
    \Lambda_n(x_{n,t})dW_{n,t}
\end{bmatrix}
$$

where $\sigma^2(x_t)$ is referred to as a spot variance. The functions $\Lambda_i(x_{i,t})$ stand for diffusion terms in the dynamics of the states $x_t$, and the functions $\kappa_i(x_{i,t})$ are drifts. For notational simplicity, denote $\sigma^2(x_t) \equiv \sigma^2_t$. 

3
We are interested in forecasting the ex-post variability of returns as measured by the integrated variance:

$$IV_t^{t+T} \equiv \int_t^{t+T} \sigma_s^2 ds,$$

which is assumed to be well-defined. Intuitively, IV defines the variance of the T-period return, i.e., $\text{Var}(s_{t+T} - s_t)$, if the drift in prices $\mu(x_t)$ is a predictable process of finite variation. In general, the same relation holds approximately, since the variation in the drift $\mu(x_t)$ is negligible in comparison to the variation in the diffusion part $\sigma_t dW^s_t$. This implies that by forecasting the integrated variance, we seek to forecast the variability of the asset price over the next T periods. For example, variances that are inputs to Sharpe ratios of the assets can be taken from the values of $IV_t^{t+T}$.

Note that the system (1) has a very general form. Any known Markov continuous dynamics of stochastic volatility can be represented in this form, e.g., affine, GARCH-SV, and log-volatility models, to be defined more formally below.

For the purpose of later derivations, we will use another representation of the same system (1). This representation is referred to as the ESV representation and was introduced by Meddahi (2001). He showed that any square integrable spot variance $\sigma^2(x_t)$ appearing in the SDE system (1) admits an Eigenfunction Stochastic Volatility (ESV) representation:

$$\sigma^2(x_t) = a_0 + \sum_{i=1}^{\infty} a_i P_i(x_t),$$

where processes $P_i(x_t)$ are called factors. These are square-integrable processes with the following properties:

(i) zero-mean: $\text{EP}_i(x_t) = 0$;

(ii) uncorrelated, with a unit variance: $\text{Cov}(P_i(x_t), P_j(x_t)) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$;

(iii) if discretely observed, each factor follows an AR(1) process:

$$\text{E}(P_i(x_{t+T})|x_\tau, \tau \leq t) = e^{-k_i T} P_i(x_t).$$

In the following discussion, a process will be referred to as a one-factor process if $a_i = 0, \forall i > 1$, and two-factor if $a_i = 0, \forall i > 2$. In general, for a p-factor process the variance can be represented by the sum, $\sigma^2(x_t) = a_0 + \sum_{i=1}^{p} a_i P_i(x_t)$. Note, that the number of uncorrelated factors $p$ in the ESV representation is generally not the same as the number of independent processes that drive the dynamics of volatility. We emphasize that the above representation is equivalent to the representation (1), and was introduced only because it facilitates further analytical derivations.

We now define the two types of forecasts for study. The first one will be called “model-based” or simply the “best” forecast. The second will be referred to as “model-free” or “reduced-form”.

2.2 The Model-Based Forecast

In this section, we define the model-based forecast. In addition, we show how the prediction error from this model-based forecast can be decomposed into three components: “genuine” forecast error, error from estimates of the parameters and states, and error from model misspecification.
The model-based forecast is defined as the “best” forecast in terms of the mean squared error (MSE). That is, given the information set \( F_t \) at time \( t \), the model-based forecast of the integrated variance \( IV_{t+1} \) minimizes the square-loss function:

\[
IV_{t+1, t}^{model} = \arg \min_{IV_{t+1} \in F_t} E(IV_{t+1} - IV_{t+1}^t | F_t)^2. \tag{5}
\]

Despite a controversy surrounding the choice of the most appropriate performance measure (see Patton and Timmerman, 2007a, b), mean squared error remains a common reference point for a forecast performance.

The function that satisfies condition (5) is the expectation of \( IV_{t+1} \) conditional on information at time \( t \), i.e., \( E(IV_{t+1} | F_t) \). For the model described by system (1), let \( F_t \) be a \( \sigma \)-algebra generated by prices and states up to time \( t \), i.e., \( p_\tau, x_\tau, \forall \tau \leq t \). Due to the Markovian dynamics of log-prices and variances given by the system (1), the model-based forecast depends only on the latest realization of the state

\[
IV_{t+1|t}^{model} = E(IV_{t+1|t} | x_t), \tag{6}
\]

and the model specification (1). The latter classifies it as “model-based”. The ESV framework (3) yields a closed-form representation for the model-based forecast:

\[
IV_{t+1|t}^{model} = \int_t^{t+1} \left[ a_0 + \sum_{i=1}^{\infty} a_i e^{-k_i(s-t)} P_i(x_t) \right] ds = a_0 + \sum_{i=1}^{\infty} a_i \frac{1 - e^{-k_i}}{k_i} P_i(x_t). \tag{7}
\]

However, this forecast is infeasible, since it is based on an unknown model, parameters and states. To define the feasible version, it is common to proceed as follows. First, we choose a model and derive a corresponding formula for the model-based forecast. We can then estimate the parameters and unobserved states using a general method such as MLE, exact or an approximation. Finally, we substitute recovered states and parameters into the formula. This resulting structure is termed a feasible model-based forecast. Its form can be derived from the ESV representation:

\[
IV_{t+1|t}^{model} = \tilde{a}_0 + \sum_{i=1}^{\infty} \tilde{a}_i \frac{1 - e^{-\tilde{k}_i}}{\tilde{k}_i} \tilde{P}_i(\hat{x}_t). \tag{8}
\]

The feasible model-based forecast given by (8) and the corresponding forecast error are objects of interest for the rest of this subsection.

In the ideal case, the exact model, parameters and latent states are all known. In this case, the only error of the model-based forecast is the “genuine” forecast error:

\[
\text{GFE}^{model} = E[IV_{t+1} - IV_{t+1|t}(\Theta, x_t)]^2, \tag{9}
\]

where \( \Theta \) are the parameters in the model. However, the total forecast error of the feasible model-based forecast involves two extra components. First, the component to be termed the error in forecast \( IV_{t+1|t} \) due to errors in the parameters and states:

\[
F(\hat{x}_t - x_t, \hat{\Theta} - \Theta) = E[IV_{t+1|t}^t(\Theta, x_t) - IV_{t+1|t}(\hat{\Theta}, \hat{x}_t)]^2, \tag{10}
\]

where \( \Theta \) are estimated parameters in the model, and \( \hat{x}_t \) are estimated states in the model. This component of the error is uncorrelated with the “genuine” error, since the “genuine” error is unpredictable at time \( t \), i.e., \( E[IV_{t+1} - IV_{t+1|t}| F_t] = 0 \), and the error in parameters and states
is a function of the information available at time $t$, i.e., $IV_{t+1|t}(\Theta,x_t) - IV_{t+1|t}(\hat{\Theta},\hat{x}_t) \in F_t$. Hence, the total error of predicting $IV_{t+1}^t$ – the total out-of-sample mean-squared prediction error (MSPE) – is simply a sum of the two components,

**Decomposition of the Model-based Forecast:**

$$
\text{Total MSPE} = E[IV_{t+1}^t - IV_{t+1|t}(\hat{\Theta},\hat{x}_t)]^2 = GFE_{\text{model}} + F(\hat{x}_t - x_t, \hat{\Theta} - \Theta).
$$

Second, an additional part of the error comes from model misspecification. Model misspecification is practically unavoidable, since the true model is generally not known for any observed data set. Hence, any chosen model is at best an approximation of the true one.

This second additional component of the error is important, since each step to form a feasible forecast incorporates the knowledge of the model. First, we use the model to define the functional form of the model-based forecast. Then we use the model to estimate the parameters and extract the spot values of states, e.g., by MLE and particle filters, EMM and the reprojection method, or by Bayesian methods. (See Jacquier, Polson and Rossi, 1994, Johannes and Polson, 2003, Stroud, Müller and Polson, 2003, Gallant and Tauchen, 1998, and Pitt and Shephard, 1999.) Hence, the model misspecification is a critical factor that affects all the steps above and can weaken the performance of the model-based forecast.

To summarize, although the model-based forecast minimizes the mean-squared error in (5), it does so only under the assumption that the true model with the parameters is known, and all the states are observable. However, it may not do so under a more realistic assumption that many inputs are to be estimated. In this latter case, the total error consists of several parts and the “feasible” model-based forecast as defined by (8) may not perform the best under the mean-squared error loss.

We seek circumstances in which the above factors eliminate the advantages of model-based forecasting. As a natural competing forecast, we use a model-free reduced-form forecast, as originally advocated by Andersen, Bollerslev, Diebold and Labys (2003) and formally analyzed by Meddahi (2003).

### 2.3 The Reduced-Form Forecast

In this section, we define a benchmark reduced-form forecast. This forecast can be derived starting from formula (7), which reports the conditional expectation for the integrated variance $IV_{t+1}^t$ based on the ESV representation. The right-hand side of this formula is a sum of $p$ autoregressive processes. This property has an important implication; it implies that for a $p$-factor ESV model, the integrated variance $IV_{t+1}^t$ is a sum of $p$ AR(1) processes and a white noise term:

$$
IV_{t+1,t} = a_0 + \sum_{i=1}^{p} a_i \frac{1 - e^{-k_i}}{k_i} P_i(x_t) + \varepsilon_{t+1},
$$

where $E(\varepsilon_{t+1}|F_t) = 0$. This decomposition suggests that the integrated variance is an ARMA(p,p) process. In general, ARMA(p,p) for the integrated variance will take the form:

$$
\prod_{i=1}^{p} (1 - e^{-k_i} L)(IV_{t+1}^t - \theta) = \eta_{t+1} - \sum_{i=1}^{p} \beta_i \eta_{t+1-i},
$$

where $\eta_t$ is heteroscedastic white noise, and $k_1, ..., k_p$ are mean-reversions in the ESV representation (4). We can find the parameters of the ARMA model (12) if we know the parameters of the base...
model (1). Alternatively, we may also estimate the same parameters in a “reduced-form” manner. Since $IV_{t+1}^{t+1}$ can be described as an ARMA process, we may simply fit the linear time-series model to $IV_{t+1}^{t+1}$. One could also use a simple autoregressive model instead of ARMA. The resulting forecast yields a model-free $IV_{t+1}^{t+1}$ predictor based on the past realizations of the integrated variance.

**Definition:** The reduced-form forecast of the integrated variance $IV_{t+1}^{t+1}$ is a linear projection of $IV_{t+1}^{t+1}$ onto the space generated by its past realizations $IV_{\tau}^{\tau+1}$, $\tau \leq t$:

$$IV_{t+1}^{rf} | t = P(IV_{t+1}^{t+1} | IV_{\tau}^{\tau+1}, \tau \leq t).$$

The above forecast is not feasible because the integrated variance is not observed. However, we may construct a feasible version of the same forecast, using a close proxy of the integrated variance – realized variance:

**Definition** Suppose log-prices $s_t$ are observed at discrete times $0$, $h$, $2h$, etc. Then the realized variance over the period $[t, t+1]$ is defined as

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

The realized variance is a directly observable measure of the intra-day variance of the price. The asymptotic behavior of RV and its consistency as a proxy for IV is discussed by Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), and Meddahi (2002).

To construct a feasible version of the reduced-form forecast, we project $RV_t^{t+1}$ on its past values $\hat{P}(RV_t^{t+1} | RV_{\tau}^{\tau+1}, \tau \leq t)$. This projection can be constructed in a strictly model-free manner using only the data that are observable. For zero drifts in returns, this forecast will be equivalent to the projection of $IV_t^{t+1}$ on the past values of RV, i.e., $\hat{IV}_{t+1}^{rf} | t = \hat{P}(IV_t^{t+1} | RV_{\tau}^{\tau+1}, \tau \leq t)$. This equivalence follows from the fact that with no drift in returns, the difference $RV_t^{t+1} - IV_t^{t+1}$ is unpredictable. (See Barndorff-Nielsen and Shephard, 2002.) The same equivalence holds approximately for intra-day data, since the drift in asset prices is negligibly small between finely sampled observations.

The ARMA representation for RV follows from the ARMA representation of integrated variance (see Meddahi, 2003):

$$\prod_{i=1}^{p} (1 - e^{-k_i L})(RV_t^{t+1} - \theta) = \eta_{t+1}(h) - \sum_{i=1}^{p} \beta_i(h) \eta_{t+1-i}(h),$$

where $\eta(h)$ is heteroscedastic white noise that depends on the distance between observations $h$. If the parameters of the ARMA representation (14) are known, then the “genuine” error of the reduced-form forecast is the difference between the variance of the shock $\eta_{t+1}$ in (14) and the variance of the noise term $RV_t^{t+1} - IV_t^{t+1}$:

$$GFE^{rf} = E[IV_t^{t+1} - P(IV_t^{t+1} | RV_{\tau}^{\tau+1}, \tau \leq t)]^2 =$$

$$= E[RV_t^{t+1} - P(RV_t^{t+1} | RV_{\tau}^{\tau+1}, \tau \leq t)]^2 - E[RV_t^{t+1} - IV_t^{t+1}]^2.$$

(15)
The decomposition above follows from the fact that the difference between the realized and integrated variance is unpredictable.

If the parameters of the ARMA representation are unknown, then the total Mean Squared Prediction Error of the reduced-form forecast includes an additional part,

**Decomposition of the Reduced-form Forecast:**

\[ \text{Total MSPE}^{rf} = \text{GFE}^{rf} + F^{rf} (\hat{\Theta}^{rf} - \Theta), \]  

where the second part \( F^{rf} (\hat{\Theta}^{rf} - \Theta) \) comes from the errors in the coefficients and \( \Theta^{rf} \) are the parameters in the regression of \( RV_{t+1}^{f} \) on its past values. Note that the covariance term is absent from the above decomposition. Since the infeasible error is orthogonal to the linear space spanned by \( RV_{t+1}^{f}, \tau \leq t \), therefore orthogonal to the error from parameter estimation.

Observe that the total Mean Squared Prediction Error of the feasible reduced-form forecast does not include errors from estimating the instantaneous unobservable states \( x_t \).

### 3 Analytical Comparison of Feasible Forecasts

In this section we present a simple framework that allows us to compare feasible model-based and reduced-form forecasts analytically. Special attention will be paid to the effects of errors in the state estimates \( \hat{x}_t \) and model misspecification.

Throughout this section, we will assume that the data are generated from a multifactor model. Formally, this implies that for the ESV representation (3, 4) \( a_2 \neq 0 \). This assumption was tested for financial time-series and the hypothesis of a one-factor model (\( H_0 : a_2 = 0 \)) was consistently rejected. For example, Chernov et al. (2003) reject the one-factor hypothesis for stock-index data using \( \chi^2 \)-statistics within the EMM estimation. Bollerslev and Zhou (2002) also reject the one-factor hypothesis within the GMM estimation that matches conditional first and second moments of the realized variance for foreign exchange rates.

The success of multifactor models is explained by their dual implications; they can simultaneously generate a large variability in the variance and a high persistence over long horizons. This combination of properties is an attribute of asset price series. Since any multifactor process is a mixture of at least two components, one of these components adds to the variability of the variance, and the other accounts for the high persistence. Using the ESV notation (3), this implies that for a two-factor model with \( a_1 \neq 0, a_2 \neq 0 \) and \( a_i = 0, i > 2 \), the mean-reversion of one factor \( k_1 \) is significantly higher than \( k_2 \approx 0 \).

Nonetheless, in the literature, models with only one component are often used instead of multi-component models. For instance, the models that are listed in Table 1 are often chosen due to the simplicity of their handling and estimation.

Therefore, although the true generating process may be governed by several components, econometricians often assume simple one-component processes. In general, there are few research papers that focus on models with two components and even fewer that consider models with three components. (See Barndorff-Nielsen and Shephard, 2002.) We may expect that the true number of components is not limited to two or even three for real data, implying that the number of components is underestimated in many applications. Hence, the effect of the underestimation of the number of factors (components) can be a common source of forecasting error.

In this section, we consider a simple case in which an econometrician assumes a one-factor model of the general form:

\[ d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \Lambda(\sigma_t)dW_t. \]  

(17)
In the above model, the only factor is the spot variance. Therefore, we can make a link between the model (17) and the ESV representation (3) by defining the ESV parameters $a_0 = \theta$, $\sigma_t^2 = \text{Var}\sigma_t^2$, $P_t(x_t) = \frac{2^{\gamma t-a_0}}{|a_1|}$, and $k_1 = k$.

Hence, the model-based forecast can be derived as a special case of the general formula (7):

$$E_tIV_{t+1} = \theta + \frac{1 - e^{-k}}{k}(\sigma_t^2 - \theta).$$

(18)

The model-based forecast is linear in the last realized spot variance and the slope is a function of the mean-reversion coefficient $k$. To implement this forecast in practice the following problems are to be solved: parameter estimation and estimation of the latent spot variances $\sigma_t^2$.

The first problem of implementing (18) – estimation of the parameters – will not be fully addressed in this paper. We assume that a large span of data is available, thus driving errors in the parameters to zero. In the following analysis, we are concerned only with the errors that arise from the finite number of observations per unit of time (infill asymptotic).

Nevertheless, not knowing the parameters still poses a problem for the model-based approach, since if a true model were known, then regardless of the estimation procedure, the estimates would converge to their true unique values. However, in our case the model is misspecified. Consequently the limits of the estimated parameters may depend heavily on the estimation procedure. Therefore, the estimation procedure for the model parameters also plays an important role in the feasible model-based forecast.

The second problem of implementing (18) is filtering the state $\sigma_t^2$. As a result of this next step, another error will be added to the total prediction error. For illustration, assume that the parameters of (18) are estimated using the moment conditions $\hat{\theta} = E\sigma_t^2$ and $\frac{1 - e^{-k}}{k} = \frac{\text{cov}(IV_{t+1}^t, \sigma_t^2)}{\text{Var}\sigma_t^2}$

(19)

When we substitute the estimate $\hat{\sigma}_t^2$ into the formula (18), we get the following decomposition of the error:

$$IV_{t+1}^{t+1} - \hat{\theta} - \frac{1 - e^{-k}}{k}(\hat{\sigma}_t^2 - \hat{\theta}) = [IV_{t+1}^t - P_tIV_{t+1}^t] + \left[\frac{\text{cov}(IV_{t+1}^t, \sigma_t^2)}{\text{Var}\sigma_t^2}(\sigma_t^2 - \hat{\sigma}_t^2)\right],$$

where $P_tIV_{t+1}^t$ is a linear projection of $IV_{t+1}^t$ on $\sigma_t^2$. The first part of the above expression is the infeasible forecast error of the “last-spot” forecast, i.e., the linear forecast based on $\sigma_t^2$. The second part is due to the error in the spot variance.

For consistent estimates of the spot variance $\hat{\sigma}_t^2$, the error from the second part in (19) converges to zero as the sampling frequency increases to infinity, e.g., Aït-Sahalia et al. (2005). However, at very high frequencies, microstructure effects may blur the results. To avoid such microstructure effects the minimum distance between observations is often less than 5 minutes for financial time-series studies. Therefore, despite the asymptotic negligibility of the error in spot variance, it remains a nontrivial part of the total error in (19) under usual conditions. The error in the spot variance will also depend on the particular filtering technique used to extract the spot variance.

There are several ways to extract the spot variance $\sigma_t^2$ from past prices $s_{t-i}$, $i \geq 1$. In this paper, we will focus on the efficient ARCH filters of Nelson and Foster (1994). ARCH filters give consistent variance estimates under very general conditions, most importantly even for misspecified models. (See Nelson, 1992.) For example, for GARCH-SV models with $\Lambda(\sigma_t^2) = \eta(\sigma_t^2)$ in (17) and zero leverage, the efficient ARCH filter takes the form of a discrete-time GARCH(1,1):

$$\sigma_{t+h}^2 = \phi_h + a_h\sigma_t^2 + b_h\xi_{t+h}. $$

(20)

$^3$The limits of the estimated parameters $\theta$ and $k$ in the model (17) will be denoted by $\hat{\theta}$ and $\hat{k}$. For more details on parameter estimation see the proof of Proposition 1 in the Appendix.

$^4$We refer to this as a “no-bias” case, since the forecast is unbiased conditionally on $\sigma_t^2$. 

9
where \( \xi_{t+h} = \frac{a_{t+h}-a_t}{\sqrt{h}} \) is a normalized innovation in prices. The parameters of the filter \( \phi_h, a_h, \) and \( b_h \) are chosen optimally based on an estimated model. In particular, they will depend on the estimates of the persistence \( k \) and the volatility-of-volatility \( \lambda = \frac{\operatorname{Var}\sigma_t^2}{E\sigma_t^2} \).

In the Appendix, we prove the following statement:

**Proposition 1.** Let \( \sigma_t^2 \) be square integrable with the correlation function: \( \operatorname{corr}(\sigma_{t+h}^2, \sigma_t^2) = \sum_{i=1}^{p} a_i^2 e^{-k_i h} \sum_{i=1}^{p} a_i^2 \). Suppose we apply the ARCH filter of the form (20) to extract the spot variance. The log price process is described by (1) with no leverage effect and zero drift. Then the comparison of the reduced-form forecast and the forecast based on a one-factor model depends only on the following set of parameters \( \Xi \):

- **Volatility-of-volatility** – \( \lambda = \frac{\operatorname{Var}\sigma_t^2}{E\sigma_t^2} \);
- **Relative weights of factors**: \( \frac{a_i}{a_1}, i = 2, ..., p; \)
- **Persistence of factors**: \( k_i, i = 1, p; \)
- **Sampling frequency** – \( h \).

All the steps of the analytical comparison are given and proved in the Appendix. Notably, the assumption about the correlation structure of the variance process is very general, since ESV representation for any square-integrable process satisfies this assumption. Hence, the comparison of the forecasts is simplified within the ESV framework.

To understand the effects of the parameters \( \lambda, a_2/a_1, ..., a_p/a_1, k_1, ..., k_p, \) and \( h \) on the performance of the forecasting techniques, we consider two cases – the comparison of the forecasts when the true data-generating process is a one-factor model, i.e., model is correct, and the comparison of the forecasts under model misspecification. The importance of relevant parameters will be addressed in these subsections. To appreciate the results this section also includes an illustration using the GARCH-SV example from Andersen, Bollerslev and Meddahi (2004). In particular, we will replicate the comparison of the infeasible forecasts and update it by comparison of the feasible forecasts. The last two subsections extend results in two essential directions – long-horizon forecasting and forecasting with the data contaminated by the microstructure noise.

### 3.1 Comparison if the Model is Correctly Specified

If the model were correct, that is the forecast (18) were optimal, than out of two components of the error discussed above – model misspecification error and error in the latent state – only the latter would be left. Formally, the total MSPE of the model-based forecast would contain only two parts:

\[
\text{Total MSPE} = \text{GFE}_{\text{model}} + \left( \frac{1-e^{-k}}{k} \right)^2 E(\hat{\sigma}_t^2 - \sigma_t^2)^2.
\]

Note that the second part is inherently related to the quality of variance filtering. Thus only those parameters that define the accuracy of the estimate \( \hat{\sigma}_t^2 \) will also affect the decrease in the accuracy of the model-based forecast. As was mentioned above, the second part of the error is zero if the data are observed continuously. However, since the usual data are discrete, the comparison of infeasible forecasts depends on how model-based and reduced-form approaches are affected by the sampling frequency. The asymptotic behavior of an MLE-estimated spot variance is discussed by Nelson and Foster (1994), and Gloter and Jacod (2001a, b). In this paper we derive the exact formula for the error in the estimate \( \hat{\sigma}_t^2 \) extracted by the ARCH filter (20). The formula is given by (53) in the
Appendix and can be further simplified through approximation around \( h \approx 0 \), i.e., for infinitely active sampling.

From (53) in the case of a correct model, it follows that the error in \( \hat{\sigma}_t^2 \) equals approximately:

\[
E(\hat{\sigma}_t^2 - \sigma_t^2 | k)^2 \approx h \text{Var}^2 \frac{k}{b_h} + b_h E \sigma^4.
\] (21)

The formula above clarifies the trade-off in estimating \( \hat{\sigma}_t^2 \). The choice is between using as much data as possible to make the estimator more efficient, or using an estimation window as narrow as possible to reduce the bias. Both characteristics are functions of the same parameter \( b_h \) and the error is minimized by the following value \(^5\) of the GARCH-parameter \( b_h = \sqrt{\lambda k} \). The resulting error in the variance estimate equals

\[
E(\hat{\sigma}_t^2 - \sigma_t^2)^2 \approx 2 \text{Var}\sigma_t^2 \sqrt{\frac{kh}{\lambda}}.
\] (22)

Therefore, approximation of the total MSPE around \( h \approx 0 \) will contain the “genuine” forecast error plus the part coming from the term in (22):

\[
\frac{\text{Total MSPE}}{\text{Var}\sigma_t^2} = 2 \left[ \frac{1}{k} - 1 - \frac{1 - e^{-k}}{k^2} \right] - \left[ 1 - \frac{e^{-k}}{k} \right]^2 + 2 \sqrt{\frac{hk}{\lambda}} \left[ 1 - \frac{e^{-k}}{k} \right]^2 + O(h).
\] (23)

The analogous Taylor decomposition of the reduced-form forecast error given by equation (58) in the Appendix yields:

\[
\frac{\text{Total MSPE}}{\text{Var}\sigma_t^2} = \frac{\text{Var}[IV_{t+1}^t - P(IV_{t+1}^t | IV_{t+1-	au}^t, \tau = 1, 2, \ldots)]}{\text{Var}\sigma_t^2} + O(h).
\] (24)

The \( O(h) \) term appears in the above equation, since to predict \( IV_{t+1}^t \), we use past realizations of \( RV \) instead of the latent \( IV \)-series.

It is worth noting that the two errors above have different convergence rates with respect to the distance between observations \( h \). The reduced-form forecast is of a stochastic order \( O(h) \), while the model-based forecast of a stochastic order \( O(\sqrt{h}) \). Hence, we can conclude that \(^6\)

**Corollary 1.** The decrease in sampling frequency has a higher negative effect on the performance of the model-based forecasts, than on the performance of the reduced-form forecasts.

Another interesting observation is that the increase in the volatility-of-volatility \( \lambda \) affects the forecast comparison in favor of the model-based forecast. This observation can be explained as follows. Volatility series are filtered from noisy squared returns. The level of noise in the squared return is proportional to \( E\sigma^4 = \text{Var}\sigma^2 + E^2\sigma^2 \). Hence, a reduction in the mean \( E\sigma \) would reduce the noise in the returns while keeping the same variability of the spot variance. Therefore, an increase in the volatility-of-volatility makes square returns better proxies for the spot variance.

\(^5\)The parameter of the efficient GARCH filter is derived by Nelson and Foster (1994) for a one-factor GARCH-SV process. As follows from the Appendix, for a general ESV model the optimal choice of \( b_h \) is \( \sqrt{\lambda k}h \), where \( \tilde{k} \) is an average mean-reversion across volatility factors.

\(^6\)The same result holds under model misspecification, since for the model-based forecast

\[
\frac{\text{Total MSPE}}{\text{Var}\sigma_t^2} = 2 \left[ \frac{1}{k} - 1 - \frac{1 - e^{-k}}{k^2} \right] - \left[ 1 - \frac{e^{-k}}{k} \right]^2 + \sqrt{\frac{h}{\lambda}} \left( \frac{k}{\sqrt{\lambda}} + \sqrt{\tilde{k}} \right) \left[ 1 - \frac{e^{-k}}{k} \right]^2 + O(h),
\] (25)

where \( \tilde{k} \) is an estimate of \( k \) and \( \tilde{k} \) is an average mean-reversion across factors.
Corollary 2. As the sampling frequency decreases, the comparative performance of the model-based forecasts deteriorates less for higher levels of volatility-of-volatility \( \lambda = \frac{Var_{t}^{2}}{E\sigma_{t}^{4}} \).

Going back to Proposition 1, we conclude that the volatility-of-volatility, \( \lambda \), and sampling frequency, \( h \), affect the comparative performance of the forecasts if the model is correctly specified.

3.2 The Effect of Model Misspecification

In the above discussion, model misspecification resulted from application of one-factor models to the modeling of multifactor processes. Consider the case of a two-factor process. Within the ESV framework the true variance process includes two components with different mean-reversions:

\[
\sigma_{t}^{2} = a_{0} + a_{1}P_{1}(x_{t}) + a_{2}P_{2}(x_{t}),
\]

where \( a_{1} > 0 \) and \( a_{2} > 0 \). However, the econometrician wrongly assumes that either \( a_{1} = 0 \) or \( a_{2} = 0 \). Therefore, the measure of the model misspecification is the ratio \( \ln(\frac{a_{1}}{a_{2}}) \). This ratio is equal to \( \pm \infty \) when the model is truly one-factor, and thus the contribution of one of the factors is zero. At the other extreme, this ratio is zero when both factors are of equal importance, i.e., \( a_{1} = a_{2} \).

In this subsection, we investigate the effect of the model misspecification as measured by \( \ln(\frac{a_{1}}{a_{2}}) \) on the comparison between the model-based and reduced-form forecasts. The comparison is carried out for different values of the other parameters that influence the outcome: mean-reversions \( k_{1} \) and \( k_{2} \), volatility-of-volatility \( \lambda = \frac{Var_{t}^{2}}{E\sigma_{t}^{4}} \) and the sampling frequency.

First, we fix the estimates for \( k_{1} \) and \( k_{2} \) from studies that fitted two-factor models to financial series. These studies are summarized in Table 2 and include four examples: three papers estimated coefficients of multifactor models for foreign exchange rates and one dealt with stock indices. We derived the parameters of interest – mean-reversions \( k_{1} \) and \( k_{2} \), volatility-of-volatility \( \lambda \) – from the parameters reported in those papers.

For each pair of \( k_{1} \) and \( k_{2} \) from Table 2, Figure 1 shows the regions in the space \( (\ln(\frac{a_{1}}{a_{2}}), \lambda) \) where the reduced-form forecast outperforms the model-based forecast, i.e., total MSPE is lower for the reduced-form forecast. The regions corresponding to different sampling frequencies are shown in different intensities of grey. Small square marks inside the graph indicate representative parameter values taken from the studies in Table 2. For example, for the model of Alizadeh, Brandt and Diebold (2002) the coefficients are \( k_{1} \approx 0.9 \), \( k_{2} \approx 0.02 \), \( \lambda \approx 0.6 \), and \( \ln(\frac{a_{1}}{a_{2}}) \approx -1 \). From the bottom left panel in Figure 1, we see that the intersection of 0.6 for ordinates and –1.0 for abscissas is dark-grey, which corresponds to 15-minute sampling frequencies. This implies that the model-based forecast renders smaller errors for 5-minute sampling, but cedes efficiency to the reduced-form forecast for 15-minute and 30-minute frequencies.

Figure 1 illustrates the statements from the previous section. First, the graph shows that the decrease in sampling frequency adversely affects the model-based forecast, making it less appealing than the simpler alternative. As the distance between observations increases from 5 to 30 minutes, a larger area of parameter sets yields MSPE(reduced-form) < MSPE(model-based).

Second, the graph confirms that the effect of finite sampling on the performance of the model-based forecast is lower for high volatility-of-volatility \( \lambda \). For instance, consider the bottom left panel with \( k_{1} = 0.9 \) and \( k_{2} = 0.02 \) and choose the level of misspecification to be \( \ln(\frac{a_{1}}{a_{2}}) = -1 \). For \( \lambda = 0.1 \), the model-based forecast results in higher errors for all the chosen frequencies. For \( \lambda = 0.4 \), the model-based forecast is the most efficient for 5-minute but not for 15-minute or for 30-minute sampling. Finally, for \( \lambda = 0.8 \), the model-based forecast is more accurate for the highest frequencies (5-minute and 15-minute) and becomes less accurate only for 30-minute sampling.
Third, a higher persistence favors the reduced-form forecast. For each of the four panels, $k_1$ is higher than $k_2$. The “average” mean-reversion of the variance is equal to $a_1^2(a_1^2 + a_2^2)^{-1}k_1 + a_2^2(a_1^2 + a_2^2)^{-1}k_2$. Thus, the mean-reversion of the variance is increasing in $\ln\left(\frac{a_1}{a_2}\right)$. Figure 1 shows that the areas with a better performance of the reduced-form forecast are located to the left relative to the symmetric case $\ln\left(\frac{a_1}{a_2}\right) = 0$. Thus, other factors being equal, a higher persistence of the volatility process gives an advantage to the reduced-form forecast.

Finally, Figure 1 demonstrates how the model misspecification affects the forecast comparison. Model-misspecification is measured by the ratio $|\ln\left(\frac{a_1}{a_2}\right)| \in [0, +\infty)$, with no-misspecification cases located at $\ln\left(\frac{a_1}{a_2}\right) = \pm\infty$. Notably, the areas where the reduced-form forecast is better (shown in grey) are centered at $|\ln\left(\frac{a_1}{a_2}\right)| \approx 0$ for three of the four graphs, implying that the model misspecification acts in favor of the reduced-form forecast.

It can be formally shown that the “genuine” error in the model-based forecast (44) is quadratic in the ratio $\frac{a_2^2}{a_1^2}$ and achieves its global maximum at a finite value of $|\ln\left(\frac{a_1}{a_2}\right)|$. That is, the predictive power of the model-based forecast is at its minimum when the spot variance consists of two components: slow-moving and fast-moving, and the contributions of each of these components are non-trivial, i.e., in the case of misspecification. Thus, in reference to the results of the Proposition 1, we conclude that the ratios $a_2/a_1$, ..., $a_p/a_1$ define the comparative performance of the forecasts under model misspecification, with the performance of the model-based forecast adversely affected if values of these ratios are close to one.

These results indicate that the reduced-form forecast performs similarly to the model-based forecast due to two effects. First, the efficiency of the “best forecast” (model-based) vanishes for finite sampling frequencies. The second effect results from model misspecification and explains why the reduced-form forecast eventually outperforms the model-based forecast for certain parameter values. (See Figure 1.)

Note an important connection between the findings of this section and earlier findings by Nelson (1992), and Nelson and Foster (1995). The common theme is the performance of model-based techniques under model misspecification. A related result from these studies is the consistency of the estimate $\hat{\sigma}_t^2$ under a variety of misspecification types. Nelson and Foster (1995) also argue that under certain conditions ARCH filters yield consistent forecasts. However, the assumptions of Nelson and Foster (1995) are violated if the number of factors is underestimated. For this reason, contrary to Nelson and Foster (1995), we find a significant loss in predictive power for misspecified model-based forecasts.

3.3 Numerical Example

In this section we continue a numerical example from Andersen, Bollerslev, and Meddahi (2004). The example compares the model-based and reduced-form forecasts for the one-factor GARCH-SV process calibrated to daily series of the spot DM/USD exchange rate from 1987 to 1992. This exercise assumes that states are observable and the GARCH-SV is the true model for the data. We extend this example, first by relaxing the assumption that the states are observable, and second by relaxing the assumption that the true model is one-factor. Instead, we assume that the true model is two-factor, with parameters calibrated to the same DM/USD series.

Therefore, we use the same assumption made in the previous subsection: an econometrician employs a one-factor model of the type (17). Specifically, following Andersen, Bollerslev, and

---

7For comparison, the “genuine” error is linear in $\frac{a_1^2}{a_2^2}$ for a correctly specified two-factor model (see (45)), and achieves maximum for $\ln\frac{a_1^2}{a_2^2} = -\infty$ if $k_1 > k_2$. 

Meddahi (2004), he estimates the parameters to be \( k = 0.035, \theta = 0.636 \) and \( \Lambda(\sigma_t) = \sigma \sigma_t^2 \), where \( \sigma \) is chosen so that the volatility-of-volatility is \( \lambda = 0.296 \). However, the true data-generating process is two-factor and described by either of the following two candidates.

The first candidate is the SR-SV model estimated by Bollerslev and Zhou (2002) for high-frequency 5-minute spot returns on DM/USD from 1986 until 1996. The second candidate is the CEV-SV model estimated by Barndorff-Nielsen and Shephard (2002) for the same data. Thus, we have three different scenarios. Under the first scenario, the returns follow the one-factor GARCH-SV model. This case corresponds to the first row in Table 3. The second row in Table 3 is for the version where the true model is two-factor, as suggested by Bollerslev and Zhou (2002). Finally, the third row of the table corresponds to the case when the true model is from Barndorff-Nielsen and Shephard (2002). Thus, for the first row the model is specified correctly, but in the second and third rows the model is misspecified.

The data reported in Table 3 are the performances of the forecast based on the one-factor GARCH-SV model (the first two columns), and the reduced-form forecast (the last two columns). Table 3 uncovers the error-in-latent-states effect. In particular, the first and the third columns correspond to the case when the spot variance is observable, while the second and the fourth columns are for the case when the spot variance is latent. The results in the table assume a 5-minute distance between observations for “feasible” forecasts and are calculated using formulas derived in the Appendix (59 - 61).

Table 3 demonstrates the following results. First, the model-based forecast is the most efficient if the model is correct and the variance is observed. This is in accordance with the definition of the model-based forecast. In this case, the MSPE is a mere 2.3%, while the reduced-form forecast yields an error of 4.3%. Second, when the model is correct but the variance is unobserved, the quality of the model-based forecast deteriorates but this approach remains the most efficient with an MSPE of 6.2%.

Third, when the model is wrong, the performance of the model-based forecast is affected much more strongly than in the case of unobserved variance. For the 2F-SR-SV(II) model, for example, instead of an MSPE of 2.3% (if the model were correct), the model-based forecast delivers an MSPE of 67.2%. However, the model-based forecast keeps the leading position, as the reduced-form forecast gives 68.6%. Finally, the combination of two effects – misspecification in the model and unobserved variance – drives the performance of the model-based forecast below the reduced-form forecast. Despite its simplicity, the reduced-form forecast gives 68.8% MSPE versus 115.6% MSPE of the model-based forecast.

In the case of the 2F-SR-SV(I) model, the results are qualitatively similar, and the reduced-form forecast renders smaller errors, although the difference in errors is less dramatic: 35.4% error in the reduced-form forecast versus 39.3% error in the model-based forecast.

The most remarkable conclusion from Table 3 is that irrespective of the case we considered, Mincer-Zarnowitz \( R^2 \) of the feasible model-based forecast and the feasible reduced-form forecast are generally close. This finding indicates that the day-ahead reduced-form forecast is successful in capturing the same information that is available to the model-based forecast. Moreover, in terms of the MSPE, the reduced-form forecast can be significantly more accurate, since it is unbiased by

---

\[ \text{To construct } R^2, \text{ data on realizations of } IV_{t+1}^j \text{ are regressed on the forecasts } IV_{t+1|t}. \text{ The } R^2-\text{statistic of this regression is a measure of the forecast efficiency. There is a link between } R^2 \text{ and MSPE, which is a main performance measure in this paper:} \]

\[ \frac{\text{MSPE}}{\text{Var} IV_{t+1}^j} = 1 - R^2 + \frac{\left[ P(IV_{t+1}^j|IV_{t+1|t}) - IV_{t+1|t} \right]^2}{\text{Var} IV_{t+1}^j}, \]

\[ \text{where } P(IV_{t+1}^j|IV_{t+1|t}) \text{ is the projection of } IV_{t+1}^j \text{ on its forecast } IV_{t+1|t}. \text{ It follows that } R^2 \text{ is a valid measure of the forecast performance that does not, however, takes into account the bias part: } P(IV_{t+1}^j|IV_{t+1|t}) - IV_{t+1|t}. \]
construction. To summarize, both the possible model misspecification and the finite frequency of price observations should be taken into account when comparing the model-based forecast and reduced-form forecast. The combination of these factors produces a different ranking of the forecasts. This example illustrates how the comparative ranking of the model-based and the reduced-form forecasts can reverse given realistic assumptions.

3.4 Long-Horizon Forecasting

It can be of separate interest to investigate how model-based and model-free approaches perform over longer horizons. Multi-factor models may exhibit long-memory-like properties, but one-factor models cannot. Therefore, a different behavior can be expected from one-factor and multifactor models for longer-horizon predictions.

In this section, we consider the forecasting of the integrated variance \( IV_{t+T} \) if \( T > 0 \). A long-horizon model-based forecast for the ESV models is a generalization of the formula (7):

\[
E_t IV_{t+T} = a_0 + \sum_{i=1}^{p} a_i \frac{1 - e^{-k_i}}{k_i} e^{-k_i T} P_i(x_t). \tag{26}
\]

For example, for a one-factor model, the long-horizon forecast takes the form:

\[
E_t IV_{t+T} = \theta + \frac{1 - e^{-k}}{k} e^{-k T} (\sigma^2_t - \theta). \tag{27}
\]

The formulas above imply that as the horizon \( T \) increases, the only error, whose effect may be amplified, is the error from misspecification. Indeed, the error from the estimates of latent states remains constant, since the input \( \hat{x}_t \) is the same for all \( T \). Moreover, as \( T \to \infty \), its contribution to the MSPE fades away. On the other hand, as \( T \) increases the true forecast converges to

\[
\lim_{T \to \infty} \frac{E_t IV_{t+T} - a_0}{e^{-k_j T}} = a_j \frac{1 - e^{-k_j}}{k_j} P_j(x_t), \tag{28}
\]

where \( k_j = \min_i k_i \). Therefore, the factors with the smallest mean-reversions, \( k_i \), will dominate the forecast. If the number of factors is underestimated, the components that play a dominant role for weekly and longer-horizon forecasts will not be extracted properly. For example, for one-factor models, these dominant components are hidden within the estimate of the spot variance \( \hat{\sigma}_t^2 \).

We extend the analysis of Section 3.2 to find the parameters for which the long-horizon reduced-form forecast outperforms the model-based forecast. Algebraic computations in the general case are given in the Appendix, and the results for \( T = 1 \) are presented in Figure 2. As in the \( T = 0 \) case, the parameter sets for which the reduced-form forecast performs best are colored in grey (light grey, dark grey and black), starting from black for 5-minute frequencies.

Quite remarkably, all of the actual model estimates from Table 2, indicated by squares, now fall into this dark-shaded area, suggesting that for long-term forecasts the reduced-form forecast is more efficient even for the highest frequencies. It follows that for all these models, the underestimation of the number of factors weakens the power of the model to predict for longer horizons, while the performance of the reduced-form forecast is affected far less.

3.5 Effect of Micro-Structure Noise

We previously assumed that the observed prices are not contaminated by microstructure effects, i.e., the model (1) described both the dynamics for observed and efficient, “true”, prices. For
available data, however, it is fairer to model the asset price as a sum of the efficient price and a component often referred to as the microstructure noise, \( n_t \), i.e.,

\[ p_t = p_t^* + n_t, \]

where now only \( p_t^* \) follows the dynamics \( (1) \), and \( n_t \) are zero-mean i.i.d shocks. The reasons for the presence of the microstructure effects are bid-ask spreads, discrete grid of quotes, heterogeneity of transactions etc.

To construct the total MSPE for the model-based forecast and reduced-form forecast, the effect of this small contamination has to be considered. It is known that the effect of the microstructure noise on the realized variance grows indefinitely as \( h \to 0 \), see Zhang, Mykland, and Aït-Sahalia (2005). Therefore, we would expect that the prior results will be largely distorted at high frequencies. The performance of various reduced-form forecasts under micro-structure noise is studied in detail by Andersen, Bollerslev, and Meddahi (2009). This section looks at a similar set up from a new perspective, comparing those to the model-based approach for a continuum of parameters.

To evaluate the effect of the microstructure noise on the forecasts, we make two simplifying assumptions. First, as in the main-stream literature (see Aït-Sahalia and Mancini, 2008), the microstructure noise is modeled by a normally distributed i.i.d sequence. In the proofs, the normality assumption defines only the link between the variances of \( n_t^2 \) and \( n_t \), and therefore, this assumption can be easily relaxed by calibrating these quantities separately. The second assumption is that the biases of the model-based and the reduced-form forecasts are equal. This assumption holds when \( \hat{\theta} \) in the model is estimated by a sample mean of spot variance estimates, \( \hat{\sigma}^2_t \). The derivations of the modified formulas for MSPEs are outlined in the Appendix.

In addition to the RV-based reduced-form forecast, this section also reports the results for the reduced-form forecast based on TSRV – a non-parametric measure of variance that is robust to the microstructure noise,

\[ TSRV_{t+1}^* = \frac{1}{K} \sum_{j=1}^{K} RV_{t+(j-1)h}^{t+1} - \frac{1}{K} RV_{t, all}^{all, t+1} \]  

(29)

where realized variances \( RV_{t+(j-1)h}^{t+1} \) are based on observations recorded at \( K \times h \)-intervals starting at \( t + (j-1)h \). The “tick-by-tick” realized variance, \( RV_{t, all}^{all, t+1} \) is computed using the data at higher-frequency observations recorded at each \( h \)-interval. Studies show the superior performance of TSRV over RV as a measure of the return variability, see Zhang, Mykland, and Aït-Sahalia (2005) and Aït-Sahalia and Mancini (2008). The results of these studies motivate us to include this as an alternative in the comparison of the forecasts. Similar to RV, the reduced-form forecast based on TSRV is a linear projection of the future TSRV on its past values.

Figure 3 shows the comparison between the model-based forecast and two reduced-form forecasts: left panels correspond to the RV case, and right panels correspond to the TSRV case. To highlight the effect of the microstructure noise, we assume that the available data are finely recorded at 5-minute intervals, and the forecasts use all the available data. For the TSRV forecast, we assume that \( K = 12 \). Thus, the graph fixes the frequencies but varies the intensity of the microstructure noise measured by the noise-to-signal ratio, \( \frac{En^2}{EIV_1} \). The selected noise intensities are from the same range as in Andersen, Bollerslev, and Meddahi (2009). Different shades of grey denote parameter sets for which reduced-form forecasts outperform the model-based forecast for different choices of the micro-structure noise intensity.

The first question addressed by Figure 3 is whether the model-based or the reduced-form forecasts is more sensitive to the microstructure noise. Note, that the parameter sets for which reduced-form forecasts outperform the model expand with the increase in the noise intensity. Even at fine
5-minute frequencies, the RV-reduced-forecast outperforms the model-based forecast for a large set of parameters if the noise-to-signal ratio is at least 1%. And the TSRV-reduced-form forecast is almost always the best choice.

The second question addressed by Figure 3 is whether the use of TSRV instead of RV yields a substantial improvement. Comparing the left and right panels of the graph, we can see that the parameter sets for which TSRV outperforms the model are significantly larger than these of RV. Thus, this study confirms the superior performance of TSRV-reduced-form forecasts in the presence of micro-structure noise. See Andersen, Bollerslev, and Meddahi (2009) for a thorough study of this topic.

Note, that as in the one-period case without the noise, the result in Figure 3 is driven mostly by the error in estimation of the spot variance, since \( E_t IV_{t+1} \approx \sigma_t^2 \). Thus, to gain quick insight into the result from Figure 3 we can analyze the decomposition of \( RV \), see Aıt-Sahalia and Mancini (2008), for \( h \approx 0 \),

\[
RV_{t+1}^t \approx IV_{t+1}^t + 2 \frac{E_{n_t}^2}{h^2} + Z \sqrt{4E_{n_t}^4 h^{-1} + O(h)},
\]

where \( Z \) is a standard normal variable. From formula (62) in the Appendix it follows that

\[
\text{Var}(\sigma_t^2 - \sigma_t^2) \approx \frac{2b_h^2}{h^2} E_{n_t}^4 \approx \frac{2\sqrt{\lambda k}}{h^{3/2}} E_{n_t}^4.
\]

Hence, we conclude that the variance of the model-based forecasting error is \( h^{-3/2} \) - sensitive to the microstructure effects, while the sensitivity of the reduced-form forecasting error is of the order \( h^{-1} \).

4 Empirical Example

We have demonstrated that the efficiency of the model-based forecast vanishes when we consider its feasible version. This result was established theoretically for the models that are calibrated to observed data. In this section we demonstrate the same effect with actual DM/USD exchange rate 5-minute returns from December 2, 1986, through June 30, 1999. The chosen market is open twenty-four hours a day, yielding a total of 288 observations per day. This data has been extensively studied before, and for a thorough description, one can refer to Andersen and Bollerslev (1998). In this study, we use raw data to form RV series. For filtration of states in the model-based forecasts, we adjust the intra-day returns by the seasonal component \(^9\).

We assess the performance of three model-based forecasts: one-, two- and three-factor. The goal of this section is to compare these to the reduced-form approach. The object of the forecasting in all these cases will be the realized variance, RV.

4.1 One-Factor Model-Based Forecasts

To estimate the parameters in (18) we rely on the method of moments that is independent of the other parameters in (1). Specifically, we match the mean of the T-period realized variance \( \theta = \frac{1}{T} E RV_{t+T}^t \), the correlation structure of the realized variance \( e^{-kT} = \text{corr} \left( RV_{t-T}^t, RV_{t+T}^t \right) \times \left( \text{corr} \left( RV_{t-T}^t, RV_{t+2T}^t \right) \right)^{-1} \), and the variance of the realized variance to get the volatility-of-volatility \( \lambda \). To avoid any biases in favor of the reduced-form forecast, we report parameters

\(^9\)We also verified that the results of this section do not change when using raw data.
and the forecast performance for different horizons $T$. Specifically, we will work with parameters calibrated to daily ($T = 1$), weekly ($T = 5$), monthly ($T = 20$), and quarterly data ($T = 60$).

The performance of this forecasting procedure is reported in the first half of Table 5, which contains Mincer-Zarnowitz $R^2$s and normalized MSPEs for different parameter sets in Table 4.

The last line of the table corresponds to the performance of the reduced-form forecast that is a simple AR-model for RV, with the number of lags chosen by the BIC-criterion. The reduced-form forecast is purely out-of-sample, as the number of lags and values of parameters are calculated using only the data available at the time of the forecast.

Forecast performances were evaluated using the whole data sample excluding the first year, i.e., for the period of 1987 - 2007. The Mincer-Zarnowitz $R^2$ of the model-based forecast is 35.6% for parameters calibrated to quarterly data. This result is close to the corresponding statistic for the reduced-form forecast (35.7%). Moreover, the difference is slightly in favor of the latter. For the other parameter values, the model-based forecast performs noticeably worse. In terms of the MSPE, the reduced-form forecast has a clear advantage with the error of 64.6% versus 76.5%.

4.2 Multi-Factor Model-Based Forecasts

The two-factor model considered here is the affine model suggested and estimated by Bollerslev and Zhou (2002) for the same data set on the DM/USD exchange rates. In contrast to the one-factor case, there are two latent states $(x_{1t}, x_{2t})$ in this model. To extract these components we rely on the efficient Particle filter following Durham (2004)\(^\text{10}\). We also include the results for a three-factor CEV-SV model with parameters estimated by Barndorff-Nielsen and Shephard (2002) for the same data set with 10-minute observations. Barndorff-Nielsen and Shephard (2002) do not choose a specific family of models. The CEV-SV model is one of possible choices.

The second half of Table 5 reports the forecasting performance of the multi-factor models. The table shows that in terms of $R^2$ the two-factor model-based forecast appears to be the best choice, explaining 37.0% of the variance of RV. However, the reduced-form forecast is very close, with an $R^2$ of 35.7%. Moreover, in terms of the mean-squared error, the reduced-form forecast is better than the model (64.6% vs. 69.8%). The performance of the three-factor model is an example of the situation when model-based forecasting could be extremely inaccurate. A possible explanation is that the CEV-SV model does not properly extract volatility components, hence, leading to the $R^2$ of 26.7%.

The conclusion from Table 5 is that, despite a better fitting of the data, the multi-factor models produce forecasts that do not significantly outperform the one-factor model in terms of $R^2$. This could be due to the fact that the large part of variation in the next-day realized variance is explained by the short-run components, and therefore one-factor models are sufficient to capture the dynamics of the variance over short horizons. As was indicated in the previous section and as also follows from Table 5, the reduced-form forecast succeeds in capturing the same information, and therefore renders the close $R^2$ result. Moreover, owing to its unbiasedness, it gives the lowest MSPE (64.6%).

4.3 Long-Horizon Forecasts

Table 6 extends empirical results from the previous subsection to long-horizon forecasts – to predict variance after one day, $T = 1$, after one week, $T = 5$, and after two weeks, $T = 10$. The one-factor forecast in the table is “quarterly”- calibrated, i.e., $k$ equals 0.018. This parametrization proved

\(^{10}\)In this study, the states $x_{1t}, x_{2t}$ are extracted using 5-minute, 15-minute and hourly returns. Here we report results only for 15-minute frequencies, since they led to the smallest forecast errors.
to be the most successful for day-ahead forecasting. However, for longer horizons, its performance gradually decays, with MSPE rising to 110.8%.

In contrast to the one-factor model, the drop in the quality of the reduced-form forecast and of the multi-factor models is less dramatic. The MSPE for the two-factor model reaches the level of 94.6% and $R^2$ remains at the level of 11.0%. The three-factor model yields a lower MSPE of 92.0%, but also lower $R^2$ of 10.2%. The reduced-form forecast yields MSPE of 88.7% and $R^2$ of 11.5% for the two-week forecast.

For the multi-factor models the long-horizon forecasts can be represented by the formula (26). Just as in the case of the day-ahead forecasting, their implementation requires time-consuming estimation of the parameters and states $x_t$. In return, they perform only on par with the reduced-form forecast, which is much simpler to implement - $R^2$ of 10 - 11% versus $R^2$ of 11.4%. Moreover, their MSPE is objectively higher than that of the reduced-form forecast. Thus, based on the results in this section, we find that even more sophisticated multi-factor models, which are better at explaining the data, yield higher forecasting errors than the reduced-form forecasts.

5 Conclusion

In this paper, we compare the performances of model-based and reduced-form forecasts of integrated variance assuming availability of intra-day data. We show that when it comes to the feasible versions of the forecasts, reduced-form forecasts can outperform model-based forecasts. Since the model-based forecast requires the knowledge of both the true model and the latent instantaneous volatility, model misspecification and the errors in instantaneous volatility estimates can in effect combine to make the model-based perform worse than the reduced-form forecast. We also confirmed these results with actual high-frequency foreign exchange rates and several popular stochastic volatility models.

This paper challenges the conventional wisdom that models always render the most efficient forecasts. It shows that simpler approaches do not necessarily fail in comparison with this benchmark. Our results lead us to challenge the belief that, though estimation and forecasting within SV-models are notoriously time-consuming, the resulting gain in efficiency justifies it. On the contrary, based on the analytical and empirical evidence in this study, we conclude that reduced-form forecasts perform similar to supposedly efficient forecasts and are often better. And though our analysis has been limited to particular assumptions, the latter are still quite general. We anticipate, therefore, that our results may carry over to more complex settings.

References


A Moments of Volatility Measures

• To keep formulas in this appendix parsimonious, we use the following notation. For multi-factor models of the form (3) we denote a weighted average of an arbitrary function \( f(k_i) \), \( i = 1, \ldots, p \) over all factors with a bar operator:

\[
\bar{f}(k) = \frac{\sum_{i=1}^{p} a_i^2 f(k_i)}{\sum_{i=1}^{p} a_i^2},
\]

where \( k_i, i = 1, \ldots, p \) are persistence parameters and \( a_i, i = 1, \ldots, p \) are factor weights. Notably any function of this type depends only on persistence parameters \( k_i \), and ratios \( \frac{a_i}{a_1}, i = 2, \ldots, p \).

• The following moments for the integrated variance were derived by Andersen, Bollerslev, and Meddahi (2004):

\[
\text{Cov}(IV_{t+1}^t, \sigma^2_t) = \frac{1 - e^{-k}}{k} \text{Var} \sigma^2_t, \tag{30}
\]

\[
\text{Var} IV_{t+1}^t = 2 \text{Var} \sigma^2_t \left[ \frac{1}{k} - \frac{1 - e^{-k}}{k^2} \right], \tag{31}
\]

\[
\text{Cov}(IV_{t+1}^t, IV_{t+l+1}^t | l > 0) = \text{Var} \sigma^2_t \left( \frac{1 - e^{-k}}{k} \right)^2 e^{-k(l-1)}. \tag{32}
\]

Hence, the correlation structure of the integrated variance depends only on the persistence parameters \( k_i \) and ratios \( \frac{a_i}{a_1}, i = 2, \ldots, p \).

• For the no-leverage case, Meddahi (2003) obtained the following moments of the realized variance:

\[
\text{Cov}(RV_{t+1}^t, RV_{t-j+1}^t) = \text{Cov}(IV_{t+1}^t, IV_{t-j+1}^t), \quad j \neq 0,
\]

\[
\text{Var} RV_{t+1}^t = \text{Var} IV_{t+1}^t + \text{Var} u_{t+1}, \tag{32}
\]

where the last component is the variance of the noise \( u_{t+1} \) in \( RV_{t+1}^t \). Its variance has been derived by Barndorff-Nielsen and Shephard (2002):

\[
\text{Var} u_{t+1} = 2h \left[ E^2 \sigma^2 + \frac{\text{Var}(\sigma^2)}{h^2} \int_0^h \int_0^h \phi(s-\tau)dsd\tau \right] = \frac{2h}{h} \text{Var}(\sigma_t^2) \left[ \frac{1 - \lambda}{\lambda} + \frac{2}{h^2} \left( \frac{h}{k} - 1 - e^{-kh} \right) \right]. \tag{33}
\]

• We also will need to find the second and fourth moments of the returns. Denote the \( h \)-period demeaned return by \( \xi_{t+h} = h^{-1/2} \int_t^{t+h} \sigma_\tau dW_\tau^p \). From Itô–isometry for square-integrable prices, it follows that

\[
E\xi_{t+h}^2 = \frac{1}{h} E \int_t^{t+h} \sigma_\tau^2 d\tau = E\sigma_t^2.
\]
Similar to the proof of Itô–isometry, we can show that for the square-integrable spot variance under the no-leverage condition, the forth moment of the return equals

\[
E\xi^4_{t+h} = \frac{3}{h^2} E \left[ \int_t^{t+h} \int_t^{t+h} \sigma^2_x \sigma^2_s d\tau ds \right] = 3E^2 \sigma_t^2 + 3 \text{Var}^2 \sigma_t^2 \int_t^{t+h} \phi(\tau, s)d\tau ds, \\
E[\xi^2_{t+h-jh} \xi^2_{t+h-ih}]_{i \neq j} = \frac{1}{h^2} E \left[ \int_{t-(j+1)h}^{t-jh} \int_{t-(i+1)h}^{t-ih} \sigma^2_x \sigma^2_s d\tau ds \right] = E^2 \sigma_t^2 + \text{Var}^2 \sigma_t^2 \int_{t-(j+1)h}^{t-jh} \phi(\tau, s)d\tau ds.
\]

A variant of the above is proved by Meddahi (2002) for ESV models. From the ESV representation, it follows that the correlation function of the square-integrable process can be decomposed in the form \(\text{corr}(\sigma^2_{t+h}, \sigma^2_t) = \sum_{a_i} a^2 e^{-k_i h} / \sum_{a_i} a_i^2\). Therefore, for ESV models \(\phi(t, s) = e^{-k|t-s|}\) and

\[
E\xi^4_{t+h} = 3E^2 \sigma_t^2 + 6 \frac{\text{Var}^2 \sigma_t^2}{h^2} \frac{1 - e^{-kh}}{k^2}, \\
E[\xi^2_{t+h-jh} \xi^2_{t+h-ih}]_{i \neq j} = E^2 \sigma_t^2 + \frac{\text{Var}^2 \sigma_t^2}{h^2} \left[ \frac{1 - e^{-kh}}{k} \right]^2 e^{-k(j-i-1)h}.
\] (34)

Finally, we will derive the covariance structure between factors \(P_t(x_t)\) and past \(h\)-period squared returns \(\xi^2_{t-jh}\). Similarly to Itô–isometry, since covariances are preserved under \(L^2\) convergence, we can show that

\[
\text{cov}(P_t(x_t), \xi^2_{t-jh}) = \frac{1}{h} \int_{t-(j+1)h}^{t-jh} \text{cov}[P_t(x_t), \sigma^2_x] d\tau = \frac{1}{h} \int_{t-(j+1)h}^{t-jh} \text{cov}[P_t(x_t), P_t(x_{\tau})] d\tau = \frac{a^2}{h} \int_{t-(j+1)h}^{t-jh} e^{-k|t-\tau|} d\tau = \frac{a^2}{h} e^{-k_jh} \frac{1 - e^{-k_i h}}{k_i}.
\] (35)

Summing up over the factors, we obtain the formula for the covariance of the spot variance and past squared returns:

\[
\text{cov}(\sigma^2_t, \xi^2_{t-jh}) = \text{Var}^2 \sigma_t^2 e^{-k_jh} \frac{1 - e^{-kh}}{kh}.
\] (36)

**B Proof of Proposition 1**

In this section we identify a complete set of parameters that affect the comparison of the model-based forecast and the reduced-form forecast. The model-based forecast employs an arbitrary stochastic volatility (SV) model with a variance dynamics of the form (17). The GMM approach is used to estimate the parameters \((k, \theta)\), and the states are extracted by the ARCH filter. The reduced-form forecast employs the ARMA(p,p) model for realized variances RV to predict IV.

For the GMM estimate, our first choice is a procedure that leaves forecast (18) unbiased irrespective of the model. This is achieved by a GMM procedure that matches the expectation

\[
E(IV_{t+1}^t - \tilde{\theta}) = 0,
\] (37)
and the regression slope:
\[ \frac{\text{cov}(IV_{t+1}, \sigma_t^2)}{\text{Var} \sigma_t^2} = 1 - e^{-\bar{k}}. \] (38)

The case when parameters are defined by the conditions above will be further referred to hereafter as the “no-bias” case: in Mincer-Zarnowitz regressions, this forecast would yield the zero intercept and the slope equal to one. However, instead of the above moments, other moment conditions can be used to define \( \tilde{k} \) and \( \tilde{\theta} \). In general, estimation procedures based on those other moments may introduce a forecast bias.

For the ARCH filter, \( \hat{\sigma}_t^2 \) follows the discrete time GARCH(1,1):
\[ \hat{\sigma}_{t+h}^2 = \phi \hat{\sigma}_t^2 + a_h \sigma_t^2 + b_h \xi_{t+h}, \] (39)
where \( \xi_{t+h} = \frac{p_{t+h} - p_{t-h}}{\sqrt{h}} \). The parameters of the model are chosen optimally to minimize asymptotic MSE.

**Proposition 1.** Let \( \sigma_t^2 \) be square integrable with the correlation function: \( \text{corr}(\sigma_{t+h}^2, \sigma_t^2) = \frac{\sum p_i a_i^2 e^{-k_i h}}{\sum a_i^2} \).

Suppose we apply the ARCH filter of the form (20) to extract the spot variance. The log price process is described by (1) with no leverage effect and zero drift. Then the comparison of the reduced-form forecast and the forecast based on a one-factor model depends only on the following set of parameters \( \Xi \):

- **Volatility-of-volatility** – \( \lambda = \frac{\text{Var} \sigma_t^2}{E \sigma_t^2} \);
- **Relative weights of factors**: \( \frac{a_i}{a_1}, i = 2, \ldots, p; \)
- **Persistence of factors**: \( k_i, i = 1, \ldots, p; \)
- **Sampling frequency** – \( h \).

**Proof.** The proof is organized as follows. First, we find the total MSPE of the “no-bias” model-based forecast. Second, we find the total MSPE of the reduced-form forecast. Finally, we compare the errors and consider a few extensions, namely (1) model-based forecast with a bias and (2) long-horizon forecasts.

### B.1 MSPE of the Model-Based Forecast

As was shown in Section 3, the total MSPE for the “no-bias” forecast based on a one-factor model is the sum of the “genuine” forecast error,

\[ \text{GFE}^{\text{model}} = E \left[ IV_{t+1} - E \sigma_t^2 - \frac{\text{cov}(IV_{t+1}, \sigma_t^2)}{\text{Var} \sigma_t^2} (\sigma_t^2 - E \sigma_t^2) \right]^2, \] (40)

and the part that is due to the error in spot variance,

\[ F(\hat{\sigma}_t^2 - \sigma_t^2) = \left[ \frac{\text{cov}(IV_{t+1}, \sigma_t^2)}{\text{Var} \sigma_t^2} \right]^2 E (\sigma_t^2 - \hat{\sigma}_t^2)^2. \] (41)

If the model is correctly specified, then the “genuine” error is independent of all the past information and, therefore, of the error in the spot variance. However, under model misspecification this is not the case and the additional part of the error is the covariance between the above two components:

\[ \text{Cov}(\text{GFE}^{\text{model}}, F(\hat{\sigma}_t^2 - \sigma_t^2)) = \text{Cov}(IV_{t+1} - \frac{\text{cov}(IV_{t+1}, \sigma_t^2)}{\text{Var} \sigma_t^2} \sigma_t^2, \frac{\text{cov}(IV_{t+1}, \sigma_t^2)}{\text{Var} \sigma_t^2} (\sigma_t^2 - \hat{\sigma}_t^2)). \] (42)
Forecast is formed using the estimates \( \tilde{k} \) and \( \tilde{\theta} \) that match the moments (37) and (38). Using formula (30), those moments imply that

\[
\frac{1 - e^{-\tilde{k}}}{\tilde{k}} = \frac{1 - e^{-k}}{k}.
\]

Further analysis of the model-based forecast error will be organized into three steps. First, we will devise the formula for the “genuine” error (40). Second, we will find \( F(\hat{\sigma}_t^2 - \sigma_t^2) \) from (41). Third, we will provide the formula for the covariance term (42).

**Step 1: “Genuine” forecast error \( \text{GFE}^\text{model} \)**

The expression for \( R^2 \) from the Mincer-Zarnowitz regression obtained by Andersen, Bollerslev, and Meddahi (2002) can be readily converted into the expression for the corresponding mean squared error:

\[
\text{GFE}^\text{model} = \text{Var} IV_{t+1}^p (1 - R^2) = \text{Var} IV_{t+1}^p - \frac{1}{\text{Var} \sigma_t^2} \left[ \sum_{i=1}^{p} a_i^2 \left( \frac{1 - e^{-k_i}}{k_i} \right) \right]^2.
\]

Substituting the moment (31) into the above expression yields the formula for the “genuine” error:

\[
\text{GFE}^\text{model} = \text{Var} \sigma_t^2 \left( 2 \left( \frac{1}{k} - \frac{1 - e^{-k}}{k^2} \right) - \left[ \frac{1 - e^{-k}}{k} \right]^2 \right).
\]

For comparison, we derive the “genuine” error of a forecast based on a multifactor model. It also follows from the expression for \( R^2 \) from the Mincer-Zarnowitz regression obtained by Andersen, Bollerslev, and Meddahi (2004) and equals

\[
\text{GFE}^\text{p-Fmodel} = \text{Var} \sigma_t^2 \left( 2 \left( \frac{1}{k} - \frac{1 - e^{-k}}{k^2} \right) - \left[ \frac{1 - e^{-k}}{k} \right]^2 \right).
\]

**Step 2: Error due to volatility estimation**

Rewrite the filter for the spot volatility in the following form:

\[
\hat{\sigma}_t^2 = b_h \sum_{j=0}^{\infty} a_h^j \xi_{t-j}^2 + \frac{\phi_h}{1 - a_h},
\]

where the innovation is equal to \( \xi_{t+h} = \frac{\int_{t+h} \sigma_c dW_p}{\sqrt{h}} \). That is, our estimation of the spot volatility is a weighted average of past squared demeaned returns and a constant that later will be defined to converge to zero as \( h \to 0 \).

After squaring the estimate \( \hat{\sigma}_t^2 \) and taking the expectation, we find that the second moment of the estimator \( \hat{\sigma}_t^2 \) equals

\[
E \hat{\sigma}_t^4 = \frac{\sigma_h^2}{(1 - a_h)^2} + 2 \frac{\phi_h}{1 - a_h} \sum_{j=0}^{\infty} a_h^j E(\xi_{t-j}^2) + b_h^2 E \left( \sum_{j=0}^{\infty} a_h^j \xi_{t-j}^2 \right)^2.
\]

\footnote{See Meddahi (2001) for the related discussion of the misspecification case.}
After substituting moments from (34) we find that

\[
E(\hat{\sigma}_t^2) = \frac{\phi_h^2}{(1 - a_h)^2} + \frac{2\phi_h b_h}{(1 - a_h)^2} E\sigma_t^2 + 3 \frac{b_h^2}{1 - a_h^2} \left[ E^2\sigma_t^2 + 2 Var\sigma_t^2 \frac{h}{k} \frac{1 - e^{-kh}}{h^2} \right] + 2 \frac{b_h^2}{1 - a_h^2} \left[ \frac{E^2\sigma_t^2 a_h}{1 - a_h} + Var\sigma_t^2 \frac{(1 - e^{-kh})^2}{kh^2} \frac{a_h}{1 - a_h e^{-kh}} \right].
\] (47)

Analogously, the corresponding cross product equals

\[
E(\hat{\sigma}_t^2 \sigma_t^2) = \frac{\phi_h}{1 - a_h} E\sigma_t^2 + b_h \sum_{j=0}^{\infty} a_h^j E(\xi_t^2 - jh\sigma_t^2).
\]

Substituting the correlation term from equation (36) yields that

\[
E(\hat{\sigma}_t^2 \sigma_t^2) = \frac{\phi_h}{1 - a_h} E\sigma_t^2 + \frac{b_h}{1 - a_h} E^2\sigma_t^2 + b_h Var\sigma_t^2 \frac{1 - e^{-kh}}{kh} \frac{1}{1 - a_h e^{-kh}}.
\] (48)

Thus, the mean-squared error in the spot variance estimate equals

\[
E(\hat{\sigma}_t^2 - \sigma_t^2)^2 = E(\hat{\sigma}_t^4) + E\sigma^4 - 2E(\hat{\sigma}_t^2 \sigma_t^2).
\] (49)

Under what conditions is the estimate \(\hat{\sigma}_t^2\) consistent? Suppose, \(\phi(h) = O(h), \ b_h = O(\sqrt{h}), \ a_h + b_h = 1 - O(h)\). Then, taking limits of the expressions (47),(48) as \(h \to 0\) yields that:

\[
\lim_{h \to 0} E(\hat{\sigma}_t^4 \mid k) = E\sigma^4,
\]

\[
\lim_{h \to 0} E(\hat{\sigma}_t^2 \sigma_t^2 \mid k) = E\sigma^4.
\]

Hence, \(\lim_{h \to 0} E(\hat{\sigma}_t^2 - \sigma_t^2 k)^2 = 0\). Therefore, any GARCH filter satisfying the assumptions outlined above is consistent, irrespective of the model. This result was first proved by Nelson(1992) for a general class of ARCH filters.

What set of coefficients \(a_h, b_h, \phi(h)\) renders the most efficient estimate \(\hat{\sigma}_t^2\)? To answer this question, we minimize the error (49) approximated around \(h = 0\). Asymptotically, the following relation holds:

\[
E(\hat{\sigma}_t^2 - \sigma_t^2 k)^2 \approx h Var\sigma^2 \frac{\tilde{k}}{b_h} + b_h E\sigma^4.
\]

\[
\tilde{k} = \frac{\sum_{i=1}^{p} a_i^2 k_i}{\sum_{i=1}^{p} a_i^2}.
\]

The above error is minimized at \(b_h = \sqrt{\frac{Var\sigma^2}{E\sigma^2 kh}} = \sqrt{\lambda kh}\). For any one-factor model, the efficient coefficients are \(b_h = \sqrt{\lambda kh}, \ a_h = 1 - b_h - \tilde{k} h, \) and \(\phi_h = k\theta h\). After substituting the limit for the estimate of \(k\) given by (43), we find the following values for the “optimal” coefficients:

\[
b_h = \sqrt{\lambda kh},
\]

\[
a_h = 1 - b_h - \tilde{k} h,
\]

\[
\phi_h = k\theta h.
\] (50)
Step 3: Covariance term.
Since the covariance between the genuine forecast error and the spot variance is zero for the unbiased forecast, the covariance term (42) simplifies to:

\[ \text{Cov}(\text{GFE}_t, F(\hat{\sigma}_t^2 - \sigma_t^2)) = -\frac{1 - e^{-k}}{k} \text{Cov}(IV_{t+1}, \text{Var}\sigma_t^2, \sigma_t^2). \]

The covariance can be simplified further, since the integrated variance inside the covariance can be replaced by its conditional expectation \( E_t(IV_{t+1}) = \theta + \sum_{i=1}^{p} \frac{1 - e^{-k_i}}{k_i} P_i(x_t). \)

\[ \text{Cov}(\text{GFE}_t, F(\hat{\sigma}_t^2 - \sigma_t^2)) = -\frac{1 - e^{-k}}{k} \text{Cov}\left[ \sum_{i=1}^{p} \left( \frac{1 - e^{-k_i}}{k_i} - \frac{1 - e^{-k}}{k} \right) P_i(x_t), \sigma_t^2 \right]. \]

Combining the formula (35) with the definition of the filter (46) yields the expression for the above covariance:

\[ \text{Cov}(P_i(x_t), \sigma_t^2) = b_h a_{ih}^2 \frac{1 - e^{-k_i}}{k_i} \sum_{j=0}^{\infty} a_{ij} e^{-k_i j} = b_h a_{ih}^2 \frac{1 - e^{-k_i}}{k_i} \frac{1}{1 - a_h e^{-k_i}}. \]

Hence, the covariance between the genuine error and the error in \( \hat{\sigma}_t^2 \) is equal to

\[ \text{Cov}(\text{GFE}_t, F(\hat{\sigma}_t^2 - \sigma_t^2)) = \text{Var}\sigma_t^2 \frac{1 - e^{-k}}{k} \left( \frac{1 - e^{-k}}{k} - \frac{1 - e^{-k}}{k} \right) \frac{1 - e^{-k h}}{k h} \frac{b_h}{1 - a_h e^{-k h}}. \] (51)

Total MSPE of the model-based forecast.
In our final step, we assemble three parts of the total error of the unbiased model forecast and substitute for the efficient parameters \( a_h, b_h, \phi_h \) from (50). Also, we replace the expectation of the spot variance which enters formulas (47, 48) by \( \sqrt{\frac{1 - e^{-k}}{k}} \text{Var}\sigma_t^2. \)

\[ \text{Total MSPE} = \text{GFE}_t + F(\hat{\sigma}_t^2 - \sigma_t^2) + 2\text{Cov}(\text{GFE}_t, F(\hat{\sigma}_t^2 - \sigma_t^2)) \]

\[ \text{GFE}_t = \text{Var}\sigma_t^2 \left( 2 \left( \frac{1 - e^{-k}}{k} - \frac{1 - e^{-k}}{k^2} \right) \right) \]

\[ F(\hat{\sigma}_t^2 - \sigma_t^2) = \left[ \frac{1 - e^{-k}}{k} \right]^2 E(\hat{\sigma}_t^2 - \sigma_t^2)^2 \]

\[ \text{Cov}(\text{GFE}_t, F(\hat{\sigma}_t^2 - \sigma_t^2)) = \text{Var}\sigma_t^2 \frac{1 - e^{-k}}{k} \left( \frac{1 - e^{-k}}{k} - \frac{1 - e^{-k}}{k} \right) \frac{1 - e^{-k h}}{k h} \frac{b_h}{1 - a_h e^{-k h}}. \] (52)
where:

$$\frac{E(\hat{\sigma}_t^2 - \sigma_t^2)^2}{\text{Var}\sigma_t^2} = \frac{1 - \lambda}{\lambda} \left( \frac{k^2 h^2}{1 - a_h^2} + \frac{2 k b h a_h}{(1 - a_h)^2} - 2 \frac{b_h}{1 - a_h} \right) - 2 b_h \frac{1 - e^{-kh}}{kh} \frac{1}{1 - a_h e^{-kh}} + \frac{1}{\lambda}$$

$$+ \frac{b_h^2}{1 - a_h^2} \left[ \frac{1 - \lambda}{\lambda} \frac{3 - a_h}{1 - a_h} + 6 \frac{\frac{h}{k} - \frac{1 - e^{-kh}}{k^2 h^2}}{h^2} + 2 \left( \frac{1 - e^{-kh}}{k^2 h^2} - \frac{a_h}{1 - a_h e^{-kh}} \right) \right]$$

(53)

$$b_h = \sqrt{\lambda kh}$$  (54)

$$a_h = 1 - b_h - \tilde{k} h$$  (55)

$$\frac{1 - e^{-\tilde{k}}}{\tilde{k}} = \frac{1 - e^{-k}}{k}$$

It follows from the expression above that (1) the total MSPE of the model-based forecast is proportional to the variance of the spot variance and (2) the coefficient of proportionality is a function of parameters $\Xi$ only.

### B.2 MSPE of the Reduced-Based Forecast

In this part, we will estimate the error from reduced-form forecasting. As follows from (15) the total Mean-Squared Prediction Error for the no-leverage case is decomposed as follows:

$$\text{Total MSPE} = \text{Var}\eta_{t+1}(h) - \text{Var}u_{t+1}^2 + E^2 u_{t+1}$$

where the first component is an innovation in the ARMA representation for realized variance (14). The second component $u_{t+1}$ is the difference between the realized variance and the integrated variance. Its variance is expressed by (33) and expectation is typically negligible, and is zero for the case of no-drift in returns.

To obtain the variance of the innovation in the ARMA representation for RV, convert ARMA into an infinite AR-representation:

$$\left[ \prod_{i=1}^{p} (1 - e^{-k_i L}) \right] \left[ \frac{1}{1 - \sum_{i=1}^{p} \beta_i L^i} \right] (RV_{t+1}^t - \theta) = \eta_{t+1}(h).$$

(56)

Slope coefficients in the ARMA representation $\beta_i$ are naturally functions of the correlation structure of RV. On the other hand, from formulas for the moments of IV and RV (30 - 33) the correlation structure of the realized variance depends only on the parameters $\Xi$. Hence, the parameters of the ARMA representation are the functions of only $\Xi$. In particular, Meddahi (2003) derived coefficients for the ARMA representation of realized variance for the case of one-factor and two-factor models.

It then follows that the variance of the innovation is equal to

$$\text{Var}\eta_{t+1}(h) = \text{Var} \left[ \frac{\prod_{i=1}^{p} (1 - e^{-k_i L})}{1 - \sum_{i=1}^{p} \beta_i L^i} \right] RV_{t+1}^t = \text{Var}\sigma_t^2 \Psi(\Xi).$$

(57)

The exact formulas for the above variance in the case of one-factor and two-factor models are also derived by Meddahi (2003).

Summing up the parts of the error and substituting the variance of the noise (33), the total error of the reduced-form forecast equals:

$$E(IV_{t+1}^t - P(IV_{t+1}^t | RV))^2 = \text{Var}\sigma_t^2 \left[ \Psi(\Xi) - 2h^2 \frac{1 - \lambda}{\lambda} + \frac{2 h}{k} \frac{\frac{h}{k} - \frac{1 - e^{-kh}}{k^2}}{1 - a_h e^{-kh}} \right].$$

(58)
B.3 Forecast Comparison

Hence, the total MSPE of both forecasts is proportional to the variance of the spot \( \sigma_t^2 \). The corresponding coefficients of proportionality derived in (52) and (58) are functions of the coefficients \( \Xi \). Hence, the comparison of the forecasts depends only on the parameters from the set \( \Xi \).

B.4 Extension 1: Bias in the Model-Based Forecast

In the above proof, we defined the parameter \( \tilde{k} \) in a way that ensures that the resulting model-based forecast is unbiased. In general, we may assume that the estimated persistence parameter \( \tilde{k} \) satisfies another moment condition:

\[
f(\tilde{k}) = f(k).
\]

For example, the method that matches \( n \)-period covariances of the spot variance results in \( \tilde{k} \), defined by the condition:

\[
e^{-nk} = e^{-nk}.
\]

Here, we derive the contribution of this bias to the total MSPE. To keep the algebra simple, we will assume that for all estimation procedures \( \theta \) matches the unconditional mean of the integrated variance \( IV_t^{t+1} \) as in (37). In this case, irrespective of the choice for the other GMM moments, the “genuine” forecast error will be a sum of two parts

\[
E \left[ IV_t^{t+1} - \hat{\theta} - \frac{1 - e^{-k}}{k} (\sigma_t^2 - \theta) \right]^2 = E \left[ IV_t^{t+1} - P_t(IV) \right]^2 + \text{Var}\sigma_t^2 \left[ \frac{\text{cov}(IV_t^{t+1}, \sigma_t^2)}{\text{Var}\sigma_t^2} - \frac{1 - e^{-k}}{k} \right]^2,
\]

where the first term includes a linear projection of the integrated variance on the most recent spot variance \( P_t(IV_t^{t+1}) \). Therefore, once we allow for model misspecification, the “genuine” forecast error includes two terms: a forecast error from the linear forecast based on the last observed spot and the “bias”. This term is absent if the forecast is based on the parameters defined by (38), i.e., in the “no-bias” case.

The redefined \( F(\hat{\sigma}_t^2 - \sigma_t^2) \) is

\[
F(\hat{\sigma}_t^2, \text{Bias}) = E \left[ E\sigma_t^4 + \frac{\text{cov}(IV_t^{t+1}, \sigma_t^2)}{\text{Var}\sigma_t^2} (\sigma_t^2 - E\sigma_t^2) - \hat{\theta} - \frac{1 - e^{-k}}{k} (\hat{\sigma}_t^2 - \hat{\theta}) \right]^2,
\]

and the covariance term is equal to

\[
\text{Cov}(\text{GFE}_{\text{model}}, F(\hat{\sigma}_t^2, \text{Bias})) = \text{Cov}(IV_t^{t+1} - \frac{\text{cov}(IV_t^{t+1}, \sigma_t^2)}{\text{Var}\sigma_t^2} \sigma_t^2, \frac{\text{cov}(IV_t^{t+1}, \sigma_t^2)}{\text{Var}\sigma_t^2} \sigma_t^2 - \frac{1 - e^{-k}}{k} \sigma_t^2).
\]

Following the same steps from the previous subsection, we find that the total MSPE of the model-based forecast is equal to:

\[
\text{Total MSPE} = \text{GFE}_{\text{model}} + F(\hat{\sigma}_t^2, \text{Bias}) + 2 \text{Cov}(\text{GFE}_{\text{model}}, F(\hat{\sigma}_t^2, \text{Bias})), \quad (59)
\]

\[
F(\hat{\sigma}_t^2, \text{Bias}) = \left[ \frac{1 - e^{-k}}{k} \right]^2 E\sigma_t^4 + \left[ \frac{1 - e^{-k}}{k} \right]^2 E\sigma_t^4 - 2 \left[ \frac{1 - e^{-k}}{k} \right] \left[ \frac{1 - e^{-k}}{k} \right] E[\sigma_t^2 \sigma_t^2], \quad (60)
\]

\[
\text{Cov}(\text{GFE}_{\text{model}}, F(\hat{\sigma}_t^2, \text{Bias})) = \text{Var}\sigma_t^2 \frac{1 - e^{-k}}{k} \left( \frac{1 - e^{-k}}{k} - \frac{1 - e^{-k}}{k} \right) \frac{1 - e^{-kh}}{kh} \frac{b_h}{1 - a_h e^{-kh}}, \quad (61)
\]
where

\[
\frac{E(\hat{\sigma}_t^2)}{\text{Var} \sigma_t^2} = \left[ \frac{k^2 h^2}{1 - a_h^2} + \frac{2k b h}{(1 - a_h)^2} \right] \frac{1 - \lambda}{\lambda} + \frac{b_h^2}{1 - a_h^2} \left[ \frac{1}{\lambda} + 2 \frac{h - 1 - e^{-kh}}{h^2} \right],
\]

\[
E(\hat{\sigma}^2_\sigma^2) = \left[ \frac{kh}{1 - a_h} + \frac{b_h}{1 - a_h} \right] \frac{1 - \lambda}{\lambda} + \frac{b_h}{kh} \frac{1 - e^{-kh}}{1 - a_h e^{-kh}}.
\]

Since the above error takes the form \( \text{Var} \sigma_t^2 f(\Xi) \), then Proposition 1 is still valid for this more general case.

**B.5 Extension 2: Drift in Returns**

Suppose there is a constant drift in returns \( \mu dt \). This generalization of the base case will not change the formula for the errors in the model-based forecast, since we demeaned returns before extracting spot variances. However, the total prediction error of the reduced-form forecast will now be equal to:

\[
\frac{\text{Total MSPE } \text{“Reduced-Form”}}{\text{Var} \sigma_t^2} = \Psi(\Xi) - 2h \frac{1 - \lambda}{\lambda} - \frac{2}{h^2} \left( \frac{h}{k} - \frac{1 - e^{-kh}}{k^2} \right) + \frac{E^2 u_{t+1}}{\text{Var} \sigma_t^2},
\]

where the expectation of the noise is

\[ E u_{t+1} = \mu^2 h. \]

The last term also affects the comparison. However, it is normally small for intra-day data and can be omitted from consideration. In general, constructing \( RV \) as a sum of the squared demeaned returns will result in the same forecast comparison as in the case of zero drift.

**B.6 Extension 3: Long-Horizon Forecast**

For the model-based forecast (27), we decompose the error in the sum of the genuine part and the error coming from the error in spot variance. The genuine part follows from the prediction error of regressing IV on the last observable spot:

\[
\text{GFE}^{\text{model}} = \text{Var} IV_{t+T + 1} = \text{Var} \sigma_t^2 \left( e^{-kT} \frac{1 - e^{-k}}{k} \right) = \text{Var} \sigma_t^2 \left[ \frac{2}{k} \left( 1 + \frac{1 - e^{-k}}{k} \right) \right].
\]

The total error equals:

\[
\text{Total MSPE} = \text{GFE}^{\text{model}} + F(\hat{\sigma}_t^2) + 2 \text{Cov}(\text{GFE}^{\text{model}}, F(\hat{\sigma}_t^2)),
\]

\[
F(\hat{\sigma}_t^2) = \left[ e^{-kT} \frac{1 - e^{-k}}{k} \right]^2 E \hat{\sigma}_t^4 + \left[ e^{-kT} \frac{1 - e^{-k}}{k} \right]^2 E \sigma_t^4 - 2e^{-kT} \frac{1 - e^{-k}}{k} \text{Cov}(\hat{\sigma}_t^2, \sigma_t^2),
\]

\[
\text{Cov}(\text{GFE}^{\text{model}}, F(\hat{\sigma}_t^2)) = \text{Var} \sigma_t^2 e^{-kT} \frac{1 - e^{-k}}{k} \left[ \frac{e^{-kT} - 1 - e^{-k}}{k} \right] \left( \frac{1 - e^{-kh}}{kh} \frac{b_h}{1 - a_h e^{-kh}} \right).
\]
For the reduced-form forecast based on the ARMA, the total MSPE is constructed in the same manner as for the short-term forecasts:

\[ E^2(IV_{t+T}^{T+1} - RV_{t+T}^{T+1}) + \text{Var}(RV_{t+T}^{T+1} - P(RV_{t+T}^{T+1} | RV_{t-j}^{T-j})) - \text{Var}(IV_{t+T}^{T+1} - RV_{t+T}^{T+1}). \]

The unexpected part of the realized variance is calculated by iterating the ARMA model for RV:

\[ RV_{t+T}^{T+1} - P(RV_{t+T}^{T+1} | RV_{t-j}^{T-j}) = \left[ F^T 1 - \sum_{i=1}^{p} \beta_i(h) L^p \prod_{i=1}^{p}(1 - e^{-k_i L}) \right]_{\eta_t+1}(h). \]

Since parameters of the ARMA representation are functions of \( \Xi \) and the variance of the innovation \( \eta_t \) takes the form (57), the variance of the above expression is equal to:

\[ \text{Var}(RV_{t}^{T} - P(RV_{t}^{T} | RV_{t-j}^{T-j})) = \Psi_T(\Xi)\text{Var}\sigma_t^2. \]

The resulting total MSPE equals

\[ \frac{\text{Total MSPE} \text{ "Reduced-Form"}}{\text{Var}\sigma_t^2} = \Psi_T(\Xi) - 2h \frac{1 - \lambda}{\lambda} - \frac{2}{k} \frac{h}{k} - \frac{1 - e^{-kh}}{k^2}. \]

Thus, the comparison for long-term forecasts is similar to the case of short-term forecasts. As before, this comparison depends solely on the parameters in \( \Xi \).

**B.7 Extension 4: Micro-Structure Noise**

**Model-Based Forecast**

As before the model-based forecast is based on the assumption

\[ E_t IV_t^{t+1} = \theta + \frac{1 - e^{-\kappa}}{\kappa}(\sigma^2_t - \theta). \]

If \( \hat{\theta} \) is equal to the sample average of \( \hat{\sigma}_t^2 \), which is the same as the average of \( \xi_t^2 \), then the squared bias of the above forecast is equal to \( BIAS = (E\xi_t^2 - \theta)^2 \) with \( E\xi_t^2 = \theta + \frac{2E\xi_t^2}{h} \). The total MSPE can be decomposed in the following way,

\[ \text{total MSPE} = \text{GFE}^{model} + \text{Var}(P - \hat{P}) + 2\text{cov}(IV_t^{t+H} - P, P - \hat{P}) + BIAS, \]

where \( P \) is a population projection of \( IV_t^{t+1} \) on \( \sigma_t^2 \), and \( \hat{P} \) is the estimated projection. Note, that \( \text{GFE}^{model} \) will remain the same as in the case with no microstructure noise. Denote the original forecast based on observations without noise by \( P^* = \hat{\theta} + \frac{1 - e^{-\kappa}}{\kappa}(\sigma_t^{*2} - \hat{\theta}) \), where \( \sigma_t^{*2} \) is a noise-free filter of the spot variance. Then, the MSPE can be further decomposed,

\[ \text{total MSPE} = \text{GFE}^{model} + \text{Var}(P - P^*) + \text{Var}(P^* - \hat{P}) + 2\text{cov}(P - P^*, P^* - \hat{P}) + 2\text{cov}(IV_t^{t+H} - P, P^* - \hat{P}) + 2\text{cov}(IV_t^{t+H} - P, P^* - \hat{P}) + BIAS. \]

Since \( \hat{P} - P^* = \frac{1 - e^{-\kappa}}{\kappa}(\hat{\sigma}_t^2 - \sigma_t^{*2}) \) and

\[ \hat{\sigma}_t^2 - \sigma_t^{*2} = \frac{b}{h} \sum_{j=0}^{\infty} \sum_{j=0}^{\infty} b \sum_{j=0}^{\infty} \int_{t-jh}^{t-jh} \int_{t-jh}^{t-jh} \sigma_t dW_t, \]
covariances of the difference $P^* - \hat{P}$ with noise-free measures, $IV_t^{t+H}$, $P$, $P^*$ are all zeros. That is, the error has only one extra term, namely the variance $\text{Var}(P^* - \hat{P})$:

$$\text{Var}(P^* - \hat{P}) = \left(1 - e^{-\tilde{H}/\tilde{k}}\right)^2 \frac{b_2^2}{\tilde{k}^2} \left(\frac{2(1 + a_h)\text{Var}_t^2 + 4E^2n_t^2}{1 - a_h^2} + 8\theta E_n^2\right).$$

Reduced-Form Forecast

The arguments of Meddahi (2003) apply to $RV_t^{t+1}$ and $TSRV_t^{t+1}$ under micro-structure noise. These arguments remain valid since corresponding conditional means are linear in factors $P_t(x_t)$. For example, for $TSRV_t^{t+1}$ it holds that:

$$E_tTSRV_t^{t+1} = \theta + \frac{\sum_{i=1}^{p} \frac{1 - e^{-k_i}}{k_i} P_i(x_t) \frac{1}{K} \left[\sum_{j=2}^{K} e^{-k_i(j-1)\frac{h}{K}}\right]}{1 - \frac{1}{K}}.$$ 

Thus, the dynamics of both $RV_t^{t+1}$ and $TSRV_t^{t+1}$ can be described by ARMA$(p, p)$ models. The coefficients of the ARMA representations can be derived in the same way as in Meddahi (2003) by inserting first and second moments of $RV_t^{t+1}$ and $TSRV_t^{t+1}$ into the formulas. The moments of $RV_t^{t+1}$ and $TSRV_t^{t+1}$ are as follows. The expectation of $RV_t^{t+1}$ under micro-structure noise is,

$$ERV_t^{t+1} = \theta + \frac{2E(n_t^2)}{h}.$$ 

That is, $RV_t^{t+1}$ is a biased estimator of the variance. The variance of $RV_t^{t+1}$ is,

$$\text{Var}RV_t^{t+1} = \text{Var}RV_t^{t+1} + (4/h - 2)\text{Var}(n_t^2) + 4/hE^2(n_t^2) + 8\theta E(n_t^2),$$

where $\text{Var}RV_t^{t+1,*}$ is the same variance under no micro-structure noise. The covariances are the same as in the case with no microstructure noise.

In contrast to $RV_t^{t+1}$, $TSRV_t^{t+1}$ is unbiased, i.e., $E TSRV_t^{t+1} = \theta$. Its variance:

$$\text{Var}TSRV_t^{t+1} = \text{Var}TSRV_t^{t+1} + \frac{2(K - 1)\text{Var}(n_t^2) + 8K\theta}{{(K - 1)^2}}$$

where $\text{Var}TSRV_t^{t+1,*}$ is a variance of $TSRV_t^{t+1}$ if $n_t \equiv 0$. Second moments of $TSRV_t^{t+1}$ are derived by Andersen, Bollerslev, and Meddahi (2009).

Using the results of Meddahi (2003), we can find coefficients in the following ARMA representations,

$$\left[\prod_{i=1}^{p} \frac{1 - e^{-k_i}L}{1 - \sum_{j=1}^{p} \beta_{i,j} L^j}\right] RV_t^{t+1} - \theta - \frac{2E(n_t^2)}{h} = \eta_{rv,t+1}(h)$$

Then the MSPE of the RV-reduced-form forecast is

$$\text{total MSPE} = \text{Var} (\eta_{rv,t+1}(h)) - \text{Var}(RV_t^{t+1} - IV_t^{t+1}) + BIAS$$

$$BIAS = \left(\frac{2E(n_t^2)}{h}\right)^2$$

$$\text{Var}(RV_t^{t+1} - IV_t^{t+1}) = (4/h - 2)\text{Var}(n_t^2) + 4/hE^2(n_t^2) + 8\theta E(n_t^2) + \text{Var}(RV_t^{t+1} - IV_t^{t+1}),$$
and the MSPE of the TSRV-reduced-form forecast is

$$\text{total MSPE} = \text{Var} \hat{TSRV}_{t+1} - \text{Var} (\eta_{tsrv,t+1}(h)) + \text{Var} IV_{t+1} - 2\text{cov}(\hat{TSRV}_{t}, IV_{t+1}),$$

where $\hat{TSRV}_{t}$ is a linear projection of $TSRV_{t}$ on past values. From the ARMA representation of TSRV and the definition (29) it follows that

$$\text{cov}(\hat{TSRV}_{t}, IV_{t+1}) \approx \frac{1}{K-1} \text{Var}^2 \sum_{j=2}^{K} \left( \frac{1 - e^{-k}}{k} \right)^2 e^{k(1+\frac{1}{K})} \left( 1 - \left[ \prod_{i=1}^{p} \left( 1 - e^{-k_i} e^{-k} \right) \right] \right).$$
Colored areas denote parameter sets for which the MSPE of the reduced-form forecast is lower than the MSPE of the model-based forecast. Different intensities of grey correspond to different sampling frequencies $h$. The X axis is the ratio of factor loadings $\ln \frac{a_1}{a_2}$. The Y axis is the volatility-of-volatility $\frac{\text{Var}(\eta_t)}{E(\eta_t)}$. The squares correspond to the parameter values from Table 2.
Figure 2: Comparison of Long-Horizon Forecasts (Day after Tomorrow)

Colored areas denote parameter sets for which the MSPE of the reduced-form forecast is lower than the MSPE of the model-based forecast. Different intensities of grey correspond to different sampling frequencies \( h \). The X axis is the ratio of factor loadings \( \ln \left( \frac{a_1}{a_2} \right) \). The Y axis is the volatility-of-volatility \( \text{Var} \left( \frac{\sigma_t}{\sigma_t} \right) \). The squares correspond to the parameter values from Table 2.
Colored areas denote parameter sets for which the MSPE of the reduced-form forecast is lower than the MSPE of the model-based forecast. Reduced-form forecast is constructed using RV for left panels and TSRV for right panels. Different intensities of grey correspond to different levels of microstructure noise measured by the noise-to-signal ratio $\frac{\text{Var}_{\epsilon I\epsilon^t}}{\text{Var}_{\epsilon I\epsilon^t} + 1}$. The X axis is the ratio of factor loadings $\ln \left( \frac{a_1}{a_2} \right)$. Prices are observed at 5-minute intervals. The Y axis is the volatility-of-volatility $\lambda = \frac{\text{Var}_{\epsilon I\epsilon^t}}{\text{Var}_{\epsilon I\epsilon^t} + 1}$. The squares correspond to the parameter values from Table 2.
Table 1: Models with One Component in the Variance

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Specification of the Variance Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square-Root SV</td>
<td>$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta \sigma_t dW_t$</td>
</tr>
<tr>
<td>Log-Volatility</td>
<td>$d\ln \sigma_t^2 = k(\theta - \ln \sigma_t^2)dt + \eta dW_t$</td>
</tr>
<tr>
<td>GARCH SV</td>
<td>$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta \sigma_t^2 dW_t$</td>
</tr>
<tr>
<td>Ornstein-Uhlenbeck</td>
<td>$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta dW_t$</td>
</tr>
</tbody>
</table>

Table 2: Parameters for Two-Factor Models

| Forecast:                   | Model Type | $k_1$  | $k_2$  | $LN|a_1/a_2|$ | $\lambda$ |
|-----------------------------|------------|--------|--------|----------------|------------|
| Bollerslev, Zhou(2002)      | Affine     | 0.57   | 0.07   | 0.025          | 0.103      |
| Data: 5-min DM/USD spot     |            |        |        |                |            |
| Huang, Tauchen(2005)        | Log-normal | 1.386  | 0.00137| 0.976          | 0.860      |
| Data: stock indices         |            |        |        |                |            |
| Alizadeh, Brandt, Diebold(2002) | Log-normal | 0.81/0.95 | 0.02/0.03 | -1.3/-0.3 | 0.66/0.68 |
| Data: daily exchange rates  |            |        |        |                |            |
| Barndorff-Nielsen, Shephard(2002) | CEV        | 3.74   | 0.04   | 0.59           | 0.64       |
| Data: 5-min DM/USD spot     |            |        |        |                |            |

Table reports parameters of the ESV representation for multifactor models estimated on observed data sets; mean-reversions ($k_1$, $k_2$), model misspecification ($LN|a_1/a_2|$), and volatility-of-volatility ($\lambda = \frac{\text{Var}\sigma_t^2}{\text{Ex}\sigma_t^4}$). All the parameters are in daily units. For log-normal models with $\ln \sigma_t^2 = x_{1t} + x_{2t}$, we define $a_1^2 = \text{Var}(\sigma_t^2|x_{1t} = \text{Ex}_{1t})$, $a_2^2 = \text{Var}(\sigma_t^2|x_{2t} = \text{Ex}_{2t})$.

Table 3: GARCH-SV example

<table>
<thead>
<tr>
<th>Forecasting Model</th>
<th>True Model</th>
<th>Model-Based Infeasible</th>
<th>Model-Based Feasible</th>
<th>Reduced-Form Infeasible</th>
<th>Reduced-Form Feasible</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F-GARCH-SV</td>
<td>1F-GARCH-SV</td>
<td>0.977(0.023)</td>
<td>0.938(0.062)</td>
<td>0.958(0.043)</td>
<td>0.910(0.090)</td>
</tr>
<tr>
<td>1F-GARCH-SV</td>
<td>2F-SR-SV(I)</td>
<td>0.819(0.181)</td>
<td>0.674(0.393)</td>
<td>0.699(0.301)</td>
<td>0.646(0.354)</td>
</tr>
<tr>
<td>1F-GARCH-SV</td>
<td>2F-SR-SV(II)</td>
<td>0.328(0.672)</td>
<td>0.295(1.156)</td>
<td>0.315(0.686)</td>
<td>0.312(0.688)</td>
</tr>
</tbody>
</table>

Table reports $R^2$ from the Mincer-Zarnowitz regressions and the normalized MSPE in parenthesis. Feasible forecasts are based on 5-minute returns. 2F-SR-SV(I) is calibrated from Andersen, Bollerslev, and Meddahi (2004) and 2F-SR-SV(II) is calibrated from Barndorff-Nielsen and Shephard (2002). The results are calculated analytically from formulas presented in the Appendix.
Table 4: Parameter Estimates of the One-Factor Model for DM/USD data

<table>
<thead>
<tr>
<th>GMM Based on:</th>
<th>( \hat{k} )</th>
<th>( \hat{\lambda} )</th>
<th>( \hat{\theta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily RV</td>
<td>0.250</td>
<td>0.459</td>
<td>0.508</td>
</tr>
<tr>
<td>Weekly RV</td>
<td>0.047</td>
<td>0.325</td>
<td>0.508</td>
</tr>
<tr>
<td>Monthly RV</td>
<td>0.033</td>
<td>0.257</td>
<td>0.508</td>
</tr>
<tr>
<td>Quarterly RV</td>
<td>0.018</td>
<td>0.216</td>
<td>0.509</td>
</tr>
</tbody>
</table>

Table reports the GMM parameter estimates for the model (17) with \( \lambda = \frac{\text{Var}(\hat{\sigma}_t^2)}{\text{E}(\sigma_t^4)} \). Time is measured in daily units and returns are in percentage form.

Table 5: Mincer-Zarnowitz \( R^2(\text{MSPE}) \) for Day-Ahead Forecasts

<table>
<thead>
<tr>
<th>Forecast</th>
<th>( R^2(\text{MSPE}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model-Based 1-F:</td>
<td></td>
</tr>
<tr>
<td>Daily ( \hat{k} = 0.250 )</td>
<td>0.236 (1.233)</td>
</tr>
<tr>
<td>Weekly ( \hat{k} = 0.047 )</td>
<td>0.324 (0.976)</td>
</tr>
<tr>
<td>Monthly ( \hat{k} = 0.033 )</td>
<td>0.341 (0.866)</td>
</tr>
<tr>
<td>Quarterly ( \hat{k} = 0.018 )</td>
<td>0.356 (0.765)</td>
</tr>
<tr>
<td>Model-Based 2-F:</td>
<td>0.370 (0.698)</td>
</tr>
<tr>
<td>Model-Based 3-F:</td>
<td>0.267 (0.782)</td>
</tr>
<tr>
<td>Reduced-Form</td>
<td>0.357 (0.646)</td>
</tr>
</tbody>
</table>

“Model-Based 1-F” is a forecast based on the one-factor model (17). “Model-Based 2-F” is a forecast based on the two-factor SR-SV model by Bollerslev and Zhou (2002). “Model-Based (3-F)” is a forecast based on the three-factor CEV-SV model with parameters from Barndorff-Nielsen and Shephard (2002). “Reduced-Form” is a linear projection of the realized variance on its past values, with the number of lags chosen by the BIC-criterion. \( \hat{k} \) are estimated mean-reversions of volatility from Table 4. Data: DM/USD spot rates. The parameter estimation period for the one-factor model is December 2, 1986, through June 30, 1999; for the two-factor model it is December 2, 1986, through December 1, 1996. The forecast evaluation period is December 2, 1987, through June 30, 1999.
<table>
<thead>
<tr>
<th>Forecast:</th>
<th>Day-Ahead $IV_T^{T+1}$</th>
<th>+1 Day $IV_T^{T+2}$</th>
<th>+1 Week $IV_T^{T+6}$</th>
<th>+2 Weeks $IV_T^{T+11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced-Form</td>
<td>0.357(0.646)</td>
<td>0.233(0.767)</td>
<td>0.137(0.863)</td>
<td>0.114(0.887)</td>
</tr>
<tr>
<td>Model-Based 1-F</td>
<td>0.356(0.765)</td>
<td>0.216(0.996)</td>
<td>0.115(1.557)</td>
<td>0.105(1.108)</td>
</tr>
<tr>
<td>Model-Based 2-F</td>
<td>0.370(0.698)</td>
<td>0.247(0.815)</td>
<td>0.146(0.914)</td>
<td>0.110(0.946)</td>
</tr>
<tr>
<td>Model-Based 3-F</td>
<td>0.267(0.782)</td>
<td>0.195(0.849)</td>
<td>0.129(0.901)</td>
<td>0.102(0.920)</td>
</tr>
</tbody>
</table>

“Model-Based (1-F)” is a forecast based on the one-factor model. “Model-Based (2-F)” is a forecast based on the two-factor SR-SV model by Bollerslev and Zhou (2002). “Model-Based (3-F)” is a forecast based on the three-factor CEV-SV model with parameters from Barndorff-Nielsen and Shephard (2002). “Reduced-Form” is a linear projection of the realized variance on its past values, with the number of lags chosen by the BIC-criterion. Data: DM/USD spot rates. The parameter estimation period for the one-factor model is December 2, 1986, through June 30, 1999; for the two-factor model it is December 2, 1986, through December 1, 1996. The forecast evaluation period is December 2, 1987, through June 30, 1999.