Integrated Variance Forecasting:

Model-Based vs. Reduced-Form

Natalia Sizova*

Duke University

January 13, 2009

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.

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

JEL classification: C22, C53.

*Please address correspondence to: Department of Economics, Duke University, Box 90097, Durham, NC 27708. Phone: (713)775-5365. Fax: (919) 684-8974. E-mail: natalia.sizova@duke.edu.

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 – has 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 relying 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 availability of high-frequency data. However, only model-based techniques were initially considered somewhat 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 allows researchers to define better proxies for the daily volatility. For example, one of non-parametric proxies – the realized variance – is a sum of squared returns.

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

[2]Reviewed by Tauchen (2004).

It is observable, and hence it admits a direct modeling, as in e.g. 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 investigate this comparison 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.

In order to guarantee results that have the greatest generality, we formulate this comparison analytically within the framework of Meddahi's ESV-models. The path-breaking 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. E.g. daily data can be used to estimate the parameters of the model. Then, we use this model to predict the future integrated variance. For example, SV-models that can be used to form the forecast include affine models by Heston (1993) and 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 and 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 the following parameters: sampling frequency and volatility-of-volatility.

Similar to 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 the theoretical explanation why these parameters 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 will introduce definitions for error components, that will be used throughout the paper. Section 3 studies analytical comparison of feasible forecasts. 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 $W_{i,t}$, $i = \overline{1,n}$, are independent standard Brownian motions. W_t^s is a standard Brownian motion which may correlate with $W_{i,t}, i = \overline{1,n}$, i.e. $\mathrm{corr}(dW_t^s, dW_{i,t}) = \rho_i(x_t)$

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

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

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 $\sigma^2(x_t) \equiv \sigma_t^2$.

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

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

which is assumed to be well-defined. Intuitively, IV defines the variance of the T-period return, i.e. $\mathrm{Var}_t(s_{t+T} - s_t)^2 = E_t(IV_t^{t+T})$, 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 $\sigma_t dW_t^s$. This implies that by forecasting the integrated variance, we aim 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:

$$\sigma^2(x_t) = a_0 + \sum_{i=1}^{\infty} a_i P_i(x_t), \tag{3}$$

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

(i) zero-mean: $EP_i(x_t) = 0$;

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

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

$$E(P_i(x_{t+T})|x_\tau, \tau \leq t) = e^{-k_i T} P_i(x_t). \tag{4}$$

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, $\sigma^2(x_t) = a_0 + \sum_{i=0}^{p} a_i P_i(x_t)$. It should be emphasized that the above representation is equivalent to the representation (1), and was introduced only because it facilitates further analytical derivations.

At this point, we proceed to the definition of two types of forecasts under 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. Additionally, 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^{t+1} minimizes the square-loss function:

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

Mean-squared error is easy to analyze and is the most popular loss-function in the literature. 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^{t+1} conditional on information at time t, i.e. $E(IV_t^{t+1}|F_t)$. For the model described by system (1), let F_t be a σ-algebra generated by prices and states up to time t, i.e. $p_\tau, x_\tau, \forall \tau \leq t$. Therefore, the model implies that $IV_{t+1,t}^{model} =$

$E(IV_t^{t+1}|p_\tau, x_\tau, \forall\tau \le t)$. Moreover, 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:

$$IV_{t+1|t}^{model} = E(IV_t^{t+1}|x_t). \tag{6}$$

The above forecast depends on the current state x_t 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:

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

As was anticipated in (6), the model-based forecast depends only on the latest state realization x_t.

However, the model-based forecast described by (7) is infeasible, since it is based on an unknown model, parameters and states. To define the feasible version of the same forecast, it is common to proceed in the following steps. In the first step, we choose a model and derive a corresponding formula for the model-based forecast. In the next step, we can estimate the parameters and unobserved states using a general method such as MLE, exact or an approximation. And finally, we can plug recovered states and parameters into the formula for the model-based forecast. This resulting structure is referred to as a feasible model-based forecast.

The feasible version of the model-based forecast based on the ESV-representation will take the form:

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

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

The starting point is the ideal case, under which the exact model, parameters and latent states are known. In this case, the only error of the model-based forecast is the "genuine" forecast error:

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

where Θ are the parameters in the model, and x_t are the factors in the model. The "best" forecast minimizes the above value (9) by definition.

However, the total forecast error of the feasible model-based forecast involves two extra parts. The first comes from not knowing the states and coefficients. This part will be referred to as the error in forecast $IV_{t+1|t}$ due to errors in the parameters and states:

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

where $\hat{\Theta}$ are estimated parameters in the model, and $\hat{x}_t$ are estimated states in the model. Hence, the total error of predicting IV_t^{t+1} – the total mean-squared prediction error (MSPE) – has two components:

$$\text{Total MSPE} = E[IV_t^{t+1} - IV_{t+1|t}(\hat{\Theta}, \hat{x}_t)]^2 = \text{GFE}^{model} + F(\hat{x}_t - x_t, \hat{\Theta} - \Theta). \tag{11}$$

The two components of the error are uncorrelated, since the "genuine" error is unpredictable based on information at time t, i.e. $E[IV_t^{t+1} - IV_{t+1|t}|F_t] = 0$, and the error in parameters and states is a function of the information available at time t, i.e. $IV_{t+1|t}(\Theta, x_t) - IV_{t+1|t}(\hat{\Theta}, \hat{x}_t) \in F_t$. The total MSPE is thus simply a sum of two variances.

The second 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 are going to investigate when the above factors may outweigh 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). The formula for the model-based forecast (7) reports the conditional expectation for the integrated variance IV_t^{t+1} 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^{t+1} 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.

Meddahi (2003) derives the coefficients of ARMA models for IV_t^{t+1} in the case of one-factor and two-factor models. In general, ARMA(p,p) for the integrated variance will take the form:

$$\prod_{i=1}^{p}(1 - e^{-k_i}L)(IV_t^{t+1} - \theta) = \eta_{t+1} - \sum_{i=1}^{p} \beta_i \eta_{t+1-i}, \tag{12}$$

where η_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). However, we may also estimate the same parameters from a reduced-form model. That is, since IV_t^{t+1} can be described as a reduced-form ARMA process, we may simply fit a linear time-series model to IV_t^{t+1}. The resulting coefficient estimates will be "model-free". Therefore, the forecast based on this reduced-form time-series model yields a model-free IV_t^{t+1} predictor based on the past realizations of the integrated variance.

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

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

The above definition describes an infeasible version of the reduced-form forecast. This forecast is not feasible, since 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

$$RV_t^{t+1} \equiv \sum_{i=1}^{1/h} (s_{t+ih} - s_{t+(i-1)h})^2. \tag{13}$$

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 Barndorff-Nielsen and Shephard (2004, 2005).

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 absolutely model-free 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. $\hat{IV}_{t+1|t}^{rf} = \hat{P}(IV_t^{t+1}|RV_\tau^{\tau+1}, \tau \leq t)$. This follows from the fact that if there is no drift in returns, then 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):

$$\prod_{i=1}^{p} (1 - e^{-k_i} L)(RV_t^{t+1} - \theta) = \eta_{t+1}(h) - \sum_{i=1}^{p} \beta_i(h) \eta_{t+1-i}(h), \tag{14}$$

where $\eta_t(h)$ is heteroscedastic white noise. In contrast to the ARMA representation of the integrated variance, the ARMA representation for RV depends on the distance between observations h. Coefficients for the above ARMA-representation are derived by Meddahi(2003) for one-factor and two-factor models with no leverage, i.e. when the contemporaneous shocks to the factors x_t and to the return are uncorrelated: $\rho_i(x_t) = 0, \forall i = \overline{1,n}$ in (1) .

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 η_{t+1} in (14) and the variance of the noise term $RV_t^{t+1} - IV_t^{t+1}$:

$$
\begin{aligned}
\text{GFE}^{rf} &= E\left[IV_t^{t+1} - P(IV_t^{t+1}|RV_\tau^{\tau+1}, \tau \le t)\right]^2 = \\
&= E\left[RV_t^{t+1} - P(RV_t^{t+1}|RV_\tau^{\tau+1}, \tau \le t) - (RV_t^{t+1} - IV_t^{t+1})\right]^2 = \\
&= E\left[RV_t^{t+1} - P(RV_t^{t+1}|RV_\tau^{\tau+1}, \tau \le t)\right]^2 - E\left[RV_t^{t+1} - IV_t^{t+1}\right]^2 .
\end{aligned}
\tag{15}
$$

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:

$$
\text{Total MSPE}^{rf} = \text{GFE}^{rf} + F^{rf}(\hat{\Theta}^{rf} - \Theta), \tag{16}
$$

where the second part $F^{rf}(\hat{\Theta}^{rf} - \Theta)$ comes from the errors in the coefficients and Θ^{rf} are the parameters in the regression of RV_t^{t+1} 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 $RV_\tau^{\tau+1}$, $\tau \le t$, therefore it is orthogonal to the error from parameter estimation. When expressing the reduced-form forecast as

$$
I\hat{V}_t^{t+1rf} = \hat{\theta} + \sum_{i=1}^{\infty} \hat{\alpha}_i \left(RV_{t-i}^{t-i+1} - \theta\right), \tag{17}
$$

the error term due to parameter estimation error uncertainty takes the form

$$
F^{rf}(\hat{\Theta}^{rf} - \Theta) = E\left[\hat{\theta} + \sum_{i=1}^{\infty} \hat{\alpha}_i \left(RV_{t-i}^{t-i+1} - \hat{\theta}\right) - \theta - \sum_{i=1}^{\infty} \alpha_i \left(RV_{t-i}^{t-i+1} - \theta\right)\right]^2 . \tag{18}
$$

It is worth noting 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 given to the effects of errors in the state estimates $\hat{x}_t$ and model misspecification.

Throughout this section, we will assume that the data is generated from a multi-factor 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 χ^2-statistics within the EMM estimation. Also, 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 multi-factor 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 multi-factor 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 $k_2 \approx 0$. One example of such a specification is the model by Huang and Tauchen (2005), which mimics the S&P 500 index behavior. In their model, one volatility component has mean-reversion with a half-life of 1/2 days and the other component has mean-reversion with a half-life of two years.

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:

$$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \Lambda(\sigma_t)dW_t. \tag{19}$$

In the above model, the only factor is the spot variance. Therefore, we can make a link between the model (19) and the ESV-representation (3) by defining the ESV-parameters $a_0 = \theta, \quad a_1 = \sqrt{\text{Var}\sigma_t^2}, \quad P_1(x_t) = \frac{\sigma_t^2 - a_0}{a_1}$, and $k_1 = k$.

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

$$E_t IV_t^{t+1} = \theta + \frac{1 - e^{-k}}{k}(\sigma_t^2 - \theta). \tag{20}$$

The model-based forecast is linear in the last realized spot variance and the slope is a function of the mean-reversion coefficient k.

Formula (20) defines an infeasible form of the model-based forecast. To implement this forecast in practice the following problems are to be solved: parameter estimation and estimation of the latent spot variances σ_t^2.

The first problem – estimation of the parameters – will not be fully addressed in this paper. We assume that a large span of data is available, thus bringing 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. For example, for a dummy-model:

$$y_t = \sigma\varepsilon_t, \ \varepsilon_t \propto N(1,1) \tag{21}$$

MLE-estimate of the parameter σ converges to the root of the following equation:

$$1 = E\left[\left(\frac{y_t}{\sigma} - 1\right)\frac{y_t}{\sigma}\right]. \tag{22}$$

On the other hand, a GMM-estimate based on the second moment converges to the variance Var y_t. Unless the above model is correct, those two different moment conditions will generally yield different parameters.

Therefore, the estimation procedure for the model parameters also plays an important role in the feasible model-based forecast. Our first choice is a procedure that leaves forecast (20) unbiased irrespective of the model. This is achieved by a GMM procedure that matches the expectation

$$E(IV_t^{t+1} - \theta) = 0, \tag{23}$$

and the regression slope:

$$\frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2} = \frac{1-e^{-k}}{k}. \tag{24}$$

The case when parameters are defined by the conditions above will be further referred to hereafter as the “no-bias” case. However, instead of the above moments, other moment conditions can be used to define k and θ. In general, estimation procedures based on those other moments may introduce a forecast bias. 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 θ matches the unconditional mean of the integrated variance IV_t^{t+1} as in (23). In this case, irrespective of the choice for the other GMM moments, the "genuine" forecast error will be a sum of two parts

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

where the first term includes a linear projection of the integrated variance on the most recent spot variance:

$$P_t(IV_t^{t+1}) = E\sigma_t^2 + \frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2}\left[\sigma_t^2 - E\sigma_t^2\right] \tag{25}$$

Therefore, once we allow for model misspecification, the "genuine" forecast error includes two terms. The first term is a forecast error from the linear forecast based on the last observed spot. The second term is the "bias". This term is absent if the forecast is based on the parameters defined by (24), i.e. in the "no-bias" case.

The second problem of implementing the model-based forecast is filtering the state σ_t^2. As a result of this next step, another error will be added to the total prediction error. As we plug the estimate $\hat{\sigma}_t^2$ into formula (20), we get the following decomposition of the error for the "no-bias" case:

$$IV_t^{t+1} - \theta - \frac{1-e^{-k}}{k}(\hat{\sigma}_t^2 - \theta) = \left[IV_t^{t+1} - P_t IV_t^{t+1}\right] + \left[\frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2}(\sigma_t^2 - \hat{\sigma}_t^2)\right] \tag{26}$$

The first part of the above expression is the infeasible forecast error of the "last-spot" forecast, i.e. the linear forecast based on σ_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 (26) 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 not lower 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 (26) under typical 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 σ_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 estimates of the variance 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 (19), the efficient ARCH-filter takes the form of a discrete-time GARCH(1,1):

$$\hat{\sigma}_{t+h}^2 = \phi_h + a_h\hat{\sigma}_t^2 + b_h\xi_{t+h} \tag{27}$$

where $\xi_{t+h} = \frac{s_{t+h} - s_t - \mu h}{\sqrt{h}}$ is a normalized innovation in prices. The parameters of the filter ϕ_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 $\lambda = \frac{Var\sigma^2}{E\sigma^4}$.

In the Appendix, we prove the following statement:

Proposition 1. *Let σ_t^2 be square integrable with the correlation function: $corr(\sigma_{t+h}^2, \sigma_t^2) = \frac{\sum_{i=1}^p a_i^2 e^{-k_i h}}{\sum_{i=1}^p a_i^2}$. Suppose we apply the efficient ARCH-filter of the form (27) 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 Ξ:*

- *Volatility-of-volatility – $\lambda = \frac{Var\sigma_t^2}{E\sigma_t^4}$;*
- *Relative weights of factors: $\frac{a_i}{a_1}, i = \overline{2,p}$;*
- *Persistence of factors: $k_i, i = \overline{1,p}$;*
- *Sampling frequency – h.*

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.

All the steps of the analytical comparison are given and proved in the Appendix. In the main body of the paper we present the following corollaries of the proof. First, we will derive asymptotic errors for $h \approx 0$, i.e. arbitrarily high sampling frequencies. In particular, we will characterize the effect of decreasing the sampling frequency on the model-based and the reduced-form forecasts. Second, we will quantify the effect of the misspecification on the model-based forecast. Finally, we will continue 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.

3.1 The Effect of Sampling Frequency

To form the feasible version of the model-based forecast, the state σ_t^2 is to be estimated. In this subsection, we show how the errors in the estimates of σ_t^2 depend on the sampling frequency. Furthermore, we summarize the effects of sampling frequency on model-based and reduced-form forecasts for $h \approx 0$, i.e infinitely active sampling.

The asymptotic behavior of an MLE-estimated spot variance is discussed by Barndorff-Nielsen et al. (2008), 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 (27). The formula is given by (65) in the Appendix and can be further simplified through approximation around $h \approx 0$. For simplicity, let's assume a "no-bias" case, i.e. the parameter $\hat{k}$ converges to $\tilde{k}$, which satisfies the moment condition(24). Then, from (65) it follows that the error in the estimate $\hat{\sigma}_t^2$ equals approximately:

$$E(\hat{\sigma}_t^2 - \sigma_t^2|k)^2 \approx hVar\sigma^2 \frac{\bar{k}}{b_h} + b_h E\sigma^4 \tag{28}$$

$$\bar{k} = \frac{\sum_{i=1}^p a_i^2 k_i}{\sum_{i=1}^p a_i^2} \tag{29}$$

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 [3] of the GARCH-parameter $b_h = \sqrt{\lambda \bar{k} h}$. However, if one assumes a one-factor model and estimates the persistence parameter k from (24), then the chosen GARCH-parameter is:

$$b_h = \sqrt{\lambda \tilde{k} h}, \tag{30}$$

where $\tilde{k}$ satisfies the moment condition in(24). The resulting error in the variance estimate equals

$$E(\hat{\sigma}_t^2 - \sigma_t^2)^2 \approx \text{Var}\sigma_t^2 \sqrt{\frac{h}{\lambda}} (\frac{\bar{k}}{\sqrt{\tilde{k}}} + \sqrt{\tilde{k}}). \tag{31}$$

The total MSPE of the model-based forecast will include the part proportional to the above term and also the term that is proportional to the covariance between the error $\sigma_t^2 - \hat{\sigma}_t^2$ and the "genuine" forecast error. (See (63) in the Appendix.) However, the latter term is of the order $O(h)$. Therefore, approximation of the total MSPE around $h \approx 0$ will contain the "genuine" forecast error, which includes the error from model misspecification, plus the part coming from the term in(31):

$$\frac{\text{Total MSPE}}{\text{Var}\sigma_t^2} = 2\overline{\left[\frac{1}{k} - \frac{1-e^{-k}}{k^2}\right]} - \overline{\left[\frac{1-e^{-k}}{k}\right]^2} + \sqrt{\frac{h}{\lambda}} (\frac{\bar{k}}{\sqrt{\tilde{k}}} + \sqrt{\tilde{k}}) \left[\frac{1-e^{-\tilde{k}}}{\tilde{k}}\right]^2 + O(h). \tag{32}$$

The analogous Taylor decomposition of the reduced-form forecast error given by equation (70) in the Appendix yields:

$$\frac{\text{Total MSPE}}{\text{Var}\sigma_t^2} = \frac{\text{Var}[IV_t^{t+1} - P(IV_t^{t+1} | IV_{t-\tau}^{t+1-\tau}, \tau = 1, 2, ...)]}{\text{Var}\sigma_t^2} + O(h). \tag{33}$$

The $O(h)$ term appears in the above equation, since to predict IV_t^{t+1}, 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

[3]The parameter of the efficient GARCH-filter is derived by Nelson and Foster(1994) for a one-factor GARCH-SV process. Here we extend the result by deriving the parameters of the efficient GARCH-filter for an arbitrary ESV-model with no leverage.

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

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 λ affects the forecast comparison in favor of the model-based forecast. This can be explained by the following consideration. Volatility series are filtered from noisy squared returns. The level of noise in the squared return is proportional to $E\sigma^4 = \text{Var}\sigma + E^2\sigma^2$. 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.

Corollary 2. *As the sampling frequency decreases, the comparative performance of the model-based forecasts deteriorates less for higher levels of volatility-of-volatility* $\lambda = \frac{Var\sigma}{E\sigma^4}$.

3.2 The Effect of Model Misspecification

In the above discussion, model misspecification resulted from application of one-factor models to the modeling of multi-factor 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 $\lambda = \frac{\mathrm{Var}\sigma_t^2}{E\sigma_t^4}$ and the sampling frequency.

First, we will 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 multi-factor models for foreign exchange rates and one dealt with stock indices. We derived the parameters of interest – mean-reversions k_i, model misspecification $\ln a_1/a_2$, and volatility-of-volatility λ – 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 $\left(\ln \frac{a_1}{a_2}, \lambda\right)$ 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 depicted 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 $\ln(a_1/a_2) \approx -1$. 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 sections. First, the graph shows that the decrease in sampling frequency adversely affects the model-based forecast, making it less appealing in comparison to 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 λ. 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 $\ln(a_1/a_2) = -1$. 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 and 30-minute samplings. Finally, for $\lambda = 0.8$, the model-based forecast dominates for the highest frequencies (5-minute and 15-minute) and retreats only for the 30-minute sampling.

Third, a higher persistence favors the reduced-form forecast. For each of the four panels, it holds that k_1 is higher than k_2. The "average" mean-reversion of the variance is equal to $a_1^2(a_1^2 + a_2^2)^{-1}k_1 + a_2^2(a_1^2 + a_2^2)^{-1}k_2$, implying that the mean-reversion of the variance is increasing in $\ln \frac{a_1}{a_2}$. Figure 1 shows that the areas with a better performance of the reduced-form forecast are located to the left from the symmetric case $\ln \frac{a_1}{a_2} = 0$. 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 $|\ln \frac{a_1}{a_2}| \in [0, +\infty)$, with no-misspecification cases located at $\ln \frac{a_1}{a_2} = \pm\infty$, Notably, the areas where the reduced-form forecast is better (shown in grey) are centered at $|\ln \frac{a_1}{a_2}| \approx 0$ 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 (56) is quadratic [4] in the ratio $\frac{a_1^2}{a_2^2}$ and achieves the global maximum at a finite value of $|\ln \frac{a_1}{a_2}|$. That is, the predictive power of the model-based forecast is at its minimum when the spot variance is comprised 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.

[4] For comparison, the "genuine" error is linear in $\frac{a_1^2}{a_2^2}$ for a correctly specified two-factor model(see (57)), and achieves maximum for $\ln \frac{a_1^2}{a_2^2} = -\infty$.

From the results of the last two subsections for day-ahead forecasts, it follows that the reduced-form forecast performs very close to the model-based forecast due to two effects. First, the efficiency of the "best forecast" (model-based) fades away for finite sampling frequencies. The second effect comes from model misspecification, which explains why the reduced-form forecast eventually outperforms the model-based forecast for certain parameter values.(See Figure 1.) This is a general result that holds for all data-generating processes described by (1).

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 proceed with the same assumption from the previous subsection: an econometrician employs a one-factor model of the type (19). Specifically, following Andersen, Bollerslev, and Meddahi (2004), he estimates the parameters to be $k = 0.035$, $\theta = 0.636$ and $\Lambda(\sigma_t) = \sigma\sigma_t^2$, where σ 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 model is estimated by Bollerslev and Zhou (2002) for high-frequency 5-minute spot returns on DM/USD from 1986 until 1996. They suggest the following two-factor affine model:

$$
\begin{aligned}
ds_t &= \sqrt{x_{1t} + x_{2t}} dW_t^s, \\
dx_{1t} &= k_1(\theta_1 - x_{1t})dt + \sigma_1 \sqrt{x_{1t}} dW_{1t}, \\
dx_{2t} &= k_2(\theta_2 - x_{2t})dt + \sigma_2 \sqrt{x_{2t}} dW_{2t}, \\
\mathrm{corr}(dW_t^s, dW_{1t}) = 0,\ & \mathrm{corr}(dW_t^s, dW_{2t}) = 0,
\end{aligned}
\tag{34}
$$

where $\theta_1 = 0.3257$ and $\theta_2 = 0.1786$ are factors' means, $k_1 = 0.5708$ and $k_2 = 0.0757$ are mean-reversions, and $\sigma_1 = 0.2286$ and $\sigma_2 = 0.1096$ define the volatility-of-volatility.

The second candidate is estimated by Barndorff-Nielsen and Shephard (2002). They fitted their CEV-SV model to 5-minute return data on the DM/USD series during the period 1986-1996. The resulting volatility dynamics is represented by the two-factor model:

$$
\sigma_t^2 = x_{1t} + x_{2t}, \tag{35}
$$

$$
Ex_{it} = 0.509\ w_i, \tag{36}
$$

$$
\mathrm{Var} x_{it} = 0.461\ w_i, \tag{37}
$$

where $w_2 = 0.212$, $w_1 = 0.788$, and the corresponding mean-reversions for the factors are $k_1 = 3.74$ and $k_2 = 0.0429$. The resulting volatility-of-volatility λ and the ratio $\ln \frac{a_1}{a_2}$ for both of the models are reported in Table 2.

Summarizing, 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). 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).

The table presents the forecast comparison from two perspectives. On the one hand, it illustrates the effect of misspecification. In particular, for the first row the model is specified correctly, but in the second and third rows the model is misspecified.

On the other hand, 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 (71 - 73).

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 example for the 2F-SR-SV(II) model, 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^2s [5] of the feasible model-based forecast and the feasible reduced-form forecast are generally close. This is an evidence 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 construction.

To summarize, when comparing the model-based forecast and reduced-form forecast, both the possible model misspecification and the finite frequency of price observations should be taken into account. The combination of these factors causes a different ranking of the forecasts. This example serves as an illustration of how the comparative ranking of the model-based and the reduced-form forecasts can switch under realistic assumptions.

3.4 Multi-period Forecasts

It can be of separate interest to investigate how model-based and model-free approaches perform for 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 multi-factor models for longer-horizon predictions.

In this section, we consider the forecasting of the integrated variance IV_t^{t+T} if $T = 5$, i.e.

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

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

where $P(IV_t^{t+1}|IV_{t+1|t})$ is the projection of IV_t^{t+1} on its forecast $IV_{t+1|t}$. It follows that R^2 is a valid measure of the forecast performance that does not, however, takes into account the bias part: $P(IV_t^{t+1}|IV_{t+1|t}) - IV_{t+1|t}$.

weekly forecasts. A multi-period model-based forecast for the ESV-models is a generalization of the formula (7):

$$E_t IV_t^{t+T} = a_0 + \sum_{i=1}^{p} a_i \frac{1 - e^{-k_i T}}{k_i} P_i(x_t). \tag{38}$$

For example, for a one-factor model, the multi-period forecast takes the form:

$$E_t IV_t^{t+T} = \theta + \frac{1 - e^{-kT}}{k} (\sigma_t^2 - \theta). \tag{39}$$

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

$$\lim_{T \to \infty} E_t IV_t^{t+T} = a_0 + \sum_{i=1}^{p} \frac{a_i}{k_i} P_i(x_t). \tag{40}$$

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, monthly, and longer-horizon forecasts will not be properly extracted. For example, for one-factor models, they are hidden within the estimate of the spot variance $\hat{\sigma}_t^2$.

We extend the analysis from Section 3.2 to find the parameters for which the multi-period reduced-form forecast outperforms the model-based forecast. Algebraic computations in the general case are given in the Appendix, and the results for $T = 5$ are presented in Figure 2. As in the one-period case, the parameter sets for which the reduced-form forecast performs the 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 multi-period forecasts the reduced-form forecast is more efficient even for the finest 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 much less.

4 Empirical Example

We have demonstrated that the efficiency of the model-based forecast dissipates, if we consider its feasible version. This 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). One of the features of the DM/USD data set is the presence of strong intra-day patterns in volatility. (See Andersen and Bollerslev, 1997.) The same paper shows that the intra-day seasonal component in DM/USD volatilities may be described as a function of only the time of day and does not depend on the level of volatility. 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 [6] :

$$s^*_{t+ih} - s^*_{t+(i-1)h} = \frac{s_{t+ih} - s_{t+(i-1)h} - \mu h}{\hat{f}(i,h)} + \mu h,$$
$$\hat{f}(i,h)^2 = \frac{\sum_{t=1}^{N}(s_{t+ih} - s_{t+(i-1)h} - \mu h)^2}{\sum_{t=1}^{N}\sum_{i=1}^{1/h}(s_{t+ih} - s_{t+(i-1)h} - \mu h)^2}.$$

We assess the performance of two model-based forecasts. The first one is constructed using a one-factor model. The other is based on a two-factor model. The goal of this section is to compare these forecasts to the reduced-form forecast. The object of the forecasting in all these cases will be the realized variance, RV.

[6] We also verified that the results of this section do not change when using raw data.

4.1 One-Factor Model-Based Forecasts

For the simplest case – a one-factor model – the log-price is assumed to follow the dynamics

$$
\begin{aligned}
ds_t &= \mu dt + \sigma_t dW_t^s, \\
d\sigma_t^2 &= k(\theta - \sigma_t^2)dt + \Lambda(\sigma_t)dW_t, \\
\text{cov}(W_t^s, W_t) &= 0.
\end{aligned} \tag{41}
$$

In the above system, the volatility-of-volatility $\Lambda(.)$ is left unspecified. This generality serves two related purposes. First, if the data are indeed generated from a one-factor model, then regardless of $\Lambda(.)$, there is no model misspecification. Second, the only type of misspecification possible in this setup is the underestimation of the number of factors.

In order to estimate the parameters in (41) we rely on the method-of-moments that is independent of the specification for $\Lambda(.)$. Specifically, we match the mean of the T-period realized variance $\theta = \frac{1}{T} ERV_t^{t+T}$, the correlation structure of the realized variance $e^{-kT} = \text{corr}\left(RV_{t-T}^t, RV_t^{t+T}\right) *$ $*(\text{corr}\left(RV_{t-T}^t, RV_{t+T}^{t+2T}\right))^{-1}$, and the variance of the realized variance to get the volatility-of-volatility $\lambda = \text{Var}\sigma_t^2 / E\sigma_t^4$ from $\text{Var}(RV_t^{t+T})$ in (46). To avoid any biases in favor of the reduced-form forecast, we report parameters 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 ($T = 60$).

Table 4 illustrates that different choices of T do indeed yield quite different sets of parameters. For instance, for daily observations $k \approx 0.25$ in daily units, which implies a half-life of less than 3 days. On the other hand, for quarterly correlations with $k \approx 0.018$ the half-life of a shock is almost 2 months. This dependence of the parameters on T is a special case of a more general result, that parameter estimates for a misspecified model depend on the estimation technique. If the underlying model were truly one-factor, then the estimated profiles $k(T)$ and $\lambda(T)$ would be flat. However, for the observed data the degree of mean-reversion $k(T)$ is decreasing in T. This pattern is easy to replicate by estimating a one-factor model on the data simulated from a multi-factor model.

For the financial volatility, the mean-reversion parameter k is typically within the $0.01 - 0.03$ range in daily units. (See e.g. Andersen et al., 2002, Eraker, Johannes and Polson, 2003.) As follows from Table 4, the mean-reversion corresponding to the quarterly data is within this range.

After estimation, we can proceed to the next step involving filtering of the latent state, i.e. the spot variance. Just as in the analytical section, we rely on the efficient ARCH-filter of Nelson given by (27) with the optimal choice of filter parameters given by (66),(67) in the Appendix. The resulting forecasts are then given by:

$$\widehat{IV}_t^{t+1} = \hat{\theta} + \frac{1 - e^{-\hat{k}}}{\hat{k}}(\hat{\sigma}_t^2 - \hat{\theta}). \tag{42}$$

The performance of this forecasting procedure is reported in the first half of Table 5, which contains Mincer-Zarnowitz R^2s and normalized MSPEs for different parameter sets in Table 4. The forecast performance was evaluated using the whole data sample excluding the first year, i.e. for the period of 1987 - 2007. 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. Just as for the model-based forecast, the performance was evaluated using the whole data set excluding the first year.

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 Two-Factor Model-Based Forecast

The sensitivity of the estimates for k to T also indirectly suggests that the one-factor model is misspecified. A more advanced way to model the variance dynamic is to consider two-factor models. This class of models can capture short-run and long-run dynamics in volatility. The two-factor model considered here is the affine model, which has been already described in Section 3.3. The parameters of this model were estimated by Bollerslev, Zhou (2002) with the same data set on the DM/USD exchange rates, using first and second conditional moments of the realized variance within the GMM framework.

In contrast to the one-factor case, there are two latent states (x_{1t}, x_{2t}) in this model, the first being the fast mean-reverting volatility component, and the second being the slow-moving volatility component. To extract these components, we choose among the methods available for estimation of these two latent series. The most popular and efficient methods are particle filtering and Bayesian methods. In contrast to them, the Kalman filter provides a simple linear updating scheme that does not employ any information about the volatility-of-volatility term. If the functional form for the volatility-of-volatility is known (e.g. in the model (34) it is $\sigma_1\sqrt{x_{it}}$ for a factor $x_{it}, i = \overline{1,2}$), we are better off using the particle filtering. Nevertheless, if the model is misspecified, then any advantage of using the particle filter can be annihilated and the Kalman-filter may produce better proxies for latent states. To avoid a bias in favor of the reduced-form forecast, we apply both methods – the Kalman-filter and the particle-filter – to extract latent states following Barndorff-Nielsen and Shephard(2002), and Durham(2004), respectively.[7]

The second half of Table 5 reports the forecasting performance of the two-factor model. The table shows that in terms of R^2 the model-based forecast with particle filtering appears to be the

[7] 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.

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%). This difference in rankings is a matter of how R^2 and MSPE account for forecast biases. The Mincer-Zarnowitz R^2 depends only on the variance of the forecast error and not on its mean, i.e. the conditional bias is irrelevant. On the other hand, MSPE includes both the variance of the error and its expected mean.

The model-based forecast with Kalman filtering performs slightly worse in terms of the R^2 (35.6%) than both the reduced-form forecast and the forecast with the particle filtering. But in terms of the MSPE, it is the best option out of the model-based forecasts (66.2 %). In order to explore the reason why the particle filter and the Kalman filter give rise to different MSPEs, we examined the dynamics of the forecasts for both of these types. We found that particle filtering delivers visibly over-smoothed predictors.

The conclusion from Table 5 is that, despite a better fitting of the data, the two-factor model produces 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 Multi-Period Forecasts

In Section 3, the analytical results were extended to the comparison of multi-period forecasts. In particular, we demonstrated that multi-period forecasts put more weight on the factors with higher persistence, thus increasing the role of the model misspecification in the forecast comparison. This change in importance between the factors explains the results highlighted by Table 6. In this table we continue the empirical example from the previous subsection and study the performance of one-factor models, the two-factor model, and the reduced-form forecast for the realized variance of the DM/USD spot rate. We focus on longer forecast horizons: weekly ($T = 5$) and monthly ($T = 22$).

Table 6 reports the Mincer-Zarnowitz R^2 and the total MSPE. The one-factor forecast in the table is "quarterly"-calibrated, i.e. k equals 0.018. This parametrization proved to be the most successful for day-ahead forecasting. However, for longer horizons, its performance gradually decays, falling to an R^2 of 27.7%.

In contrast to the one-factor models, the drop in the quality of the reduced-form forecast and of the two-factor model is less dramatic. The R^2 for the two-factor model remains at the levels of 31.6% (particle filtering) and 31.1%, (Kalman filtering). The reduced-form forecast achieves an R^2 of 32.3% for the month-ahead forecast.

For the two-factor model (34), the multi-period forecast (38) can be represented in the form:

$$E_t IV_t^{t+T} = \sum_{i=1}^{2} \left[\theta_i + \frac{1 - e^{-k_i T}}{k_i} (x_{it} - \theta_i) \right]. \tag{43}$$

Like the day-ahead forecasts, its implementation requires time-consuming estimation of the parameters and states $x_{1,t}, x_{2,t}$. In return, it performs only on a par with the reduced-form forecast, which is much simpler to implement. Moreover, its MSPE of 71.1%/82.2% is objectively higher than that of the reduced-form forecast, which is 68.1%. Thus, based on the results in this section, we find that even more sophisticated two-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, and simpler approaches are certainly destined to fail in comparison with this benchmark. That is we 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 to foresee that our results may carry over to more complex settings.

References

[1] Aït-Sahalia, Y., P. A. Mykland, and L. Zhang, 2005, How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise, The Review of Financial Studies 18(2), 351-416.

[2] Alizadeh, S., M. W. Brandt, and F. X. Diebold, 2002, Range-Based Estimation of

Stochastic Volatility Models, The Journal of Finance 57(3), 1047-1091.

[3] Andersen, T.G., L. Benzoni, E. Ghysels, and J. Lund, 2002, An Empirical Investigation of Continuous-Time Equity Return Models, The Journal of Finance 57(3), 1239 - 1284.

[4] Andersen, T.G., and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, Journal of Empirical Finance 4, 115-158.

[5] Andersen, T.G., and T. Bollerslev, 1998, Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies financial markets, The Journal of Finance 53(1), 219-265.

[6] Andersen, T.G., T. Bollerslev, P.F. Christoffersen and F.X. Diebold, 2005, Volatility Forecasting, working paper, NBER.

[7] Andersen, T.G., T. Bollerslev, P.F. Christoffersen, and F.X. Diebold, 2006, Practical Volatility and Correlation Modeling for Financial Market Risk Management, in: M. Carey and R. Stultz, (Eds.), Risks of Financial Institutions, University of Chicago Press for NBER, pp. 513-548.

[8] Andersen, T., Bollerslev, T., Diebold, F.X. and Labys, P., 2003, Modeling and Forecasting Realized Volatility, Econometrica 71(2), 529-625.

[9] Andersen, T.G., T. Bollerslev, and N. Meddahi, 2004, Analytical Evaluation of Volatility Forecasts, International Economic Review 45(4), 1079 - 1110.

[10] Barndorff-Nilsen, O.E. and N. Shephard, 2002, Econometrics Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models, Journal of the Royal Statistical Society 64(2), 253 - 280.

[11] Barndorff-Nilsen, O.E. and N. Shephard, 2004, Power and Bipower Variation with Stochastic Volatility and Jumps, Journal of Financial Econometrics 2, 137.

[12] Barndorff-Nilsen, O.E. and N. Shephard, 2005, How Accurate is the Asymptotic Approximation to the Distribution of Realized Volatility, in: D. W. K. Andrews, J. L. Powell, P. A. Ruud, and J. H. Stock, (Eds.), Identification and Inference for Econometric Models. Essays in Honor of Thomas Rothenberg, Cambridge: Cambridge University Press.

[13] Barndorff-Nielsen, O.E., P. R. Hansen, A. Lunde and N. Shephard, 2008, Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise, Econometrica 76, 1481 - 1536.

[14] Bollerslev, T. and R. Engle, 1986, Modeling the Persistence of Conditional Variances, Econometric Reviews 5, 1-50.

[15] Bollerslev, T. and H. Zhou, 2002, Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility, Journal of Econometrics 109(1), 33 - 65.

[16] Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen, 2003, Alternative Models for Stock Price Dynamics, Journal of Econometrics 116(1-2), 225 - 257.

[17] Corsi, F., 2004, A Simple Long Memory Model of Realized Volatility, working paper, University of Lugano and Swiss Finance Institute.

[18] Duffien D., J. Pan, and K. Singleton, 2000, Transform Analysis and Asset Pricing for Affine Jump-Diffusions, Econometrica 68(6), 1343-1376.

[19] Eraker, B., M.Johannes and N.Polson, 2003, The Impact of Jumps in Volatility and Returns, The Journal of Finance 58(3), 1269 - 1300.

[20] Gallant, A.R. and G. Tauchen, 1998, Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions, Journal of the American Statistical Association 93(441), 10-24.

[21] Gloter, A. and J. Jacod, 2001a, Diffusions with measurement errors. I local asymptotic normality, ESAIM: Probability and Statistics 5, 225242.

[22] Gloter, A. and J. Jacod, 2001b, Diffusions with measurement errors. II measurement errors, ESAIM: Probability and Statistics 5, 243260.

[23] Heston, S., 1993, A closed-form solution for options with stochastic volatility bond and currency options, Review of Financial Studies 6(2), 327-343.

[24] Huang, X. and G. Tauchen, 2005, The Relative Contribution of Jumps to Total Price Variance, Journal of Financial Econometrics 3(4), 456 - 499.

[25] Jacquier, E., N.G.Polson, and P.E.Rossi, 1994, Bayesian Analysis of Stochastic Volatility Models, Journal of Business and Economic Statistics 12(4), 371 - 389.

[26] Johannes, M. and N. Polson, 2003, MCMC Methods for Continuous Time Financial Econometrics, in: Y. Ait-Sahalia and L.P. Hansen (Eds.), Handbook of Financial Econometrics, University of Chicago.

[27] Meddahi, N., 2001, An Eigenfunction Approach for Volatility Modeling, working paper, Centre de Recherche et Développement en Économique.

[28] Meddahi, N., 2002, A Theoretical Comparison Between Intergrated and Realized Volatility, Journal of Applied Econometrics 17(5), 479 - 508.

[29] Meddahi, N., 2003, ARMA Representation of Integrated and Realized Variances, Econometrics Journal 6(2), 335 - 356.

[30] Nelson, D.B., 1992, Filtering and Forecasting with Misspecified ARCH models I, Journal of Econometrics 52, 61 - 90.

[31] Nelson, D.B. and D.P. Foster, 1994, Asymptotic Filtering Theory For Univariate ARCH Models, Econometrica 62(1), 1 - 41.

[32] Patton, A. J. and A. Timmermann, 2007a, Properties of optimal forecasts under asymmetric loss and nonlinearity, Journal of Econometrics 140(2), 884 - 918.

[33] Patton, A. J. and A. Timmermann, 2007b, Testing Forecast Optimality Under Unknown Loss, Journal of the American Statistical Association 102(480), 1172 - 1184.

[34] Pitt, M.K. and N. Shephard, 1999, Filtering Via Simulation: Auxiliary Particle Filters, Journal of the American Statistical Association 94(446), 590 - 599.

[35] Stroud, J.R., P.Müller, and N.G. Polson, 2003, Nonlinear State-Space Models with State-Dependent Variances, Journal of the American Statistical Association 98(462), 377 - 386.

[36] Tauchen, G., 2004, Recent Developments in Stochastic Volatilility: Statistical Modelling and General Equilibrium Analysis, working paper, Duke University.

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 $f(k_i), i =$

$\overline{1,p}$ over all factors with a bar operator:

$$\overline{f}(k) = \frac{\sum_{i=1}^{p} a_i^2 f(k_i)}{\sum_{i=1}^{p} a_i^2},$$

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

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

$$\text{Cov}(IV_t^{t+1}, \sigma_t^2) = \overline{\frac{1-e^{-k}}{k}} \text{Var}\sigma_t^2, \tag{44}$$

$$\text{Var}IV_t^{t+1} = 2\text{Var}\sigma_t^2 \overline{\left[\frac{1}{k} - \frac{1-e^{-k}}{k^2}\right]}, \tag{45}$$

$$\text{Cov}(IV_t^{t+1}, IV_{t+l}^{t+l+1} | l > 0) = \text{Var}\sigma_t^2 \overline{\left(\frac{1-e^{-k}}{k}\right)^2 e^{-k(l-1)}}.$$

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

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

$$\begin{aligned} \text{Cov}(RV_t^{t+1}, RV_{t-j}^{t-j+1}) &= \text{Cov}(IV_t^{t+1}, IV_{t-j}^{t-j+1}), \qquad j \neq 0, \\ \text{Var}RV_t^{t+1} &= \text{Var}IV_t^{t+1} + \text{Var}u_{t+1}, \end{aligned} \tag{46}$$

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

$$\begin{aligned} \text{Var}u_{t+1} &= 2h\left[E^2\sigma^2 + \frac{Var(\sigma^2)}{h^2}\int_0^h\int_0^h \phi(s-\tau)dsd\tau\right] = \\ &= 2h\text{Var}(\sigma_t^2)\left[\frac{1-\lambda}{\lambda} + \frac{2}{h^2}\overline{(\frac{h}{k} - \frac{1-e^{-kh}}{k^2})}\right]. \end{aligned} \tag{47}$$

- 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

$$E\xi_{t+h}^4 = \frac{3}{h^2}E\left[\int_t^{t+h}\int_t^{t+h}\sigma_\tau^2\sigma_s^2 d\tau ds\right] = 3E^2\sigma_t^2 + 3\mathrm{Var}\sigma_t^2\frac{\int_t^{t+h}\int_t^{t+h}\phi(\tau,s)d\tau ds}{h^2},$$
$$E[\xi_{t+h-jh}^2\xi_{t+h-ih}^2]_{i\neq j} = \frac{1}{h^2}E\left[\int_{t-(j+1)h}^{t-jh}\int_{t-(i+1)h}^{t-ih}\sigma_\tau^2\sigma_s^2 d\tau ds\right] =$$
$$= E^2\sigma_t^2 + \mathrm{Var}\sigma_t^2\frac{\int_{t-(j+1)h}^{t-jh}\int_{t-(i+1)h}^{t-ih}\phi(\tau,s)d\tau ds}{h^2}.$$

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 $\mathrm{corr}(\sigma_{t+h}^2,\sigma_t^2) = \frac{\sum_{i=1}^p a_i^2 e^{-k_i h}}{\sum_{i=1}^p a_i^2}$. Therefore, for ESV-models $\phi(t,s) = \overline{e^{-k|t-s|}}$ and

$$E\xi_{t+h}^4 = 3E^2\sigma_t^2 + 6\frac{\mathrm{Var}\sigma_t^2}{h^2}\overline{\frac{h}{k} - \frac{1-e^{-kh}}{k^2}},$$
$$E[\xi_{t+h-jh}^2\xi_{t+h-ih}^2]_{i\neq j} = E^2\sigma_t^2 + \frac{\mathrm{Var}\sigma_t^2}{h^2}\left[\frac{1-e^{-kh}}{k}\right]^2 e^{-k(|j-i|-1)h}. \tag{48}$$

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

$$\mathrm{cov}(P_i(x_t),\xi_{t-jh}^2) = \frac{1}{h}\int_{t-(j+1)h}^{t-jh}\mathrm{cov}[P_i(x_t),\sigma_\tau^2]d\tau = \frac{1}{h}\int_{t-(j+1)h}^{t-jh}\mathrm{cov}[P_i(x_t),P_i(x_\tau)]d\tau =$$
$$= \frac{a_i^2}{h}\int_{t-(j+1)h}^{t-jh}e^{-k_i|t-\tau|}d\tau = \frac{a_i^2}{h}e^{-k_i jh}\frac{1-e^{-k_i h}}{k_i}. \tag{49}$$

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

$$\mathrm{cov}(\sigma_t^2, \xi_{t-jh}^2) = \overline{\mathrm{Var}\sigma_t^2 e^{-kjh} \frac{1-e^{-kh}}{kh}}. \tag{50}$$

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 (19). The GMM approach is used to estimate the parameters, and the states are extracted by the efficient ARCH-filter. The reduced-form forecast employs the ARMA(p,p) model for realized variances RV to predict IV.

For the efficient ARCH-filter, $\hat{\sigma}_t^2$ follows the discrete time GARCH(1,1):

$$\hat{\sigma}_{t+h}^2 = \phi_h + a_h \sigma_t^2 + b_h \xi_{t+h}, \tag{51}$$

where $\xi_{t+h} = \frac{p_{t+h} - p_t - \mu h}{\sqrt{h}}$. The parameters of the model are chosen optimally to minimize asymptotic MSE.

Proposition 1. *Let σ_t^2 be square integrable with the correlation function: corr$(\sigma_{t+h}^2, \sigma_t^2) = \frac{\sum_{i=1}^p a_i^2 e^{-k_i h}}{\sum_{i=1}^p a_i^2}$. Suppose we apply the efficient ARCH-filter of the form (27) 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 Ξ:*

- *Volatility-of-volatility – $\lambda = \frac{Var\sigma_t^2}{E\sigma_t^4}$;*
- *Relative weights of factors: $\frac{a_i}{a_1}, i = \overline{2,p}$;*
- *Persistence of factors: $k_i, i = \overline{1,p}$;*

- *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 will compare the errors and consider a few extensions, model-based forecast with a bias and multi-period 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:

$$\mathrm{GFE}^{model} = E\left[IV_t^{t+1} - E\sigma_t^2 - \frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2}(\sigma_t^2 - E\sigma_t^2)\right]^2, \tag{52}$$

the part that is due to the error in spot variance:

$$F(\hat{\sigma}_t^2 - \sigma_t^2) = E\left[E\sigma_t^2 + \frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2}(\sigma_t^2 - E\sigma_t^2) - \tilde{\theta} - \frac{1-e^{-\tilde{k}}}{\tilde{k}}(\hat{\sigma}_t^2 - \tilde{\theta})\right]^2, \tag{53}$$

and twice the covariance between the above two components:

$$\mathrm{Cov}(\mathrm{GFE}^{model}, F(\hat{\sigma}_t^2 - \sigma_t^2)) = \mathrm{Cov}(IV_t^{t+1} - \frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2}\sigma_t^2, \frac{\mathrm{cov}(IV_t^{t+1}, \sigma_t^2)}{\mathrm{Var}\sigma_t^2}\sigma_t^2 - \frac{1-e^{-\tilde{k}}}{\tilde{k}}\hat{\sigma}_t^2). \tag{54}$$

The latter part is zero if the true model is one-factor. In this case, the "genuine" error is independent of all the past information and, therefore, of the error in the spot variance. However, under model misspecification the last term is no longer a null.

Since, by assumption, the model may be misspecified, the limits for parameter estimates $\tilde{\theta}, \tilde{k}$, and therefore part (53), will depend on the estimation method. For now, we restrict our consideration to the "no-bias" case. Therefore, the limits for the parameters are given by the set of moments(23) and (24). And using formula (44), those moments imply that

$$\frac{1-e^{-\tilde{k}}}{\tilde{k}} = \overline{\frac{1-e^{-k}}{k}}. \tag{55}$$

Therefore, $\tilde{k}$ is a function of Ξ only. Moreover, part (53) simplifies to

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

Further analysis of the model-based forecast error will be organized into three steps. First, we will devise the formula for the "genuine" error (52). Second, we will find $F(\hat{\sigma}_t^2 - \sigma_t^2)$ from (53). Third, we will provide the formula for the covariance term (54).

Step 1: "Genuine" forecast error GFE^{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:

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

Substituting the moment (45) into the above expression yields the formula for the "genuine" error:

$$\mathrm{GFE}^{model} = \mathrm{Var}\sigma_t^2 \left(2\overline{\left[\frac{1}{k} - \frac{1 - e^{-k}}{k^2}\right]} - \overline{\left[\frac{1 - e^{-k}}{k}\right]}^2\right). \tag{56}$$

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

$$\mathrm{GFE}^{\text{p-F}model} = \mathrm{Var}\sigma_t^2 \left(\overline{2\left[\frac{1}{k} - \frac{1 - e^{-k}}{k^2}\right] - \left[\frac{1 - e^{-k}}{k}\right]^2}\right). \tag{57}$$

Step 2: Error due to volatility estimation

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

$$\hat{\sigma}_t^2 = b_h \sum_{j=0}^{\infty} a_h^j \xi_{t-jh}^2 + \frac{\phi_h}{1 - a_h}, \tag{58}$$

where the innovation is equal to $\xi_{t+h} = \frac{\int_t^{t+h} \sigma_\tau dW_\tau^p}{\sqrt{h}}$. 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

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

After substituting moments from (48) we find that

$$E(\hat{\sigma}_t^4) = \frac{\phi_h^2}{(1-a_h)^2} + \frac{2\phi_h b_h}{(1-a_h)^2} E\sigma_t^2 + 3\frac{b_h^2}{1-a_h^2}\left[E^2\sigma_t^2 + 2\mathrm{Var}\sigma_t^2 \overline{\frac{\frac{h}{k} - \frac{1-e^{-kh}}{k^2}}{h^2}}\right] \\ +2\frac{b_h^2}{1-a_h^2}\left[\frac{E^2\sigma_t^2 \, a_h}{1-a_h} + \mathrm{Var}\sigma_t^2 \overline{\frac{(1-e^{-kh})^2}{k^2h^2} \frac{a_h}{1-a_h e^{-kh}}}\right]. \quad (59)$$

Analogously, the corresponding cross product equals

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

Substituting the correlation term from equation (50) yields that

$$E(\hat{\sigma}_t^2 \sigma_t^2) = \frac{\phi_h}{1-a_h} E\sigma_t^2 + \frac{b_h}{1-a_h} E^2\sigma_t^2 + b_h \mathrm{Var}\sigma_t^2 \overline{\frac{1-e^{-kh}}{kh} \frac{1}{1-a_h e^{-kh}}}. \quad (60)$$

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

$$E(\hat{\sigma}_t^2 - \sigma_t^2)^2 = E(\hat{\sigma}_t^4) + E\sigma^4 - 2E(\hat{\sigma}_t^2 \sigma_t^2). \quad (61)$$

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 (59),(60) as $h \to 0$ yields that:

$$\lim_{h \to 0} E(\hat{\sigma}_t^4 | k) = E\sigma^4,$$

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

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 (61) approximated around $h = 0$. Asymptotically, it holds that

$$E(\hat{\sigma}_t^2 - \sigma_t^2|k)^2 \approx hVar\sigma^2 \frac{\bar{k}}{b_h} + b_h E\sigma^4,$$
$$\bar{k} = \frac{\sum_{i=1}^p a_i^2 k_i}{\sum_{i=1}^p a_i^2}.$$

The above error is minimized at $b_h = \sqrt{\frac{Var\sigma^2}{E\sigma^4}\bar{k}h} = \sqrt{\lambda\bar{k}h}$. For any one-factor model, the efficient coefficients are $b_h = \sqrt{\lambda k h}$, $a_h = 1 - b_h - kh$,and $\phi_h = k\theta h$. After plugging in the limit for the estimate of k given by (55), the "optimal" coefficients take values

$$\begin{aligned} b_h &= \sqrt{\lambda\tilde{k}h}, \\ a_h &= 1 - b_h - \tilde{k}h, \\ \phi_h &= \tilde{k}\theta h. \end{aligned} \tag{62}$$

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 (54) simplifies to:

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

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

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

Combining the formula (49) with the definition of the filter (58) yields the expression for the above covariance:

$$\mathrm{Cov}(P_i(x_t), \hat{\sigma}_t^2) = b_h \frac{a_i^2}{h} \frac{1-e^{-k_i h}}{k_i} \sum_{j=0}^{\infty} a_h^j e^{-k_i j h} = b_h a_i^2 \frac{1-e^{-k_i h}}{k_i h} \frac{1}{1-a_h e^{-k_i h}}.$$

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

$$\mathrm{Cov}(\mathrm{GFE}^{model}, F(\hat{\sigma}_t^2 - \sigma_t^2)) = \mathrm{Var}\sigma_t^2 \overline{\frac{1-e^{-k}}{k}}\, \overline{\left(\left[\overline{\frac{1-e^{-k}}{k}} - \frac{1-e^{-k}}{k}\right] \frac{1-e^{-kh}}{kh} \frac{b_h}{1-a_h e^{-kh}}\right)}. \quad (63)$$

Total MSPE of the model-based forecast.

This is the final step, in which we assemble together three parts of the total error of the unbiased model forecast and substitute for the efficient parameters a_h, b_h, ϕ_h from (62). Also, the expectation of the spot variance which enters formulas (59, 60) is replaced by $\sqrt{\frac{1-\lambda}{\lambda}\mathrm{Var}\sigma_t^2}$.

$$\mathrm{Total\ MSPE} = \mathrm{GFE}^{model} + F(\hat{\sigma}_t^2 - \sigma_t^2) + 2\mathrm{Cov}(\mathrm{GFE}^{model}, F(\hat{\sigma}_t^2 - \sigma_t^2))$$

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

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

$$\mathrm{Cov}(\mathrm{GFE}^{model}, F(\hat{\sigma}_t^2 - \sigma_t^2)) = \mathrm{Var}\sigma_t^2 \overline{\frac{1-e^{-k}}{k}}\, \overline{\left([\overline{\frac{1-e^{-k}}{k}} - \frac{1-e^{-k}}{k}] \frac{1-e^{-kh}}{kh} \frac{b_h}{1-a_h e^{-kh}} \right)} \quad (64)$$

where:

$$\frac{E(\hat{\sigma}_t^2 - \sigma_t^2)^2}{\mathrm{Var}\sigma_t^2} = \frac{1-\lambda}{\lambda}\left(\frac{\tilde{k}^2 h^2}{1-a_h^2} + \frac{2\tilde{k} h b_h a_h}{(1-a_h)^2} - 2\frac{b_h}{1-a_h}\right) - 2b_h \overline{\frac{1-e^{-kh}}{kh} \frac{1}{1-a_h e^{-kh}}} + \frac{1}{\lambda}$$
$$+ \frac{b_h^2}{1-a_h^2}\left[\frac{1-\lambda}{\lambda}\frac{3-a_h}{1-a_h} + 6\frac{\overline{\frac{h}{k} - \frac{1-e^{-kh}}{k^2}}}{h^2} + 2\overline{\frac{(1-e^{-kh})^2}{k^2 h^2} \frac{a_h}{1-a_h e^{-kh}}}\right] \quad (65)$$

$$b_h = \sqrt{\lambda \tilde{k} h} \quad (66)$$

$$a_h = 1 - b_h - \tilde{k} h \quad (67)$$

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

From the expression above, it follows that the total MSPE of the model-based forecast is proportional to the variance of the spot variance and the coefficient of proportionality is a function of parameters Ξ 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 in the following way:

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

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 (47) 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:

$$\left[\frac{\prod_{i=1}^{p}(1-e^{-k_i}L)}{1-\sum_{i=1}^{p}\beta_i L^i}\right](RV_t^{t+1}-\theta) = \eta_{t+1}(h). \tag{68}$$

Slope coefficients in the ARMA representation β_i are naturally functions of the correlation structure of RV. On the other hand, from formulas for the moments of IV and RV (44 - 47) the correlation structure of the realized variance depends only on the parameters Ξ. Hence, the parameters of the ARMA representation are the functions of only Ξ. In particular, Meddahi(2003) derived coefficients for the ARMA representation of realized variance for the case of one-factor and two-factor models.

Then it follows that the variance of the innovation is equal to

$$\text{Var}\eta_{t+1}(h) = \text{Var}\left(\left[\frac{\prod_{i=1}^{p}(1-e^{-k_i}L)}{1-\sum_{i=1}^{p}\beta_i L^i}\right] RV_t^{t+1}\right) = \text{Var}\sigma_t^2\Psi(\Xi). \tag{69}$$

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 (47), the total error of the reduced-form forecast equals:

$$E(IV_t^{t+1} - P(IV_t^{t+1}|RV))^2 = \mathrm{Var}\sigma_t^2 \left[\Psi(\Xi) - 2h\frac{1-\lambda}{\lambda} - \frac{2}{h^2}\overline{(\frac{h}{k} - \frac{1-e^{-kh}}{k^2})} \right]. \tag{70}$$

B.3 Forecast Comparison

Hence, the total MSPE of both forecasts is proportional to the variance of the spot $\mathrm{Var}\sigma_t^2$. The corresponding coefficients of proportionality derived in (64) and (70) are functions of the coefficients Ξ. Hence, the comparison of the forecasts depends only on the parameters from the set Ξ . $\square$

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:

$$f(\tilde{k}) = \overline{f(k)}.$$

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

$$e^{-n\tilde{k}} = \overline{e^{-nk}}.$$

If we allow for a bias and follow the same steps from the previous subsection, we find that the total MSPE of the model-based forecast is equal to:

$$\text{Total MSPE} = \text{GFE}^{model} + F(\hat{\sigma}_t^2, Bias) + 2\ \text{Cov}(\text{GFE}^{model}, F(\hat{\sigma}_t^2, Bias)), \quad (71)$$

$$F(\hat{\sigma}_t^2, Bias) = \left[\frac{1-e^{-\tilde{k}}}{\tilde{k}}\right]^2 E\hat{\sigma}_t^4 + \left[\overline{\frac{1-e^{-k}}{k}}\right]^2 E\sigma_t^4 - 2\frac{1-e^{-\tilde{k}}}{\tilde{k}}\overline{\frac{1-e^{-k}}{k}}E[\hat{\sigma}_t^2\sigma_t^2], \quad (72)$$

$$\text{Cov}(\text{GFE}^{model}, F(\hat{\sigma}_t^2, Bias)) = \text{Var}\sigma_t^2\frac{1-e^{-\tilde{k}}}{\tilde{k}}\overline{\left([\overline{\frac{1-e^{-k}}{k}} - \frac{1-e^{-k}}{k}]\frac{1-e^{-kh}}{kh}\frac{b_h}{1-a_h e^{-kh}}\right)}, \quad (73)$$

where

$$\frac{E(\hat{\sigma}_t^4)}{\text{Var}\sigma_t^2} = [\frac{\tilde{k}^2h^2}{1-a_h^2} + \frac{2\tilde{k}hb_h}{(1-a_h)^2}]\frac{1-\lambda}{\lambda} + 3\frac{b_h^2}{1-a_h^2}\left[\frac{1-\lambda}{\lambda} + 2\frac{\overline{\frac{h}{k} - \frac{1-e^{-kh}}{k^2}}}{h^2}\right]$$

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

$$\frac{E(\hat{\sigma}_t^2\sigma_t^2)}{\text{Var}\sigma_t^2} = [\frac{\tilde{k}h}{1-a_h} + \frac{b_h}{1-a_h}]\frac{1-\lambda}{\lambda} + b_h\overline{\frac{1-e^{-kh}}{kh}\frac{1}{1-a_h e^{-kh}}}.$$

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 μdt. 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:

$$\frac{\text{Total MSPE "Reduced-Form"}}{\text{Var}\sigma_t^2} = \Psi(\Xi) - 2h\frac{1-\lambda}{\lambda} - \frac{2}{h^2}(\overline{\frac{h}{k} - \frac{1-e^{-kh}}{k^2}}) + \frac{E^2u_{t+1}}{Var\sigma^2},$$

where the expectation of the noise is

$$Eu_{t+1} = \mu^2 h.$$

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 : Multi-Period Forecast

For the model-based forecast (39), 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 and can be found from R^2 of the Mincer-Zarnowitz regression derived by Andersen, Bollerslev, and Meddahi(2004):

$$\begin{aligned}\mathrm{GFE}^{model} &= \mathrm{Var}IV_t^{t+T} - \mathrm{Var}\sigma_t^2\left(\overline{\frac{1-e^{-kT}}{k}}\right)^2 = \\ &= \mathrm{Var}\sigma_t^2\left[\overline{\frac{2}{k}[T+\frac{1-e^{-kT}}{k}]} - \left(\overline{\frac{1-e^{-kT}}{k}}\right)^2\right].\end{aligned}$$

The total error equals:

$$\begin{aligned}\text{Total MSPE} &= \mathrm{GFE}^{model} + F(\hat\sigma_t^2, Bias) + 2\ \mathrm{Cov}(\mathrm{GFE}^{model}, F(\hat\sigma_t^2, Bias)),\\ \mathrm{GFE}^{model} &= \mathrm{Var}\sigma_t^2\left[\overline{\frac{2}{k}[T+\frac{1-e^{-kT}}{k}]} - \left(\overline{\frac{1-e^{-kT}}{k}}\right)^2\right],\\ F(\hat\sigma_t^2, Bias) &= \left[\frac{1-e^{-\tilde{k}T}}{\tilde{k}}\right]^2 E\hat\sigma_t^4 + \left[\overline{\frac{1-e^{-kT}}{k}}\right]^2 E\sigma_t^4 - 2\frac{1-e^{-\tilde{k}T}}{\tilde{k}}\overline{\frac{1-e^{-kT}}{k}}E[\hat\sigma_t^2\sigma_t^2],\\ &\mathrm{Cov}(\mathrm{GFE}^{model}, F(\hat\sigma_t^2, Bias)) = \\ &= \mathrm{Var}\sigma_t^2\frac{1-e^{-\tilde{k}T}}{\tilde{k}}\overline{\left([\overline{\frac{1-e^{-kT}}{k}} - \frac{1-e^{-kT}}{k}]\frac{1-e^{-kh}}{kh}\frac{b_h}{1-a_h e^{-kh}}\right)}.\end{aligned}$$

For the reduced-form forecast based on the ARMA, the total MSPE is constructed in the same manner as for the one-period case:

$$\text{Total MSPE} = E^2(IV_t^{t+T} - RV_t^{t+T}) + \mathrm{Var}(RV_t^{t+T} - P(RV_t^{t+T}|RV_{t-j}^{t+T-j})) - \mathrm{Var}(IV_t^{t+T} - RV_t^{t+T}).$$

For zero drifts in returns, the difference between the integrated variance and the realized variance is unpredictable, and therefore serially uncorrelated. Hence, the variance of the noise is equal to

$$\mathrm{Var}(IV_t^{t+T} - RV_t^{t+T}) = T\mathrm{Var}u_t^{t+1}.$$

The innovation in the realized variance can be found by iterating the ARMA for the RV:

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

Since parameters of the ARMA-representation are functions of Ξ and the variance of the innovation η_t takes the form (69), the variance of the above expression is equal to:

$$\mathrm{Var}(RV_t^{t+T} - P(RV_t^{t+T}|RV_{t-j}^{t+T-j})) = \Psi_T(\Xi)\mathrm{Var}\sigma_t^2.$$

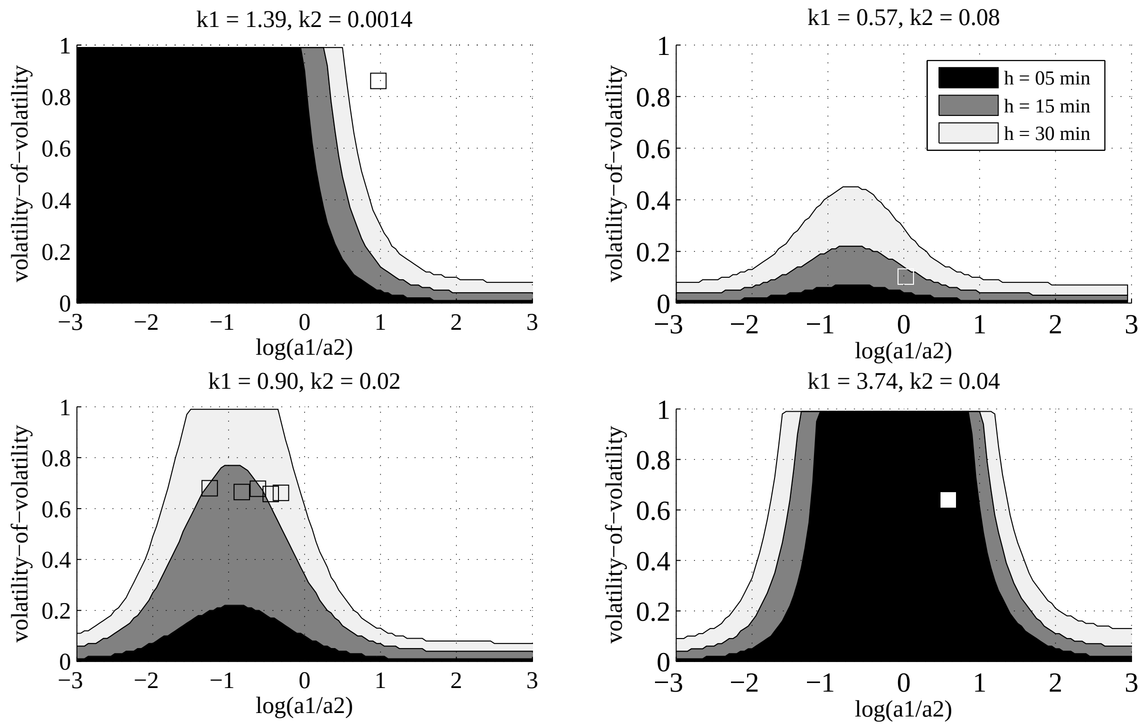
The resulting total MSPE equals

$$\frac{\text{Total MSPE “Reduced-Form”}}{\mathrm{Var}\sigma_t^2} = \Psi_T(\Xi) - 2hT\frac{1-\lambda}{\lambda} - \frac{2T}{h^2}\overline{(\frac{h}{k} - \frac{1 - e^{-kh}}{k^2}})).$$

Thus, the comparison for multi-period forecasts is similar to the case of one-period forecasts. As before, this comparison depends solely on the parameters in Ξ.

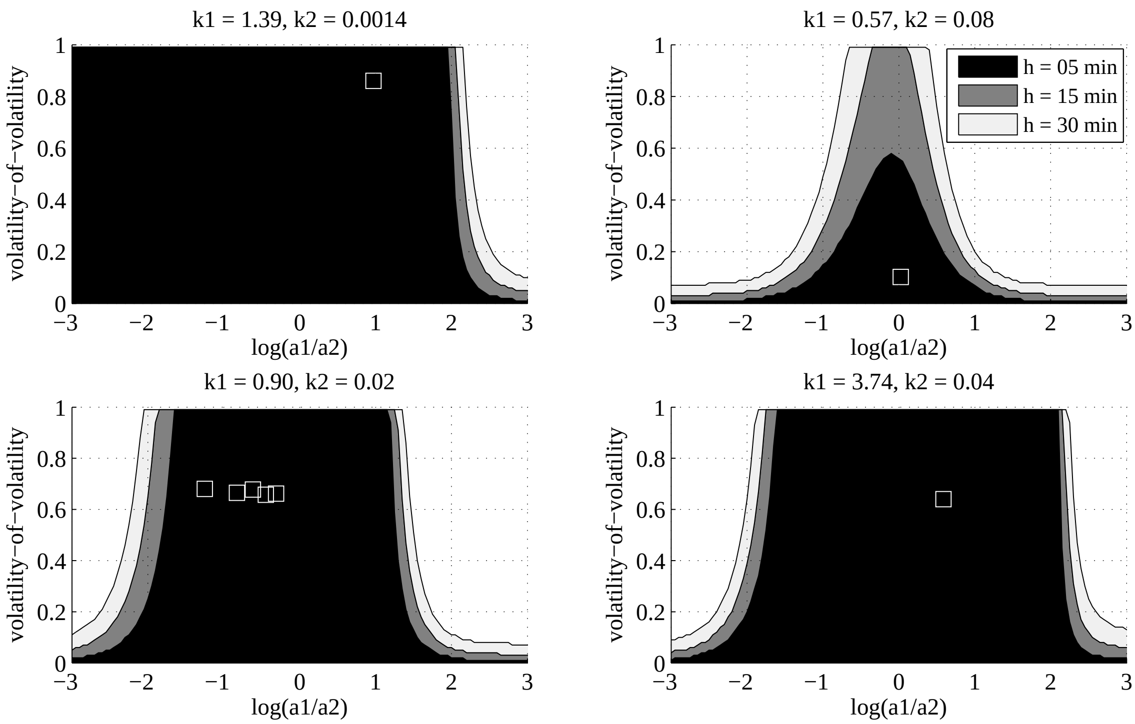
C Tables and Figures

Figure 1: Comparison of Day-Ahead Forecasts



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 gray 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 $\frac{\text{Var}\sigma_t^2}{E\sigma_t^4}$. The squares correspond to the parameter values from Table 2.

Figure 2: Comparison of Week-Ahead Forecasts



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 gray 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 $\frac{\text{Var}\sigma_t^2}{E\sigma_t^4}$. The squares correspond to the parameter values from Table 2.

Table 1: Models with One Component in the Variance

Model Name	Specification of the Variance Process
Square-Root SV	$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta\sigma_t dW_t$
Log-Volatility	$d\ln\sigma_t^2 = k(\theta - \ln\sigma_t^2)dt + \eta dW_t$
GARCH SV	$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta\sigma_t^2 dW_t$
Ornstein-Uhlenbeck	$d\sigma_t^2 = k(\theta - \sigma_t^2)dt + \eta dW_t$

Table 2: Parameters for Two-Factor Models

Forecast:	Model Type	k_1	k_2	$LN(a_1/a_2)$	λ
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 multi-factor models estimated on observed data sets; mean-reversions (k_1, k_2), model misspecification ($\ln(a_1/a_2)$), and volatility-of-volatility ($\lambda = \frac{\mathrm{Var}\sigma^2}{E\sigma^4}$). All the parameters are in daily units. For log-normal models with $\ln \sigma_t^2 = x_{1t} + x_{2t}$, we define $a_1^2 = \mathrm{Var}(\sigma_t^2 | x_{t2} = Ex_{t2})$, $a_2^2 = \mathrm{Var}(\sigma_t^2 | x_{t1} = Ex_{t1})$.

Table 3: GARCH-SV example

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

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

	$\hat{k}$	$\hat{\lambda}$	$\hat{\theta}$
GMM based on: :			
- daily RV :	0.250	0.459	0.508
- weekly RV :	0.047	0.325	0.508
- monthly RV :	0.033	0.257	0.508
- quarterly RV :	0.018	0.216	0.509

Table reports the GMM parameter estimates for model (41). Time is measured in daily units and returns are in percentage form.

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

Forecast		$R^2(MSPE)$
Model-Based(1-F):		
Daily	$\hat{k} = 0.250$	0.236 (1.233)
Weekly	$\hat{k} = 0.047$	0.324 (0.976)
Monthly	$\hat{k} = 0.033$	0.341 (0.866)
Quarterly	$\hat{k} = 0.018$	0.356 (0.765)
Model-Based(2-F):		
2-Factor(Particle)		0.370 (0.698)
2-Factor(Kalman)		0.356 (0.662)
Reduced-Form		0.357 (0.646)

"Model-Based (1-F)" is a forecast based on the one-factor model (41). "Model-Based (2-F)" is a forecast based on the two-factor SR-SV model (34). "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 6: Mincer-Zarnowitz R^2(MSPE) for Multi-Period Forecasts

Forecast:	Day Ahead	Week Ahead	Month Ahead
Model-Based 2-F (Particle)	0.370(0.698)	0.393(0.718)	0.316(0.822)
Model-Based 2-F (Kalman)	0.356(0.662)	0.345(0.666)	0.311(0.711)
Model-Based 1-F	0.356(0.765)	0.340(1.031)	0.277(1.307)
Reduced-Form	0.357(0.646)	0.377(0.623)	0.323(0.681)

"Model-Based (1-F)" is a forecast based on the one-factor model (41). "Model-Based (2-F)" is a forecast based on the two-factor SR-SV model (34). "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.