Volatility Forecasting and Explanatory Variables: A Tractable Bayesian Approach to Stochastic Volatility

Nicolas Chapados
Department of Computer Science and Operations Reasearch, University of Montreal, Montreal, Canada (chapados@iro.umontreal.ca)

Christian Dorion*
Department of Finance, HEC Montreal, Montreal, Canada (christian.dorion@hec.ca)

May 26, 2013
First Draft: June 2, 2010

Abstract We provide a formulation of stochastic volatility (SV) based on Gaussian process regression (GPR). Forecasting volatility out-of-sample, both simulation and empirical analyses show that our GPR-based stochastic volatility (GPSV) model clearly outperforms SV and GARCH benchmarks, especially at long horizons. Most importantly, our approach enables the straightforward incorporation of arbitrary covariates without requiring the specification of functional forms a priori. Augmenting the GPSV model with exogenous variables increases its performance even further. In particular, a simple set of covariates reduces the error rate on one-year out-of-sample forecasting during the 2007-09 recession by 26% relative to a benchmark range-based SV model.

Keywords Bayesian volatility models; Stochastic volatility; Macro-finance variables; Generalized autoregressive conditional heteroscedasticity models; Long memory in volatility; Multifactor volatility.

*We would like to thank Carole Bernard, Peter Christoffersen, Adlai Fisher, and participants at the 2012 IFSID's Conference on Structured Products and Derivatives for insightful comments. Thanks to the Institut de Finance Mathématique de Montréal (IFM2) for financial support.

Contents

1 Introduction

Forecasting asset price volatility at long horizons is a challenging task. Accurately forecasting volatility has, for some time, been thought to be feasible only at short horizons (West and Cho 1995; Christoffersen and Diebold 2000). Yet, recent work by Brandt and Jones (2006) and Engle, Ghysels, and Sohn (2013), amongst others, show that respectable long-horizon forecasts can be obtained by using, in the words of Brandt and Jones, "less misspecified volatility models and more informative volatility proxies." This paper pushes this logic further by showing how a Bayesian regression treatment of volatility enables high-quality long-horizon forecasts at manageable computational cost.

Our contribution is threefold. First, on the methodological front, we frame volatility modeling as a regression problem by recognizing the link between classical stochastic volatility models, combined with range-based volatility measurements, and estimation techniques from the field of Bayesian nonparametrics. A second methodological contribution is to show how to provide arbitrary covariates for volatility forecasting in a nonparametric setting, without having to specify *a priori* the functional forms incorporating them. Finally, we show the empirical benefits of being able to engage in a quasi-analytical, fully Bayesian treatment, without the complexities of MCMC inference, along with the greater estimation stability brought forth by this approach on long-horizon volatility forecasts. Our empirical results also show that *GPR-based stochastic volatility* (GPSV) models can fruitfully exploit the informational content of different economic variables to better forecast volatility.

Spawned by Schwert's (1989) seminal paper, a large body of literature now exists to understand the determinants of asset price volatility and how they can be incorporated within models to improve forecasts. The impact of macro-finance variables on volatility has been studied in a variety of frameworks including that of Flannery and Protopapadakis (2002), who use a number of dummies and macroeconomic observables that impact both the conditional expected return and variance, or in the MIDAS-GARCH model of Engle, Ghysels, and Sohn (2013), using distributed lags of mixed sampled macroeconomic data. Studies by Engle and Rangel (2008), Corradi, Distaso, and Mele (2013), and Dorion (2012) also suggest that the use of various macroeconomic indicators can improve the forecasting ability of volatility models.[1]

In all cases, explicit functional forms and concomitant parametrization need to be specified to accommodate explanatory variables, and all sources of nonlinearity must be

[1] At the intraday frequency, an equally important body of literature studies seasonalities in intraday volatility, in the footsteps of Andersen and Bollerslev (1998).

chosen by the econometrician. As additional variables and corresponding parameters are added, the estimation of such models is increasingly challenging. Moreover, with more complex models, parameter estimation uncertainty compounds when making long-horizon forecasts and may become significant relative to process uncertainty, severely impacting accuracy.[2]

This paper aims to address both the explanatory variable issue and estimation stability difficulty by placing volatility modeling and forecasting in a nonparametric Bayesian regression framework. To this end, we principally call upon the technique of *Gaussian process regression* (GPR), which has become well established over the past fifteen years in the fields of nonparametric Bayesian statistics and machine learning.[3] A Gaussian process (GP) is a (possibly uncountably infinite) collection of random variables, any finite number of which have a joint Gaussian distribution. Regression with Gaussian processes relies on an appealingly simple application of Bayes' rule: it starts by assuming that the function to be modeled belongs to a class of functions characterized by a GP, which constitutes a functional prior distribution. When combined with a likelihood function obtained from observed data, we form a posterior distribution over functions that are, in a sense, "compatible" with the observed data. Moreover, if the data likelihood is Gaussian, then the posterior is itself a GP, whose distribution at a finite number of points can be computed analytically by simple matrix operations.

There exists an intimate connection between GPs and stochastic volatility. In those models, it is often assumed that the (log-)volatility follows an Ornstein-Uhlenbeck (mean-reverting) process. This process *is* a GP, whose covariance function belongs to the so-called Matérn class (Matérn 1960). While Robinson (2001) establishes clear links between stochastic volatility models and the study of the associated Gaussian process covariance function, estimation and forecasting in a nonparametric Bayesian regression framework is virtually absent from the stochastic volatility literature.

There is, of course, a vast literature on Bayesian estimation of parametric volatility

[2] See Stambaugh (1999), Barberis (2000), Hoevenaars, Molenaar, Schotman, and Steenkamp (2011), and Pástor and Stambaugh (2012) for insightful discussions on the potential impact parameter uncertainty in forecasting applications.

[3] Gaussian process regression, as used in this paper, is closely related to the technique of "kriging" originally developed in geostatistics (Matheron 1973; Cressie 1993; Stein 1999). Helpful references include, amongst many others, O'Hagan (1978), Williams and Rasmussen (1996), and Rasmussen and Williams (2006). The mathematical theory of Gaussian processes is developed in any number of textbooks, e.g. Adler and Taylor (2007).

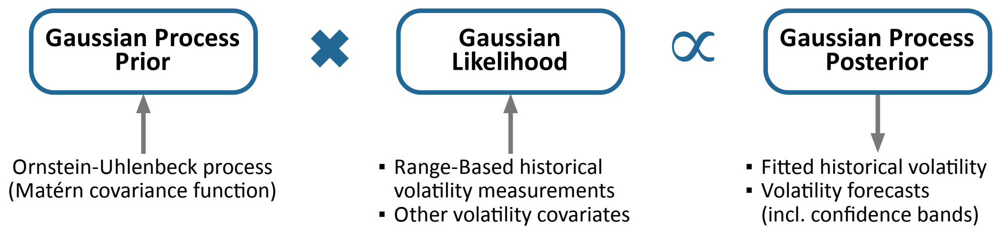


Figure 1: *Summary of the proposed Bayesian volatility modeling technique, illustrating how Gaussian process regression combines the various ingredients being combined by the application of Bayes' rule. A Gaussian process prior combined with a Gaussian likelihood yields an* ***analytical*** *Gaussian process posterior.*

models, starting with the seminal papers of Jacquier, Polson, and Rossi (1994), and Kim, Shephard, and Chib (1998). Most of these approaches involve, in estimation or forecasting, a variant of Monte Carlo Markov Chains (MCMC). Despite the powerful theoretical appeal of the Bayesian framework, this approach to stochastic volatility has not been widely adopted, in part due to the complexity and the computational burden associated with MCMC.

However, MCMC would be unnecessary if the prior and likelihood formed a conjugate pair (Figure 1). For volatility modeling with GP priors, this appears—at first—unlikely, since the usual volatility proxies such as (log) absolute or squared returns are highly non-Gaussian. Fortunately, as usefully pointed out by Alizadeh, Brandt, and Diebold (2002), there *does* exist a volatility proxy that is so nearly normally distributed that it can, for all intents and purposes, be taken as such: the *log price range*, computed as the logarithm of the difference between the high and low log-prices during a day (or other time period).[4] The present paper uses the range as a key ingredient in a Bayesian model of volatility: by virtue of their near normality, range-based measurements yield a Gaussian likelihood function, which is conjugate to the GP prior already at the heart of a great number of stochastic volatility models. Hence the "Bayesian loop" can be closed, and efficient inference—when compared to MCMC at least—carried out.

An additional benefit is to easily incorporate any number of additional volatility co-variates into a single coherent framework. Using futures data on the E-mini S&P 500, we compare the out-of-sample volatility forecasts of the SV, GARCH and GPSV models. The models are estimated using daily range data, and their out-of-sample forecasts are eval-

[4]Range-based techniques have been shown by Garman and Klass (1980) and Parkinson (1980) to be more statistically efficient for volatility estimation than squared returns, and have been applied to stochastic volatility and GARCH models respectively by Alizadeh, Brandt, and Diebold (2002) and Brandt and Jones (2006).

uated against the evolution of the five-minute realized volatility, which is derived using high-frequency stock market data. Without resorting to any observables, the GPSV model outperforms standard SV and GARCH models at forecasting over long horizons. When seasonality and macro-finance variables are accounted for, the forecasting accuracy of the augmented GPSV model, one day to one year ahead out of sample, becomes significantly better.

This article is organized as follows. We review standard dynamic volatility models in Section 2 along with related properties of the range. Section 3 provides an overview of the theory behind Gaussian processes, and links dynamic volatility processes to the Matérn covariance function, illustrating how they can be employed to build GPSV models. Section 4 presents the results of a simulation study to establish that GPSVs perform comparably to established filtering techniques when the underlying volatility generating process is known. In Section 5, we turn to the estimation and forecasting of volatility on actual stock market data and show that, in this context, GPSV models significantly outperform established models, particularly when additional macro-finance variables are used. Section 6 concludes and suggests avenues for future research.

2 Dynamic Volatility Models

Volatility of financial asset prices is now widely accepted to be time-varying and mean-reverting, and while it is highly volatile, it is also highly persistent; that is, the covariance $\mathrm{Cov}(\sigma_t^2, \sigma_{t+j}^2)$ decays slowly with j. While GARCH models (Engle 1982; Bollerslev 1986) acknowledge this issue, they rely on the assumption that shocks to asset prices and asset volatilities arise from a single source, yielding a measurable volatility process. In contrast, stochastic volatility (SV) models posit that, while they can be correlated, these shocks arise from two different sources. Volatility is thus an unobservable process driven by non-measurable shocks.[5] The typical approach to SV models involves parameters describing the speed of mean reversion, the mean-reversion level and the volatility of volatility.

This paper considers the use of a GP representation of such a volatility process. Before doing so, however, this section reviews recent advances in the estimation of dynamic volatility models, namely the use of range-based estimators, and describes two benchmarks, one from the stochastic volatility family and one from the GARCH family.

[5] On stochastic volatility, see Ghysels, Harvey, and Renault (1996), Shephard (2004) and Andersen, Bollerslev, Christoffersen, and Diebold (2006).

2.1 Dynamic Volatility and the Daily Price Range

Alizadeh, Brandt, and Diebold (ABD, 2002) demonstrate that the daily (log-) price range, defined as

$$r_t = \log(h_t - l_t) = \log(\log H_t - \log L_t)\,, \tag{2.1}$$

where h_t and l_t respectively are the logarithm of the highest (H_t) and lowest (L_t) price taken by an asset on day t, is a highly efficient, approximately-normally-distributed proxy of volatility. Consequently, they advocate its use in the estimation of a standard stochastic volatility model. Building on ABD's results, Brandt and Jones (2006) compare the use of absolute or squared returns to the range-based estimation of a variety of GARCH models. The models estimated using the range dominate their return-based counterparts both in terms of in-sample fit and, most importantly, in terms of their ability to precisely forecast volatility out of sample.

The theory behind these results is as follows. Consider a driftless Brownian motion x with constant diffusion coefficient σ and let $r_\tau \equiv \log\left(\sup_{0<t\leq\tau} x(t) - \inf_{0<t\leq\tau} x(t)\right)$ be the range over a finite interval of length τ.[6,7] Assuming that $x(0) = 0$, the density of r_τ is given by

$$P(r_\tau|\sigma) = 8\sum_{k=1}^{\infty}(-1)^{k-1}\frac{k^2 e^{r_\tau}}{\sigma\sqrt{\tau}}\phi\Big(\frac{k e^{r_\tau}}{\sigma\sqrt{\tau}}\Big),$$

where $\phi(\cdot)$ is the standard normal density (Alizadeh, Brandt, and Diebold 2002; Feller 1951). Under the above process, it can be shown that the first four moments of the range over the time interval τ are given by

$$\begin{aligned} \mathbb{E}[r_\tau] &= 0.43 + \tfrac{1}{2}\log\tau + \log\sigma\,, & \mathrm{Skew}(r_\tau) &= 0.17\,, \\ \mathrm{Var}(r_\tau) &= 0.084\,, & \mathrm{Kurt}(r_\tau) &= 2.80\,. \end{aligned} \tag{2.2}$$

Quite remarkably, the variance, skewness and kurtosis of the range distribution are independent of σ and interval length τ.[8] Moreover, both the skewness and kurtosis are very close to those of a normal distribution. Hence, given an observed range r_τ, one can use the quantity $r_\tau - \frac{1}{2}\log\tau - 0.43$ as an unbiased and approximately-normally-distributed estimator of $\log\sigma$.

[6] In the context of an analysis of daily returns, the assumption of a constant diffusion coefficient, σ, amounts to assuming that volatility is constant throughout a day.

[7] Throughout this paper, we use subscript t for discrete time indices and parentheses, (t), for continuous time.

[8] Note that, using a nonparametric estimator of volatility occupation times, Li, Todorov, and Tauchen (2012) document empirically that the interquartile range of log volatility is unrelated to the level of the volatility.

Moreover, the daily range is a much less noisy volatility proxy than absolute or squared returns (Parkinson 1980; Andersen and Bollerslev 1998), and is robust to market microstructure noise (ABD 2002). Given these features, the range proves valuable in the estimation of volatility models, as shown by ABD and Brandt and Jones (2006).

2.2 A Range-Based Stochastic Volatility Model

The building block of our analysis is the stochastic volatility model discussed in ABD.[9] The asset price S_t evolves according to a geometric Brownian motion with constant drift μ and stochastic volatility σ_t, the logarithm of the latter being modeled as a mean-reverting Ornstein–Uhlenbeck process,

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu\,\mathrm{d}t + \sigma(t)\,\mathrm{d}W_S(t) \tag{2.3}$$

$$\mathrm{d}\log\sigma(t) = \alpha\left(\log\bar{\sigma} - \log\sigma(t)\right)\mathrm{d}t + \beta\,\mathrm{d}W_\sigma(t), \tag{2.4}$$

where $W_S(t)$ and $W_\sigma(t)$ are two independent Wiener processes, $\log\bar{\sigma}$ is the mean volatility level, α is a constant governing the speed of mean-reversion and β is the constant volatility of log-volatility.

As financial data is observed at discrete intervals, any continuous-time stochastic volatility model must be discretized for estimation. Considering daily data, the Ornstein–Uhlenbeck log-volatility process of eq. (2.4) can be discretized as the following autoregressive process

$$\log\sigma_{t+1} = \log\bar{\sigma} + \rho_\tau(\log\sigma_t - \log\bar{\sigma}) + \beta\sqrt{\tau}\,\varepsilon^\sigma_{t+1}\,, \tag{2.5}$$

where t is a discrete time index such that $\log\sigma_t \equiv \log\sigma(t\tau)$, $\tau = 1/252$ is the fraction of a year taken by a trading day, $\rho_\tau \equiv 1 - \alpha\tau$, and the $\{\varepsilon^\sigma_t\}$ are uncorrelated standard normal variates.

The volatility process in (2.5) can be estimated based on the following observation equation

$$r_t = \log\sigma_t + \mathbb{E}\left[r_t - \log\sigma_t\right] + \sqrt{\mathrm{Var}\,(r_t)}\,\varepsilon^r_t\,, \tag{2.6}$$

where $\log\sigma_t$ is the latent state and ε^r_t is the observation noise. The bias and variance of r_t as a log-volatility proxy (i.e. the expectation and variance terms in (2.6)) are given analytically

[9] As pointed out by Brandt and Jones (2005), ABD's specification yields a discretized volatility model that is also studied in Duffie and Singleton (1993), Gallant, Hsieh, and Tauchen (1997), Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998), and Taylor (1994).

in (2.2). The observation equation (2.6) together with the transition equation (2.5) form a linear dynamical system for which the Kalman smoother (Kalman 1960; Hamilton 1994) provides a likelihood function that can be used to obtain maximum-likelihood estimates of $\log \bar{\sigma}, \beta$ and ρ_τ.[10] The Kalman smoother is also used to extract the latent log-volatilities in sample.

As shown in Appendix A.1, given a set of parameters and latent state estimates based on the information up to time t, this model implies a joint normal predictive distribution for observations beyond the last observed one, with mean vector and covariance matrix elements given by

$$\mathbb{E}_t\left[\log \sigma_{t+k}\right] = \log \bar{\sigma} + \rho_\tau^k \left(\mathbb{E}_t\left[\log \sigma_t\right] - \log \bar{\sigma}\right) \tag{2.7}$$

$$\mathrm{Cov}_t\left(\log \sigma_{t+k}, \log \sigma_{t+m}\right) = \rho_\tau^{k+m}\, \mathrm{Var}_t\left(\log \sigma_t\right) + \beta^2 \tau \rho_\tau^{m-k} \left(\frac{1-\rho_\tau^{2k}}{1-\rho_\tau^2}\right), \tag{2.8}$$

with $1 \leq k \leq m$, and where $\mathbb{E}\left[\log \sigma_t\right]$ and $\mathrm{Var}\left(\log \sigma_t\right)$ represent the latent log-volatility distribution at the last in-sample observation, as determined by the Kalman smoother.

2.2.1 ABD's Two-Component Model

Unsurprisingly, Alizadeh, Brandt, and Diebold (2002) find that the single-component stochastic volatility model above is severely misspecified. These results are consistent with the growing consensus that two-factor volatility processes better capture the time-series properties of volatility by separately accounting for transient and high-persistence volatility shocks.[11] To overcome the misspecification of their single-component model, ABD introduce the following two-component model, which we refer to as ABD(2),

$$\log \sigma_{t+1} = \log \bar{\sigma} + \log \sigma_{1,t+1} + \log \sigma_{2,t+1}, \tag{2.9}$$

$$\log \sigma_{1,t+1} = \rho_{1,\tau} \log \sigma_{1,t} + \beta_1 \sqrt{\tau}\, \varepsilon^\sigma_{1,t+1}, \tag{2.10}$$

$$\log \sigma_{2,t+1} = \rho_{2,\tau} \log \sigma_{1,t} + \beta_2 \sqrt{\tau}\, \varepsilon^\sigma_{2,t+1}, \tag{2.11}$$

where the volatility component innovations $\varepsilon^\sigma_{1,t+1}$ and $\varepsilon^\sigma_{2,t+1}$ are contemporaneously and serially independent standard normal variates. ABD find that this two-component model

[10] Given that ε^r_i is assumed to be a standard normal innovation while it might not actually be, we should properly resort here to quasi-maximum likelihood. Yet, as highlighted in (2.2), the log-range's distribution is very close to normal thus greatly improving the efficiency of the estimation. See Alizadeh, Brandt, and Diebold (2002) for details.

[11] On two-factor models, see, amongst others, Engle and Lee (1999); Andersen, Bollerslev, Diebold, and Ebens (2001a,2001b); Alizadeh, Brandt, and Diebold (2002); and Engle and Rangel (2008).

better describes the exchange-rate time series data than its single-component counterpart. Moreover, ABD(2) is not rejected by the specification test rejecting the single-component version of the model. Based on these results, we expect ABD(2) to be a much more stringent benchmark than the single-component model.

Note that, under ABD(2), the predictive distribution of log-volatility is still a joint normal distribution with moments, similar to those in equations (2.7) and (2.8), given in Appendix A.2.

2.3 A Range-Based GARCH Volatility Model

The vast majority of GARCH models are specified as

$$R_{t+1} := \log \frac{S_{t+1}}{S_t} = \mu_{t+1} + \sigma_{t+1}\varepsilon_{t+1} \,, \tag{2.12}$$

$$\sigma_{t+1} = f\left(\varepsilon_t^2;\, \sigma_s, s \leq t \right) \,, \tag{2.13}$$

where S is the asset price, and f is some parametric function of squared standardized log-returns ε^2 and of past filtered volatilities. Following ABD's work, Brandt and Jones (2006) developed versions of the EGARCH model building on the superior informativeness of the price range. They found all range-based models to dominate their return-based counterparts both in and out of sample.

One of these models, very similar to ABD's model, is the REGARCH model, a range-based version of Nelson's (1991) EGARCH model. The model specification is as follows,

$$R_{t+1} = \sigma_{t+1}\varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim \mathcal{N}(0,1) \,, \tag{2.14}$$

$$r_{t+1} \sim \mathcal{N}\left(0.43 + \log \sigma_{t+1}, 0.29^2\right) \,, \tag{2.15}$$

$$\log \sigma_{t+1} - \log \sigma_t = \kappa(\theta - \log \sigma_t) + \phi\, \frac{r_t - 0.43 - \log \sigma_t}{0.29} + \delta\, \varepsilon_t \,, \tag{2.16}$$

where R_t is the log asset return and r_t the range observed on day t; the normal distribution assumption for the range is based on the results of Section 2.1.[12] Estimating the model with maximum likelihood is straightforward.

As shown in Appendix A.3, given a set of parameters and filtered values of log-volatility based on the information up to time t, this model implies a joint normal predictive distribution for future log-volatility levels, with mean vector and covariance matrix elements

[12]Note that, here, $\log \sigma_t$ is measured on a daily measure, i.e. $\log \sigma_t = \log(\sigma_t^a/252)$ where $\log \sigma_t^a$ is an annualized measure of volatility. Hence, $r_t - 0.43 - \log \sigma_t = r_t - 0.43 - \log \sigma_t^a - \log \tau$, with $\tau = 1/252$, is consistent with equation (2.2).

given by

$$\mathbb{E}_t\left[\log\sigma_{t+k}-\theta\right]=\rho^{k-1}(\log\sigma_{t+1}-\theta)\,, \tag{2.17}$$

$$\mathrm{Cov}_t\left(\log\sigma_{t+k},\log\sigma_{t+m}\right)=\left(\phi^2+\delta^2\right)\rho^{m-k}\left(\frac{1-\rho^{2(k-1)}}{1-\rho^2}\right), \tag{2.18}$$

for $2\leq k\leq m$. Note that there is an important difference between these equations and equations (2.7) and (2.8) obtained for the stochastic volatility model. By definition of the model, the one-step ahead log-volatility level, $\log\sigma_{t+1}$, is considered to be known once we have observed R_t. Hence, the model assumes that there is no uncertainty around its value ($\mathrm{Var}_t\left(\log\sigma_{t+1}\right)=0$) and, consequently, a covariance of zero with any future log-volatility forecast.

2.3.1 Two-Component REGARCH Model

Like ABD, Brandt and Jones (2006) highlight that the single-component REGARCH model is severely misspecified and significantly outperformed by a two-component version of the model, REGARCH(2), defined as follows:

$$\log\sigma_{1,t+1}-\log\sigma_{2,t+1}=\rho_1(\log\sigma_{1,t}-\log\sigma_{2,t})+\phi_1\,\tilde{r}_t+\delta_1\,\varepsilon_t\,, \tag{2.19}$$

$$\log\sigma_{2,t+1}-\theta=\rho_2(\log\sigma_{2,t}-\theta)+\phi_2\,\tilde{r}_t+\delta_2\,\varepsilon_t\,, \tag{2.20}$$

$$\text{where}\quad\tilde{r}_t=\left(r_t-0.43-\log\sigma_t\right)/0.29\,. \tag{2.21}$$

As shown in Appendix A.4, given a set of parameters and filtered values of log-volatility based on the information up to time t, this model also implies a joint normal predictive distribution for future log-volatility levels, with mean and covariance much similar to those of the single-component model (equations (2.17) and (2.18)).

3 Gaussian Processes and Stochastic Volatility

We now turn to a brief review of Gaussian process regression (GPR) and then introduce the volatility forecasting methodology being put forward in this paper.

3.1 Basic Concepts for the Regression Case

A Gaussian process is a generalization of the Gaussian distribution: it represents a probability distribution over functions which is entirely specified by a mean and covariance functions. Borrowing the definition from Rasmussen and Williams (2006), we formally define a Gaussian process (GP) as follows:

Definition A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.

Let $\mathbf{x} \in \mathbb{R}^D$ index into the real process $f(\mathbf{x})$.[13] We write

$$f(\mathbf{x}) \sim \mathcal{GP}\left(M(\cdot), K(\cdot, \cdot)\right), \tag{3.1}$$

where functions $M(\cdot)$ and $K(\cdot, \cdot)$ are, respectively, the mean and covariance functions:

$$\begin{aligned} M(\mathbf{x}) &= \mathbb{E}\left[f(\mathbf{x})\right], \\ K(\mathbf{x}_1, \mathbf{x}_2) &= \mathbb{E}\left[\left(f(\mathbf{x}_1) - M(\mathbf{x}_1)\right)\left(f(\mathbf{x}_2) - M(\mathbf{x}_2)\right)\right]. \end{aligned}$$

In parallel to the normal distribution, which is fully specified by a mean vector and covariance matrix, a Gaussian process specification only requires a mean and covariance *functions*. The latter is pivotal in characterizing the behavior of functions induced by the process, and we discuss its importance for volatility modeling in the next section.

In a regression setting, we assume that we are given an estimation sample $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$, with $\mathbf{x}_i \in \mathbb{R}^D$ a vector of covariates and $y_i \in \mathbb{R}$ an associated desired response. For convenience, let $\mathbf{X}$ be the matrix of all predictor variables in the estimation sample, and $\mathbf{y}$ the vector of desired responses. We further assume that the y_i are noisy measurements from the underlying unobserved process $f(\mathbf{x})$:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, \qquad \text{with } \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_n^2). \tag{3.2}$$

As outlined in eq. (2.2), volatility measurements obtained from the log-range are nearly normally distributed around the true log-volatility, and hence eq. (3.2) can be assumed to hold, with $f(\cdot)$ representing the underlying "true" log-volatility process, and y_i derived from log-range observations, and $\mathbf{x}$ (for the purposes of the present discussion) a scalar time index.

Regression with a GP is achieved by means of Bayesian inference in order to obtain a posterior distribution over functions given a suitable prior and an observed data sample. Then, given new test observations, we can use the posterior to arrive at a predictive distribution conditional on the test observations and the estimation sample: from a GP prior and Gaussian likelihood, we obtain a posterior that is a GP; marginalizing this posterior at a finite number of test points yields a jointly normal predictive distribution over those points.

[13]Contrarily to many treatments of stochastic processes, there is no necessity for $\mathbf{x}$ to represent time; indeed, in our application of Gaussian processes, some of the elements of $\mathbf{x}$ have a temporal interpretation, but most are general macroeconomic variables used to condition the response.

Although the idea of manipulating distributions over functions may appear cumbersome, the consistency property of GPs means that any finite number of values sampled from the process f are jointly normal; hence inference over random functions can be shown to be completely equivalent to inference over a finite number of random variables.

It is often convenient, for simplicity, to assume that the GP prior distribution has a mean of zero,[14]

$$f(\mathbf{x}) \sim \mathcal{GP}(0, K(\cdot,\cdot)).$$

Let $\mathbf{f}' = [f(\mathbf{x}_1), \ldots, f(\mathbf{x})_N]$ be the vector of (latent) function values evaluated each point of the estimation sample. Their joint prior distribution is given by

$$\mathbf{f} \sim \mathcal{N}(\mathbf{0}, K(\mathbf{X}, \mathbf{X})),$$

where $K(\mathbf{X},\mathbf{X})_{i,j} \equiv K(\mathbf{x}_i, \mathbf{x}_j)$ is a matrix formed by evaluating the covariance function between all pairs $\mathbf{x}_i, \mathbf{x}_j$. This matrix is also known as the *kernel* or *Gram* matrix. We consider the joint prior distribution between the estimation sample and an additional set of test observations, denoted by the matrix $\mathbf{X}_*$, and whose unknown function values $\mathbf{f}_*$ we wish to infer. Under the GP, the joint prior distribution $P(\mathbf{f}, \mathbf{f}_*)$ between $\mathbf{f}$ and $\mathbf{f}_*$ is[15]

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X},\mathbf{X}) & K(\mathbf{X},\mathbf{X}_*) \\ K(\mathbf{X}_*,\mathbf{X}) & K(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix}\right). \tag{3.3}$$

We obtain the predictive distribution at the test points as follows. From Bayes' theorem, the joint posterior is obtained as the product of the above prior and data likelihood

$$P(\mathbf{f}, \mathbf{f}_* \,|\, \mathbf{y}) = \frac{P(\mathbf{y} \,|\, \mathbf{f}, \mathbf{f}_*) P(\mathbf{f}, \mathbf{f}_*)}{P(\mathbf{y})} = \frac{P(\mathbf{y} \,|\, \mathbf{f}) P(\mathbf{f}, \mathbf{f}_*)}{P(\mathbf{y})}, \tag{3.4}$$

where $P(\mathbf{y} \,|\, \mathbf{f}, \mathbf{f}_*) = P(\mathbf{y} \,|\, \mathbf{f})$ since, by assumption, the likelihood is conditionally independent of $\mathbf{f}_*$ given $\mathbf{f}$ and, from eq. (3.2),

$$\mathbf{y} \,|\, \mathbf{f} \sim \mathcal{N}(\mathbf{f}, \sigma_n^2 I_N) \tag{3.5}$$

with I_N the $N \times N$ identity matrix. The desired predictive distribution is obtained by marginalizing out the latent variables corresponding to the estimation sample,

$$P(\mathbf{f}_* \,|\, \mathbf{y}) = \int P(\mathbf{f}, \mathbf{f}_* \,|\, \mathbf{y}) \, \mathrm{d}\mathbf{f} = \frac{1}{P(\mathbf{y})} \int P(\mathbf{y} \,|\, \mathbf{f}) P(\mathbf{f}, \mathbf{f}_*) \, \mathrm{d}\mathbf{f}.$$

[14]This assumption has no bearing on the ability of the GP to represent distinctly non-zero posterior means, as will be obvious below.

[15]Note that the following expressions in this section are implicitly conditioned on $\mathbf{X}$ and $\mathbf{X}_*$. The explicit conditioning notation is omitted for brevity.

Since these distributions are all normal, the result of the (normalized) marginal is also normal, and can be shown to have mean vector and covariance matrix respectively given by

$$\mathbb{E}\left[\mathbf{f}_* \,|\, \mathbf{y}\right] = K(\mathbf{X}_*, \mathbf{X})\mathbf{\Lambda}^{-1}\mathbf{y}, \tag{3.6}$$

$$\mathrm{Cov}\left[\mathbf{f}_* \,|\, \mathbf{y}\right] = K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}_*, \mathbf{X})\mathbf{\Lambda}^{-1}K(\mathbf{X}, \mathbf{X}_*), \tag{3.7}$$

with

$$\mathbf{\Lambda} = K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I_N. \tag{3.8}$$

We see from these expressions that the conditional expectation of unknown function values $\mathbf{f}_*$ is linear in $\mathbf{y}$, but nonlinear in $\mathbf{X}$. Computation of $\mathbf{\Lambda}^{-1}$ (or solving the associated linear system) is the most computationally expensive step in Gaussian process regression, requiring $O(N^3)$ time and $O(N^2)$ space.[16]

Note that a natural outcome of GP regression is an expression for not only the expected value at the test points (3.6), but also the full covariance matrix between those points (3.7). This covariance matrix is the analogue of equation (2.8) in the stochastic volatility model, and of equation (2.18) in the REGARCH case.

3.2 Modeling Stochastic Volatility with Gaussian Processes

The first step in modeling stochastic volatility with a GPR model is to cast it in terms of a (highly nonlinear) regression from time indices to volatility measurements obtained from the log-range. Put differently, our estimation sample is made up of pairs $\mathcal{D} = \{(t_i, y_i)\}$, where t_i is a numerical time index and y_i is obtained from the observed log price range at that time, as shown next. It is significant that the formulation introduced here allows for arbitrary time indices; they do not have to be equally spaced in time.

Rearranging ABD's model observation equation (2.6) yields

$$r_t = \log\sigma_t + \mathbb{E}\left[r_t - \log\sigma_t\right] + \sqrt{\mathrm{Var}\,(r_t)}\,\varepsilon_t^r\,, \qquad \varepsilon_t^r \sim \mathcal{N}(0,1) \tag{3.9}$$

$$\iff \qquad r_t - \log\tau - 0.43 = \log\sigma_t + 0.29\,\varepsilon_t^r\,, \tag{3.10}$$

$$\iff \qquad y_t = f(t) + \sigma_n^2\,\varepsilon_t^r \tag{3.11}$$

[16] A substantial literature exists to study so-called sparse approximations for Gaussian process models, which aim to substantially reduce the computational load at the expense of some precision; although we do not make use of such approximations in this paper, they could be used to scale up the techniques we introduce to much larger data sets; for good overviews, see Quiñonero-Candela and Rasmussen (2005) and Snelson (2007).

where the third line, analogous to (3.2), arises by setting $y_t = r_t - \log\tau - 0.43$ and $f(t) = \log\sigma_t$. For conciseness, we denote by $\mathbf{t}$ the vector of all time indices, $\mathbf{t} = [1, 2, \ldots, T]'$, and by $\mathbf{y}$ the desired responses,

$$\mathbf{y} = \begin{bmatrix} r_1 - \log\tau - 0.43 \\ r_2 - \log\tau - 0.43 \\ \vdots \\ r_T - \log\tau - 0.43 \end{bmatrix}. \tag{3.12}$$

Using only the time indices $\mathbf{t}$ as covariates yields a GPR formulation of a standard stochastic volatility model, which we call the basic GPSV model. We later return to the extension of SV models obtained by adding more explanatory variables to the mix.

Choice of Covariance Function

In GPR modeling, the GP covariance function characterizes the functional prior distribution. A proper choice for the covariance function is determinant for encoding prior knowledge about the problem.[17] In order to yield valid covariance matrices, the covariance function should, at the very least, be symmetric and positive semi-definite, which implies that all its eigenvalues are nonnegative,

$$\int K(\mathbf{u}, \mathbf{v}) f(\mathbf{u}) f(\mathbf{v})\, \mathrm{d}\mu(\mathbf{u})\, \mathrm{d}\mu(\mathbf{v}) \geq 0,$$

for all functions f defined on the appropriate space (e.g. $\mathbb{R}^D$) and measure μ.

The covariance function corresponding to the Ornstein–Uhlenbeck (OU) process is a member of a family known as the Matérn class (Matérn 1960; Stein 1999). As derived in Appendix B, under the OU process, the prior covariance between $\log\sigma_{t_1}$ and $\log\sigma_{t_2}$ can be written as

$$K_{\mathrm{OU}}(t_1, t_2) = \eta_f \, \exp\left(-\frac{|t_1 - t_2|}{\eta_\ell}\right). \tag{3.13}$$

The constants η_f and η_ℓ are known as *hyperparameters* (which, in Bayesian statistics, refer to parameters of a prior distribution): the first one governs the overall signal variance, and the second the characteristic length-scale of variation with respect the time interval $t_1 - t_2$. We discuss in the next section how to choose these parameters. The important feature of this covariance function is its stationarity: the covariance only depends on the difference between t_1 and t_2, and not on absolute times.

[17] Several splendid examples are given in Rasmussen and Williams (2006).

Another very common choice of covariance function is the squared exponential (also known as the Gaussian or radial basis function kernel),

$$K_{\mathrm{SE}}(\mathbf{u}, \mathbf{v}; \eta_\ell) = \eta_f \exp\left(-\frac{\|\mathbf{u} - \mathbf{v}\|^2}{2\eta_\ell^2}\right). \tag{3.14}$$

The hyperparameters have the same meaning as previously.

These two covariance functions yield extremely different functional priors. Figure 1 illustrates several functions drawn from a GP, respectively under the squared exponential and the OU covariances in one dimension. The same random numbers were used for generating both sets, so the functions plotted with the same color are "alike" in a sense. We note that a function under the squared exponential prior is much smoother than the corresponding function under the OU prior; for the purpose modeling asset price volatility, the squared exponential covariance function constitutes an ill-suited prior since it would yield posteriors that are too smooth. For the time dimension, the GPSV model relies on the K_{OU} covariance function due to its relationship with standard stochastic volatility.

Moreover, we note that the sum of two covariance functions yields in turn a valid covariance function. This allows much flexibility in specifying functional priors that are matched to domain knowledge, for instance using an OU prior for the time dimension as just mentioned, along with the much smoother squared exponential covariance for incorporating other volatility covariates that can be expected to have a softer impact. We make use of this additivity property, sometimes called "covariance function engineering", in section 5.2.

Forecasting with the GPR Model

Forecasting is carried out by querying the model at time indices that extend beyond the end of estimation sample. For example, suppose that the model was estimated using data until time T. To forecast volatility one year into the future, one simply set $\mathbf{t}^* = [T+1, \ldots, T+252]'$, and computes $\mathbb{E}[\mathbf{f}_* \,|\, \mathbf{y}]$ and $\mathrm{Cov}[\mathbf{f}_* \,|\, \mathbf{y}]$ as per equations (3.6) and (3.7) to obtain the conditional volatility forecasts. This is analogous to equations (2.7) and (2.8) in the ABD model.

A Note on Missing Values

An interesting property of the GP framework is that it accommodates missing values straightforwardly. Indeed, vector $\mathbf{t}$ is in no manner required to contain consecutive time indices. Hence, if observation y_s is missing for some date s, time s simply does not appear in $\mathbf{t}$. This is extremely convenient to deal with data sampled at different frequencies.

3.2.1 Two-Component GPSV Model

Since the sum of two covariance functions is a covariance function, we consider a two-component variant as the sum of two K_{OU} covariances, in the same spirit as the two-component versions of ABD and REGARCH:

$$K_{\mathrm{OU2C}}(t_1, t_2) = K_{1,\mathrm{OU}}(t_1, t_2) + K_{2,\mathrm{OU}}(t_1, t_2)\,, \tag{3.15}$$

$$K_{1,\mathrm{OU}}(t_1, t_2) = \eta_{1,f}\, \exp\left(-\frac{|t_1 - t_2|}{\eta_{1,\ell}}\right), \tag{3.16}$$

$$K_{2,\mathrm{OU}}(t_1, t_2) = \eta_{2,f}\, \exp\left(-\frac{|t_1 - t_2|}{\eta_{2,\ell}}\right), \tag{3.17}$$

with $\eta_{1,\ell} > \eta_{2,\ell}$ to ensure that one component captures long-range volatility effects and the other short-range ones.

3.3 Optimization of Hyperparameters

Most covariance functions, including those presented above, have free parameters — termed *hyperparameters* — that control their shape, for instance the characteristic length-scale η_ℓ. It remains to explain how their value should be set.[18]

An approach that is quite efficient for GPR models consists in maximizing the *marginal likelihood*[19] of the observed data with respect to the hyperparameters. This function can be computed by introducing latent function values that are immediately marginalized over. Let θ the set of hyperparameters that are to be optimized; and let $K_{\mathbf{X}}(\theta)$ the covariance matrix computed by a covariance function whose hyperparameters are θ,

$$K_{\mathbf{X}}(\theta)_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j; \theta).$$

The marginal likelihood can be written

$$p(\mathbf{y}\,|\,\mathbf{X}, \theta) = \int p(\mathbf{y}\,|\,\mathbf{f}, \mathbf{X})p(\mathbf{f}\,|\,\mathbf{X}, \theta)\,\mathrm{d}\mathbf{f}, \tag{3.18}$$

[18] This is an instance of the general problem of model selection. For many machine learning algorithms, this problem has often been approached by minimizing a validation error through cross-validation (Stone 1974) and such methods have been proposed for Gaussian processes (Sundararajan and Keerthi 2001).

[19] Also called the *integrated likelihood* or (Bayesian) *evidence*.

where the distribution of observations $p(\mathbf{y}\,|\,\mathbf{f},\mathbf{X})$, given by eq. (3.5), is conditionally independent of the hyperparameters given the latent function $\mathbf{f}$. Under the Gaussian process prior (see eq. (3.3)), we have $\mathbf{f}\,|\,\mathbf{X},\theta \sim \mathcal{N}(0, K_{\mathbf{X}}(\theta))$, or in terms of log-likelihood,

$$\log p(\mathbf{f}\,|\,\mathbf{X},\theta) = -\frac{1}{2}\mathbf{f}' K_{\mathbf{X}}^{-1}(\theta)\mathbf{f} - \frac{1}{2}\log|K_{\mathbf{X}}(\theta)| - \frac{N}{2}\log 2\pi. \tag{3.19}$$

Since both distributions in eq. (3.18) are normal, the marginalization can be carried out analytically to yield

$$\log p(\mathbf{y}\,|\,\mathbf{X},\theta) = -\frac{1}{2}\mathbf{y}'(K_{\mathbf{X}}(\theta)+\sigma_n^2 I_N)^{-1}\mathbf{y} - \frac{1}{2}\log\left|K_{\mathbf{X}}(\theta)+\sigma_n^2 I_N\right| - \frac{N}{2}\log 2\pi\,, \tag{3.20}$$

where N is the number of in-sample observations and $|\cdot|$ denotes the matrix determinant.

This expression can be maximized numerically to yield the selected hyperparameters:

$$\theta^* = \arg\max_{\theta} \log p(\mathbf{y}\,|\,\mathbf{X},\theta).$$

The likelihood function is in general non-convex and this maximization only finds a local maximum in parameter space; however, empirically it usually works very well for a large class of covariance functions, including those covered in Section 3.2.

It should be mentioned that completely a Bayesian treatment would not be satisfied with such an optimization, since it only picks a single value for each hyperparameter. Instead, one would define a prior distribution on hyperparameters (a so-called *hyperprior*), $p(\theta)$, and marginalize with respect to this distribution. However, for many "reasonable" hyperprior distributions, this is computationally expensive, and often yields no marked improvement over simple optimization of the marginal likelihood (MacKay 1999).

The gradient of the marginal log-likelihood with respect to the hyperparameters — necessary for numerical optimization algorithms — can be expressed as

$$\frac{\partial \log p(\mathbf{y}\,|\,\mathbf{X},\theta)}{\partial\theta_i} = \frac{1}{2}\mathbf{y}' K_{\mathbf{X}}^{-1}(\theta)\frac{\partial K_{\mathbf{X}}(\theta)}{\partial\theta_i}K_{\mathbf{X}}^{-1}(\theta)\mathbf{y} - \frac{1}{2}\mathrm{Tr}\left(K_{\mathbf{X}}^{-1}(\theta)\frac{\partial K_{\mathbf{X}}(\theta)}{\partial\theta_i}\right). \tag{3.21}$$

See Rasmussen and Williams (2006) for details on the derivation of this equation.[20]

4 Simulation Analysis

Our objective is now to demonstrate the empirical performance gains obtained by casting stochastic volatility models in a *Gaussian process regression* (GPR) framework. Before considering actual market data, we here introduce some performance measures of interest and

[20]The software implementation of GPR models and hyperparameter optimization used for all results in this paper is part of the excellent GPML open-source Matlab toolbox (Rasmussen and Nickisch 2010).

consider a simulation analysis in the framework of Alizadeh, Brandt, and Diebold (2002). Working in a simulation context, we know the underlying stochastic volatility process and can thus straightforwardly evaluate the effectiveness of the *GPR-based stochastic volatility* (GPSV) model in (i) recovering this process, given only daily price highs and lows, and (ii) forecasting it out of sample.

We first sample daily volatilities from eq. (2.5). Then, in order to obtain daily log-range values, we sample the log-price process s intraday using a driftless discretization of (2.3) , i.e. on day t

$$s(t_0 + n\tilde{\tau}) = s(t_0 + (n-1)\tilde{\tau}) + \sigma_t \sqrt{\tilde{\tau}}\, \varepsilon_n^s\,, \qquad n = 1, \ldots, M\,, \tag{4.1}$$

where $t_0 = (t-1)\tau$ is the (continuous) time at the beginning of day t, $\{\varepsilon_n^s\}$ are intraday standard normal innovations (independent of $\{\varepsilon_{i\tau}^{\sigma}\}$), and $\tilde{\tau} = \tau/M$. This yields M log-prices per day, whose maximum and minimum values are used to compute r_t. As parameters for the log-volatility process, we took $\alpha = 3.855$, $\log\bar{\sigma} = -2.5$ and $\beta = 0.75$, yielding realistic volatility dynamics, largely consistent, for instance, with those observed on foreign exchange markets (Alizadeh, Brandt, and Diebold 2002). We used $M = 5184$ intraday prices.[21]

4.1 Recovering Underlying Volatility

An illustration of how well the models are capable of recovering an underlying known log-volatility process appears in Figure 2. Both the ABD and GP models exhibit nearly identical solutions, both on the inference of the underlying in-sample volatility and over the forecasting period. It should be emphasized that neither model has direct access to the underlying volatility process: on each day, they only "observe" the high and low prices.

Largely the same remark applies for the REGARCH model. The volatility filtered by this model is not as smooth as the latent states extracted by the ABD and GP models. This is a natural consequence of the specification of all GARCH models, which implements filtering rather than smoothing of estimated volatilities over the sample period. Since the one-step-ahead volatility, σ_{t+1}, is considered to be known at time t, the filtered state is not revised using the information arriving after time t, even though this information is part

[21] Per five-minute period, this yields 18 prices observations. See Alizadeh, Brandt, and Diebold (2002) for a discussion on the impact of M.

of the estimation sample and thus available to the econometrician.[22] In Figure 2, we see that the filtered nature of GARCH volatility yields a systematic lag with respect to the true underlying log-volatility.

Since the models cannot represent additional structure, the out-of-sample log-volatility trajectory has the shape of an exponential reversion to the long-run mean log-volatility. This limitation is addressed in Section 5, where we extend the covariance function of the GPSV model to account for seasonality and macro-finance variables, enabling the model to represent richer out-of-sample dynamics.

4.2 Performance Measures

We now introduce statistical measures allowing a comparison of the performance of the different models in and out of sample, over 10,000 Monte Carlo trajectories. All performance measures introduced next are computed in the log-volatility domain. A measure of forecasting up to horizon H includes the *entire subtrajectory* for $h = 1, \dots, H$. In other words, we are interested in modeling the complete future volatility trajectory rather than a point estimate at some future horizon.

The mean-squared error (MSE) is computed between the true log-volatility (which is known from the simulated trajectories) and the model forecast. Let t be the time at which the forecast is made. Let $\log \sigma^j_{t+h}$ the true underlying process log-volatility for the j-th Monte Carlo trajectory, and $\log \hat{\sigma}^j_{t+h|t}$ the corresponding conditional expectation given by some model. The MSE at horizon H is defined as

$$\text{MSE}^j_H = \frac{1}{H}(\Delta^j_{H|t})'\Delta^j_{H|t}, \tag{4.2}$$

where $\Delta^j_{H|t}$ is a vector whose h-th element is given by $\left(\Delta^j_{H|t}\right)_h = \log \hat{\sigma}^j_{t+h|t} - \log \sigma^j_{t+h}$. The overall reported MSE at horizon H is the arithmetic average of all MSE^j_H across Monte Carlo trajectories.

Whereas the MSE captures the ability of a model to correctly represent the conditional expectation, the *negative log-likelihood* (NLL) captures the ability of the model to assign greater probability density to likely events. All models considered in this paper (ABD, REGARCH, GPSV) give rise to a normal joint predictive distribution of future log volatility

[22] In this context, the conditional variance of σ_{t+1} at time t is zero; in Figure 2, the uncertainty on the filtered log-volatility at time t is based on its standard deviation conditional on the information available at time $t-2$ as obtained using equation (2.18).

(see Appendix A). The NLL at horizon H for a trajectory j is

$$\text{NLL}_H^j = \frac{1}{2}\left[(\Delta_{H|t}^j)'(\Sigma_{H|t}^j)^{-1}\Delta_{H|t}^j + \log\left|\Sigma_{H|t}^j\right| + H\log 2\pi\right], \tag{4.3}$$

where $\Delta_{H|t}^j$ is as previously, $\Sigma_{H|t}^j$ is the covariance matrix of the model forecasts up to horizon H.

Finally, the *absolute bias* evaluates whether forecasts are more likely to be wrong in one direction than another. The bias at horizon H for a trajectory j is

$$\text{Bias}_H^j = \frac{1}{H}\sum_{h=1}^{H} \log\hat{\sigma}_{t+h|t}^j - \log\sigma_{t+h}^j. \tag{4.4}$$

4.3 Performance Comparison on Data Simulated with ABD's SV Models

Table 1 shows detailed performance results, averaged on 10,000 Monte Carlo trajectories, comparing the data-generating ABD model, and the REGARCH model to the GPSV model using the K_{OU} covariance function.

We simulate 10,000 6-year-long log-volatility trajectories using equation (2.5). We estimate the (data-generating) ABD, REGARCH and GPSV models using the first 5 years of each trajectory. Each model is then used to forecast log-volatility from one day to one year ahead (horizon H, in business days). For each forecast, the NLL, MSE and Bias are computed—equations (4.2) to (4.4)—and their average (and standard error) over the 10,000 trajectories is reported in the three left-hand side columns of Table 1.

Paired t-tests are performed comparing the NLL, MSE and magnitude of the bias, |Bias|, trajectory per trajectory. In-sample, the ABD and GPSV perform comparably. As for the REGARCH, the aforementioned lag between the filtered log-volatility and the underlying process (Figure 2) leads to poor in-sample performance.

In terms of out-of-sample NLL, the ABD model dramatically outperforms both the REGARCH and the GPSV model, a fact that is not surprising given that its parametric form corresponds exactly to the data generating process. More surprising, however, is the complete reversal in how models compare in terms of out-of-sample forecasting MSE. The REGARCH, at most horizons, and the GPSV model, at all horizons, significantly outperform the ABD model in terms of forecasting MSE. Given that ABD is the data-generating model, this inferior performance with respect to MSE can only be attributed to stability issues in parameter estimation becoming significant in out-of-sample forecasting, especially at long horizons.

Given that the results of Alizadeh, Brandt, and Diebold (2002) and Brandt and Jones (2006) clearly highlight the misspecification of the single-component ABD and REGARCH model, we repeat the above Monte Carlo experiment using ABD(2) as the data generating model. Table 2 reports the results. Once again, the data-generating ABD(2) model outperforms alternate models both from the in-sample MSE and the out-of-sample NLL perspective. And once more, it is significantly outperformed by GPSV(2) model in terms of out-of-sample MSE. Hence, placing volatility modeling and forecasting in a nonparametric Bayesian regression framework seems to alleviate the estimation stability difficulty encountered in the classical framework.

5 Empirical Analysis

Obviously, the true test of a forecasting methodology is whether it can capture and exploit empirical regularities found in actual market data. We now turn to forecasting the volatility of the S&P 500 index. Contrarily to the previous simulation results, a practical hurdle for comparing alternative models is that the "true" volatility of the index is unobserved; hence, it appears that there would be no ground truth forming a basis of comparison. To sidestep this issue, we use targets based on *realized volatility*, computed from high-frequency data, which have been recognized over the past decade as providing a statistically consistent estimate of the underlying volatility.[23]

Our raw data comes from the TickData™ database and contains all futures transactions on the E-mini S&P 500 between 1997 to 2011 (nearly 431M trades). At any given time, we use data from the contract closest to maturity, and roll over to the next contract roughly two weeks prior to maturity. The data is processed to extract prices at five-minute intervals, from which intraday returns are computed.[24] Realized volatility over a day is then computed as the sum of the five-minutes returns observed during the day. The annualized realized volatility on day t is computed as

$$\mathrm{RV}_t = 252 \sum_{i=1}^{M} r_{t,i}^2, \tag{5.1}$$

where $r_{t,i}$ is the i-th logarithmic return during day t, and M is the number of returns during the day, and 252 is the number of trading days during a year.[25]

[23]References on RV abound, starting with Andersen, Bollerslev, Diebold, and Ebens (2001a, 2003) and Barndorff-Nielsen and Shephard (2002). The survey in Andersen, Bollerslev, and Diebold (2009) provide an extensive review of the literature.

[24]See Aït-Sahalia (2005), and Bandi and Russell (2008) for discussions on the sampling frequency.

[25]While RV_t is an unbiased estimator of σ_t^2, $\log\left(\sqrt{\mathrm{RV}_t}\right)$ is not an unbiased estimator of $\log\sigma_t$. Indeed, $\log\left(\sqrt{\mathrm{RV}_t}\right) = 0.5\log\left(\sigma_t^2 \sum_{i=1}^{M} \epsilon_i^2\right) = \log\sigma_t + 0.5\left[\log(1/M) + \log\left(\sum_{i=1}^{M} \epsilon_i^2\right)\right]$. We correct for this bias.

It must be emphasized that realized volatility is used in this work only to *evaluate performance*; it is not required to estimate the models. High-frequency data of the kind needed to compute realized volatility does not constitute a volatility-modeling panacea: it remains available only for a few highly-developed markets, does not have a long history, and requires a substantial data management infrastructure.

5.1 Results without Explanatory Variables

The same performance metrics as those of Section 4 are employed. In the present case, the target is the log-volatility computed based on (5.1), with out-of-sample data extending from September 1997 to March 2011. The models are first estimated using five years of daily S&P 500 data from September 16, 1992, to September 10, 1997. The models' forecasting performance is assessed on a 252-day out-of-sample horizon extending beyond the end of the estimation sample. Then, the estimation sample is rolled one week forward, models are re-estimated, and new out-of-sample performance measures are computed. Table 3 reports the average of these measures for the 653 (weekly) iterations of this procedure.

Table 3 also reports paired t-tests comparing all models to the GPSV(2). The t-stats of the paired tests are computed using Newey and West (1987) standard errors, in order to account for the autocorrelation introduced by the overlap of the out-of-sample periods.The average performance of the models, along the three criteria, is almost always worse than that of our GPSV(2) model. The performance gain obtained using the GPSV(2) model is almost always statistically significant at long horizons, especially along the NLL criterion. The five-day- to one-year-ahead forecasting NLL of the GPSV(2) is systematically significantly better than that of competing models. Recall that, whereas the MSE captures the ability of a model to correctly represent the conditional expectation, the NLL captures the ability of the model to correctly represent the conditional *distribution* of future volatility.[26] Our volatility model is thus able to better capture volatility expectations, but more importantly the uncertainty about future volatility. This could prove especially useful in value-at-risk applications or in the valuation of volatility-sensitive claims.

Note that each two-component model clearly outperforms its single-component counterpart. This confirms the results of Alizadeh, Brandt, and Diebold (2002) and Brandt and

[26]Considering out-of-sample NLL performance, both versions of the REGARCH model perform surprisingly poorly. While the REGARCH properly forecast trends in volatility (as per the MSE criterion), it improperly describes the covariance structure of potential future volatility paths. As discussed along with Figure 2, the REGARCH assumption that volatility is measurable leads to levels of variance of the log-volatility process that are too low to properly describe the data.

Jones (2006), and adds up to the overwhelming evidence in favor of models with two components (Engle and Lee 1999; Andersen, Bollerslev, Diebold, and Ebens 2001a and 2001b; Alizadeh, Brandt, and Diebold 2002; and Engle and Rangel 2008). The better performance of the two-component models is clear along the three criteria, at all horizons from five days and beyond.

It is worth noting that all models severely underestimate volatility at long horizons. Pástor and Stambaugh (2012) argue that while mean reversion contributes strongly to reducing long-horizon variance, it is more than offset by various uncertainties faced by the investor. Our bias results suggest that the volatility models considered could be improved, in terms of long-run forecasting performance, by accounting for the uncertainties faced by investors; this could be done by incorporating macro-finance variables, an issue to which we now turn.

5.2 Adding Explanatory Variables

By treating volatility as a regression problem, the Gaussian process regression (GPR) framework allows the straightforward incorporation of a number of explanatory variables, a task difficult with the Kalman filter used by traditional stochastic volatility models (including our benchmark ABD model). These variables are important because they allow for exogenous factors driving volatility, including seasonalities and macroeconomic conditions.

One of the most attractive features of the GPR framework is that it provides the ability to incorporate domain knowledge in a principled manner via the specification of a problem-specific covariance function. We augment the OU covariance (3.13) of the basic GPSV model to accommodate day-of-the-week (DOW) effects and macro-finance variables $\mathbf{z}$, assuming as before that t is a date index, to obtain the augmented covariance

$$K_{\mathrm{D,Y}}\Big((t,\mathbf{z}),(t',\mathbf{z}')\Big) = \eta_t^2 K_{\mathrm{OU2C}}(t,t') + \eta_{\mathrm{DOW}}^2 \delta(\mathrm{DOW}(t),\mathrm{DOW}(t')) + \eta_{\mathbf{z}}^2 K_{\mathrm{SE}}(\mathbf{z},\mathbf{z}'). \tag{5.2}$$

The first term is the standard two-component OU covariance for temporal dependencies. In the second term, the function $\mathrm{DOW}(t)$ returns the day of the week on which date t falls, and $\delta(i,j)$ is a Kronecker delta taking value one iff $i = j$ and zero otherwise. The purpose of this term is to potentially introduce additional covariance (thence a prior for increased similarity, depending on the magnitude of η_{DOW}) between the observations at dates t and t' if they fall on the same day of the week. The third term incorporates macro-finance variables via a smoother similarity measure induced through a squared exponential prior. All hyperparameters $\{\eta_t, \eta_{\mathrm{DOW}}, \eta_{\mathbf{z}}\}$ are estimated using the optimization technique outlined in section 3.3.

The macro-finance variables $\mathbf{z}$ used in this paper are:

- BY_t: U.S. government bond yield;
- DY_t: Long-run (10-year smoothed) U.S. equities earnings yield;
- EY_t: Long-run (10-year smoothed) U.S. equities dividend yield.

In the above notation, the vector of observables is thus $\mathbf{z} = (BY_t, DY_t, EY_t)'$. The data is sourced from Robert Shiller's meticulously-maintained historical data about the US market (Shiller 2005). We label GPSV(D,Y) this variant of the GPSV model. Versions with day-of-the-week only or yields only are respectively labeled GPSV(D) and GPSV(Y).

A difficulty must be addressed when contemporaneous volatility covariates are considered in a regression setting: in order to make a forecast of volatility, a separate forecast of those covariates must be carried out as well. To avoid undue complexity, we resort in this paper to a simplyfying assumption and make only a constant forecast, i.e. we repeat the last-known value of each of the above covariate throughout the forecasting horizon. This constancy assumption is what led us to choosing relatively slow-varying "yield" covariates in our experiments. One can obviously envision a more comprehensive approach, which would involve making separate forecasts for each covariate (including associated forecasting uncertainty) and using those within the GPR model. Although efficient approaches have been proposed to incorporate covariate uncertainty into the GPR framework (Girard 2004), we consider these to be out of the scope of the present paper.

5.3 Results with Explanatory Variables

Within the settings used to obtain the results in Table 3 (Section 5.1), we compute out-of-sample forecasting NLL, MSE and Bias measures for three simple, augmented models. Results are reported in Table 4.

Accounting only for the day-of-the-week effect, the GPSV(D) significantly outperforms, according to the three criteria, the benchmark GPSV(2) model in terms of one-day-ahead forecasts. On the other hand, the GPSV(D) systematically and significantly underperforms at long horizons. These results are intuitive since the day-of-the-week effect is most likely to impact forecast performance only at short horizons.

The GPSV(Y) model accounts for the current bond yield, dividend yield, earnings yield. The GPSV(Y) improvements over the benchmark GPSV(2) model are modest. However, combining the day-of-the-week seasonality with the set of exogenous variables, the

GPSV(D,Y) model yields interesting improvements. Long-horizon gains in NLL are statistically significant, but the gains in terms of MSE are not. This suggests that this particular set of exogenous variables helps better describing the distribution of future volatility, but not due to better capturing conditional expectations.

A Time Series Perspective

These results must however be interpreted with caution. Figure 4 unfolds, for the ABD(2), GPSV(2), and GPSV(D,Y) models, the weekly error measures summarized by the one-year-ahead measures in Table 4. This figure reports one-year out-of-sample forecasting bias (top row) and MSE (middle row) for the ABD(2), GPSV(2) and GPSV(D,Y) models. It is important to note that each performance measure is reported at the last date included in the out-of-sample period it corresponds to. The forecast that yield the error measure reported at time t is thus made one year earlier, with the information at time $t - 252$. Grey-shaded periods are NBER recessions. In all series, the end of the most recent recession marks a significant change in how the alternative models compare.

The upper row report weekly log-volatility biases. The three models underestimate volatility to some extent from 1997 to 2003. During the period of low volatility between 2004 and 2007, this figure is reversed and all models overestimate volatility. Then, as volatility rises entering in the last recession, underestimation is again prevalent. The middle row displays how these forecasting errors translate in terms of MSE.

The bottom row of Figure 4 reports the MSE reduction obtained by using the GPSV(2) model rather than the ABD(2) model (left), GPSV(D,Y) rather than GPSV(2) (middle), and GPSV(D,Y) rather than ABD(2) (right). To complement these plots, Table 5 provides ratios of each model's MSE against the MSE of the GPSV(D,Y) model. These ratios are reported for the full sample (Panel A) and for four different subsample (Panels B to E).

As can be seen in the lower-left plot of Figure 4, from Sept. 1997 to Nov. 2001, the GPSV(2) model consistently outperforms the ABD(2). On average over this period, the one-year out-of-sample MSE of the latter model is 35.6% higher than that the GPSV(2) model (1.60/1.18, Table 5, Panel B). Nonetheless, the GPSV(D,Y) further improves on the GPSV(2), reducing the error by 18%. Note that beyond the one-day horizon, no model outperforms the GPSV(D,Y) at any horizon.

During the relatively calm period between the two recessions in our sample, MSE differences in Figure 4 are dwarfed by the amplitude of the MSE differences in the later recession. Yet, variance ratios of the ABD(2) and GPSV(2) are respectively 1.10 and 1.03.

Thus, substantial performance gains are obtained by casting the stochastic volatility model in a GPR framework, and moderate incremental gains are obtained by accounting for exogenous variables. Over this period, no model consistently outperforms the GPSV(D,Y) at any horizon.

The greatest performance gains occur during the 2007-2009 recession. Once again, beyond the one-day horizon, no model consistently outperforms the GPSV(D,Y) at any horizon. Biases are sizable for all models during this period. However, the variance ratios of the ABD(2) and GPSV(2) are respectively 1.26 and 1.06 during this period. That is, forecasting (log-) volatility using the GPSV model augmented with a naive set of exogenous variables led to one-year out-of-sample forecasting errors 26% smaller than the errors made with the two-component model suggested by ABD. Moreover, given the results of Alizadeh, Brandt, and Diebold (2002), we can assume that their two-component model is of the most stringent benchmarks in this class of model.

These results are however in stark contrast with those obtained as volatility declines at the end of the recession. Interestingly, as the bias of the ABD(2) model narrows down to zero (Figure 4, upper left), that of the GPSV model with exogenous variables increases sharply to reach its worst levels. As noted earlier, the error measures reported at time t are based on forecasts made the year before, at time $t - 252$. In Figure 4, a dashed line marks the date of the last error measure based on forecasts made during the recession. It is interesting to note how the end of the relative under-performance of the GPSV(D,Y) model coincides with this date (lower-right panel).

Finally, it is worth keeping in mind the remarks of section 5.2 regarding the treatment of covariates over the forecasting horizon: in all cases, the "yield" covariates are kept at their last-known values at the time of making the forecast. It is sensible to suspect that covariates embedding expectations about the future yield values may lead to increased performance. Alternatively, modeling the dynamics of the exogenous variables and using the forecasts, accounting for uncertainty, would constitute a promising avenue for further refinements.

6 Conclusion

In this paper, we provide a formulation of stochastic volatility (SV) based on Gaussian process regression (GPR), a flexible framework for nonparametric Bayesian regression. This allows us to benefit from the tools developed in the GPR literature and apply these to the estimation and forecasting of volatility processes.

Part of our contribution is to show how to provide arbitrary covariates for volatility forecasting in a nonparametric setting, without having to specify *a priori* the functional forms incorporating them. Using a GPR-based stochastic volatility (GPSV) model augmented with a simple set of covariates, we obtain volatility forecasts that are significantly better than with the range-based SV model put forward by Alizadeh, Brandt, and Diebold (2002), at all horizons, from one-week- to one-year-ahead out-of-sample. In particular, during the 2007-2009 recession, the augmented GPSV model leads to one-year out-of-sample forecasting errors 26% smaller than the errors made with the benchmark range-based SV model.

Our work suggests several avenues for future research. First, as seasonalities and macro-finance covariates can straightforwardly be dealt with in a GPR framework, the methodological framework put forth in this paper could prove to be a powerful tool to investigate the links between stock market volatility and economic activity.

A time series analysis of the relative performance the augmented GPSV model highlights that the accuracy of long-horizon volatility forecasts could be improved by accounting for forecasts of the macro-finance variables themselves. Another promising avenue for further refinements would be to use the GPR methodology put forward here to model the dynamics of the exogenous variables and using the forecasts, accounting for uncertainty, to predict the main volatility process.

Finally, the ability to accurately forecast volatility over long horizons could prove highly relevant in pricing long-term derivative contracts and assessing the long-term risk exposures of complex portfolios.

References

Adler, R. J. and J. E. Taylor (2007). *Random Fields and Geometry*. Berlin: Springer.

Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002, June). Range-Based Estimation of Stochastic Volatility Models. *Journal of Finance 57*(3), 1047–1091.

Andersen, T., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

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

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and Correlation Forecasting. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 777–878. North-Holland.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2009). Parametric and Nonparametric Volatility Measurement. In Y. Ait-Sahalia and L. Hansen (Eds.), *Handbook of Financial Econometrics*, Amsterdam. Elsevier.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001a). The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001b). The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics 61*, 43–76.

Aït-Sahalia, Y. (2005). How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies 18*(2), 351–416.

Bandi, F. M. and J. R. Russell (2008). Microstructure Noise, Realized Variance, and Optimal Sampling. *Review of Economic Studies 75*, 339–369.

Barberis, N. (2000). Investing for the long run when returns are predictable. *The Journal of Finance 55*(1), 225–264.

Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models. *Journal of the Royal Statistical Society Series B 64*, 253–280.

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

Brandt, M. W. and C. S. Jones (2005). Bayesian Range-Based Estimation of Stochastic Volatility Models. *Finance Research Letters 2*, 201–209.

Brandt, M. W. and C. S. Jones (2006). Volatility Forecasting With Range-Based EGARCH Models. *Journal of Business and Economic Statistics 24*(4), 470–486.

Christoffersen, P. and F. Diebold (2000). How Relevant Is Volatility Forecasting for Financial Risk Management? *Review of Economics and Statistics 82*(1), 12–22.

Corradi, V., W. Distaso, and A. Mele (2013). Macroeconomic Determinants of Stock Volatility and Volatility Premiums. *Journal of Monetary Economics 60*, 203–220.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons.

Dorion, C. (2012). Business Conditions, Market Volatility and Options Prices. HEC Montreal Working Paper.

Duffie, D. and K. Singleton (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica 61*, 929–952.

Engle, R., E. Ghysels, and B. Sohn (2013). Stock Market Volatility and Macroeconomic Fundamentals. *Review of Economics and Statistics*, forthcoming.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. *Econometrica 50*, 987–1008.

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

Engle, R. F. and J. G. Rangel (2008). The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes. *Review of Financial Studies 21*, 1187–1222.

Feller, W. (1951). The Asymptotic Distribution of the Range of Sums of Independent Random Variables. *Annals of Mathematical Statistics 22*(3), 427–432.

Flannery, M. J. and A. A. Protopapadakis (2002). Macroeconomic Factors *Do* Influence Aggregate Stock Returns. *Review of Financial Studies 15*, 751–782.

Gallant, A. R., D. A. Hsieh, and G. E. Tauchen (1997). Estimation of Stochastic Volatility Models with Diagnostics. *Journal of Econometrics 81*, 159–192.

Garman, M. B. and M. J. Klass (1980). On the estimation of price volatility from historical data. *Journal of Business 53*(1), 67–78.

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic Volatility. In G. Maddala and C. Rao (Eds.), *Handbook of Statistics*, Volume 14. North-Holland.

Girard, A. (2004). *Approximate methods for propagation of uncertainty with Gaussian process models*. Ph. D. thesis, University of Glasgow.

Hamilton (1994). *Time Series Analysis*. Princeton University Press.

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

Hoevenaars, R., R. Molenaar, P. Schotman, and T. Steenkamp (2011). Strategic asset allocation for long-term investors: Parameter uncertainty and prior information. *Working paper*.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994, October). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business and Economic Statistics 12*(4), 69–87.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering 82*(Series D), 35–45.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies 65*, 361–393.

Li, J., V. Todorov, and G. Tauchen (2012). Volatility Occupation Times. Working Paper.

MacKay, D. J. C. (1999). Comparison of Approximate Methods for Handling Hyperparameters. *Neural Computation 11*(5), 1035–1068.

Matheron, G. (1973). The Intrinsic Random Functions and Their Applications. *Advances in Applied Probability 5*, 439–468.

Matérn, B. (1960). *Spatial Variation*. Meddelanden från Statens Skogsforskningsinstitut, 49, No. 5. Stockholm: Almänna Förlaget. Second edition (1986), Springer-Verlag, Berlin.

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

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703–708.

O'Hagan, A. (1978). Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society B 40*, 1–42. (With discussion).

Parkinson, M. (1980). The Extreme Value Method for Estimating the Variance of the Rate of Return. *Journal of Business 53*(1), 61–65.

Pástor, L'. and R. F. Stambaugh (2012). Are Stocks Really Less Volatile in the Long Run? *The Journal of Finance 67*(2), 431–478.

Quiñonero-Candela, J. and C. E. Rasmussen (2005). A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research 6*, 1935–1959.

Rasmussen, C. E. and H. Nickisch (2010, December). Gaussian Processes for Machine Learning (GPML) Toolbox. *Journal of Machine Learning Research 11*, 3011–3015.

Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. MIT Press.

Robinson, P. M. (2001, April). The memory of stochastic volatility models. *Journal of Econometrics 101*(2), 195–218.

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

Shephard, N. (Ed.) (2004). *Stochastic Volatility: Selected Readings*. Oxford, UK: Oxford University Press.

Shiller, R. (2005). *Irrational Exuberance* (Second ed.). Princeton University Press. Data source: `http://www.econ.yale.edu/~shiller/data.htm`.

Snelson, E. (2007). *Flexible and efficient Gaussian process models for machine learning*. Ph. D. thesis, University College London.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY: Springer.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, B 36*(1), 111–147.

Sundararajan, S. and S. S. Keerthi (2001). Predictive Approaches for Choosing Hyperparameters in Gaussian Processes. *Neural Computation 13*(5), 1103–1118.

Taylor, S. J. (1994). Modelling Stochastic Volatility. *Mathematical Finance 4*, 183–204.

West, K. D. and D. Cho (1995). The Predictive Ability of Several Models of Exchange Rate Volatility. *Journal of Econometrics 69*, 367—-391.

Williams, C. K. I. and C. E. Rasmussen (1996). Gaussian Processes for Regression. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Eds.), *Advances in Neural Information Processing Systems 8*, pp. 514–520. MIT Press.

Appendix

A Predictive Distribution

A.1 Stochastic Volatility with One Component

In the SV1 model, log-volatility follows equation (2.5), which is equivalent to

$$\log \sigma_{t+1} - \log \bar{\sigma} = \rho_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau}\, \varepsilon^{\sigma}_{t+1} \,. \tag{A.1}$$

The value of the log-volatility process two steps ahead can thus be written

$$\begin{aligned}
\log \sigma_{t+2} - \log \bar{\sigma} &= \rho_\tau(\log \sigma_{t+1} - \log \bar{\sigma}) + \beta \sqrt{\tau}\, \varepsilon^{\sigma}_{t+2} && \text{(A.2)}\\
&= \rho_\tau \left[\rho_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau}\, \varepsilon^{\sigma}_{t+1}\right] + \beta \sqrt{\tau}\, \varepsilon^{\sigma}_{t+2} && \text{(A.3)}\\
&= \rho^2_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau}\left(\varepsilon^{\sigma}_{t+2} + \rho_\tau \varepsilon^{\sigma}_{t+1}\right) && \text{(A.4)}
\end{aligned}$$

If we have that at $t + k$

$$\log \sigma_{t+k} - \log \bar{\sigma} = \rho^k_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau} \sum_{j=1}^{k} \rho^{k-j}_\tau \varepsilon_{t+j} \tag{A.5}$$

then at $t + k + 1$ we have

$$\begin{aligned}
\log \sigma_{t+k+1} - \log \bar{\sigma} &= \rho_\tau \left[\rho^k_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau} \sum_{j=1}^{k} \rho^{k-j}_\tau \varepsilon_{t+j}\right] + \beta \sqrt{\tau}\, \varepsilon^{\sigma}_{t+k+1} && \text{(A.6)}\\
&= \rho^{k+1}_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau}\left(\varepsilon^{\sigma}_{t+k+1} + \rho_\tau \sum_{j=1}^{k} \rho^{k-j}_\tau \varepsilon_{t+j}\right) && \text{(A.7)}\\
&= \rho^{k+1}_\tau(\log \sigma_t - \log \bar{\sigma}) + \beta \sqrt{\tau} \sum_{j=1}^{k+1} \rho^{k+1-j}_\tau \varepsilon_{t+j} \,. && \text{(A.8)}
\end{aligned}$$

Hence, by induction, (A.5) holds for all $k \geq 1$. Taking the expectation at time t on both sides yields equation (2.7). To obtain equation (2.8), simply note that, assuming $k < m$, we have

$$\begin{aligned}
&\operatorname{Cov}_t(\log \sigma_{t+k}, \log \sigma_{t+m})\\
&= \operatorname{Cov}_t\left(\rho^k_\tau \log \sigma_t + \beta \sqrt{\tau} \sum_{j=1}^{k} \rho^{k-j}_\tau \varepsilon_{t+j}\,,\ \rho^m_\tau \log \sigma_t + \beta \sqrt{\tau} \sum_{j=1}^{m} \rho^{m-j}_\tau \varepsilon_{t+j}\right) && \text{(A.9)}\\
&= \rho^{k+m}_\tau \operatorname{Var}_t(\log \sigma_t) + \beta^2 \tau \operatorname{Cov}_t\left(\sum_{j=1}^{k} \rho^{k-j}_\tau \varepsilon_{t+j}\,,\ \sum_{j=1}^{k} \rho^{m-j}_\tau \varepsilon_{t+j} + \sum_{j=k+1}^{m} \rho^{m-j}_\tau \varepsilon_{t+j}\right) && \text{(A.10)}\\
&= \rho^{k+m}_\tau \operatorname{Var}_t(\log \sigma_t) + \beta^2 \tau \sum_{j=1}^{k} \rho^{k+m-2j}_\tau \ = \ \rho^{k+m}_\tau \operatorname{Var}_t(\log \sigma_t) + \beta^2 \tau \sum_{n=0}^{k-1} \rho^{k+m-2(k-n)}_\tau && \text{(A.11)}\\
&= \rho^{k+m}_\tau \operatorname{Var}_t(\log \sigma_t) + \beta^2 \tau \rho^{m-k}_\tau \sum_{n=0}^{k-1} \rho^{2n}_\tau \ = \ \rho^{k+m}_\tau \operatorname{Var}_t(\log \sigma_t) + \beta^2 \tau \rho^{m-k}_\tau \left(\frac{1 - \rho^{2k}_\tau}{1 - \rho^2_\tau}\right) . && \text{(A.12)}
\end{aligned}$$

where $\mathrm{Var}_t(\log\sigma_t)$ is obtained via the Kalman filter. If this estimation error were null, i.e. $\mathrm{Var}_t(\log\sigma_t) = 0$, the correlation between the two next log-volatility levels would be given by

$$\mathrm{Corr}_t(\log\sigma_{t+1}, \log\sigma_{t+2}) = \frac{\beta^2\tau\rho_\tau\left(\frac{1-\rho_\tau^2}{1-\rho_\tau^2}\right)}{\left[\beta^2\tau\left(\frac{1-\rho_\tau^2}{1-\rho_\tau^2}\right)\ \beta^2\tau\left(\frac{1-\rho_\tau^4}{1-\rho_\tau^2}\right)\right]^{1/2}} = \frac{\rho}{\sqrt{1+\rho^2}}\,. \tag{A.13}$$

A.2 Stochastic Volatility with Two Components

In the SV2 model, we have

$$\begin{aligned}\log\sigma_{t+1} - \log\bar{\sigma} &= \log\sigma_{1,t+1} + \log\sigma_{2,t+1} &\text{(A.14)}\\ &= \rho_{1,\tau}\log\sigma_{1,t} + \beta_1\sqrt{\tau}\,\varepsilon^\sigma_{1,t+1} + \rho_{2,\tau}\log\sigma_{2,t} + \beta_2\sqrt{\tau}\,\varepsilon^\sigma_{2,t+1}\,. &\text{(A.15)}\end{aligned}$$

By induction, we can show that

$$\log\sigma_{t+k} - \log\bar{\sigma} = \rho^k_{1,\tau}\log\sigma_{1,t} + \rho^k_{2,\tau}\log\sigma_{2,t} + \beta_1\sqrt{\tau}\sum_{j=1}^{k}\rho^{k-j}_{1,\tau}\varepsilon_{1,t+j} + \beta_2\sqrt{\tau}\sum_{j=1}^{k}\rho^{k-j}_{2,\tau}\varepsilon_{2,t+j}\,. \tag{A.16}$$

Taking conditional expectations on both sides yield

$$\mathbb{E}_t[\log\sigma_{t+k}] = \log\bar{\sigma} + \rho^k_{1,\tau}\,\mathbb{E}_t[\log\sigma_{1,t}] + \rho^k_{2,\tau}\,\mathbb{E}_t[\log\sigma_{2,t}]\ , \tag{A.17}$$

where the $\mathbb{E}_t\left[\log\sigma_{p,t}\right]$ is given by the Kalman filter. The covariance between two future observations, $k < m$, is given by

$$\begin{aligned}&\mathrm{Cov}_t(\log\sigma_{t+k}, \log\sigma_{t+m})\\ &\quad= \mathrm{Cov}_t\Big(\rho^k_{1,\tau}\log\sigma_{1,t} + \rho^k_{2,\tau}\log\sigma_{2,t} + \beta_1\sqrt{\tau}\sum_{j=1}^{k}\rho^{k-j}_{1,\tau}\varepsilon_{1,t+j} + \beta_2\sqrt{\tau}\sum_{j=1}^{k}\rho^{k-j}_{2,\tau}\varepsilon_{2,t+j}\,,\\ &\qquad\qquad \rho^m_{1,\tau}\log\sigma_{1,t} + \rho^m_{2,\tau}\log\sigma_{2,t} + \beta_1\sqrt{\tau}\sum_{j=1}^{m}\rho^{m-j}_{1,\tau}\varepsilon_{1,t+j} + \beta_2\sqrt{\tau}\sum_{j=1}^{m}\rho^{m-j}_{2,\tau}\varepsilon_{2,t+j}\Big) &\text{(A.18)}\\ &\quad= \rho^{k+m}_{1,\tau}\,\mathrm{Var}_t(\log\sigma_{1,t}) + \beta_1^2\tau\rho^{m-k}_{1,\tau}\left(\frac{1-\rho^{2k}_{1,\tau}}{1-\rho^2_{1,\tau}}\right) + \rho^{k+m}_{2,\tau}\,\mathrm{Var}_t(\log\sigma_{2,t}) + \beta_2^2\tau\rho^{m-k}_{2,\tau}\left(\frac{1-\rho^{2k}_{2,\tau}}{1-\rho^2_{2,\tau}}\right),\\ &\qquad\quad + \left(\rho^k_{1,\tau}\rho^m_{2,\tau} + \rho^m_{1,\tau}\rho^k_{2,\tau}\right)\mathrm{Cov}_t\left(\log\sigma_{1,t}\,,\ \log\sigma_{2,t}\right). &\text{(A.19)}\end{aligned}$$

where $\mathrm{Var}_t\left(\log\sigma_{p,t}\right)$, $p = 1, 2$, and $\mathrm{Cov}_t\left(\log\sigma_{1,t}\,,\ \log\sigma_{2,t}\right)$ are obtained via the Kalman filter.

A.3 REGARCH Volatility with One Component

We can rewrite equation (2.16) as

$$\log\sigma_{t+1} - \theta = \rho(\log\sigma_t - \theta) + \phi\,\tilde{r}_t + \delta\,\varepsilon_t\,, \tag{A.20}$$

$$\text{where}\quad \rho = (1-\kappa)\quad\text{and}\quad \tilde{r}_t \equiv \frac{r_t - 0.43 - \log\sigma_t}{0.29} \sim \mathcal{N}\left(0.43 + \log\sigma_{t+1}, 0.29^2\right). \tag{A.21}$$

As opposed to (A.1), the above value of $\log \sigma_{t+1}$ is not a random variable at time t, given that r_t and ε_t are observed. Hence, equation (A.5) into

$$\log \sigma_{t+k} - \theta = \rho^{k-1}(\log \sigma_{t+1} - \theta) + \phi \sum_{j=1}^{k-1} \rho^{k-1-j} \tilde{r}_{t+j} + \delta \sum_{j=1}^{k-1} \rho^{k-1-j} \varepsilon_{t+j} \tag{A.22}$$

Taking the expectation on both sides at t yields

$$\mathbb{E}_t \left[\log \sigma_{t+k} - \theta\right] = \rho^{k-1}(\log \sigma_{t+1} - \theta) \tag{A.23}$$

much alike equation (2.7). To obtain an equation analogous to (2.8), we have that,

$$\begin{aligned} &\mathrm{Cov}_t\left(\log \sigma_{t+k}, \log \sigma_{t+m}\right) \\ &\quad = \mathrm{Cov}_t\left(\frac{\phi}{\rho}\sum_{j=1}^{k-1} \rho^{k-j}\tilde{r}_{t+j} + \frac{\delta}{\rho}\sum_{j=1}^{k-1} \rho^{k-j}\varepsilon_{t+j}\,,\ \frac{\phi}{\rho}\sum_{j=1}^{m-1} \rho^{m-j}\tilde{r}_{t+j} + \frac{\delta}{\rho}\sum_{j=1}^{m-1} \rho^{m-j}\varepsilon_{t+j}\right), \end{aligned} \tag{A.24}$$

where we use the fact that, in GARCH filtering, $\log \sigma_{t+1}$ is known. Using the fact that $\tilde{r}_t$ and ε_t are iid standard normal, and assuming $2 \leq k < m$, we have that

$$\begin{aligned} &\mathrm{Cov}_t\left(\log \sigma_{t+k}, \log \sigma_{t+m}\right) \\ &\quad = \frac{\phi^2}{\rho^2}\mathrm{Cov}_t\left(\sum_{j=1}^{k-1} \rho^{k-j}\tilde{r}_{t+j}\,,\ \sum_{j=1}^{m-1} \rho^{m-j}\tilde{r}_{t+j}\right) + \frac{\delta^2}{\rho^2}\mathrm{Cov}_t\left(\sum_{j=1}^{k-1} \rho^{k-j}\varepsilon_{t+j}\,,\ \sum_{j=1}^{m-1} \rho^{m-j}\varepsilon_{t+j}\right) && \text{(A.25)} \\ &\quad = \frac{\phi^2}{\rho^2}\sum_{j=1}^{k-1} \rho^{k+m-2j} + \frac{\delta^2}{\rho^2}\sum_{j=1}^{k-1} \rho^{k+m-2j} = \left(\frac{\phi^2}{\rho^2} + \frac{\delta^2}{\rho^2}\right)\sum_{n=1}^{k-1} \rho^{k+m-2(n-k)} && \text{(A.26)} \\ &\quad = \left(\phi^2 + \delta^2\right)\rho^{m-k-2}\sum_{n=1}^{k-1} \rho^{-2n} = \left(\phi^2 + \delta^2\right)\rho^{m-k}\left(\frac{1-\rho^{2(k-1)}}{1-\rho^2}\right). && \text{(A.27)} \end{aligned}$$

The correlation between the two next log-volatility levels is given by

$$\mathrm{Corr}_t\left(\log \sigma_{t+2}, \log \sigma_{t+3}\right) = \frac{\left(\phi^2 + \delta^2\right)\rho^{-1}\left(\frac{\rho^2-\rho^4}{1-\rho^2}\right)}{\left[\left(\phi^2 + \delta^2\right)\rho^{-2}\frac{\rho^2(1-\rho^2)}{(1-\rho^2)}\ \left(\phi^2 + \delta^2\right)\rho^{-2}\frac{\rho^2(1-\rho^4)}{(1-\rho^2)}\right]^{1/2}} = \frac{\rho}{\sqrt{1+\rho^2}}. \tag{A.28}$$

A.4 REGARCH Volatility with Two Components

Similarly to the one-component case, we can rewrite equations (2.19) and (2.20) as

$$\log \sigma_{1,t+1} = \log \sigma_{2,t+1} + \rho_1(\log \sigma_{1,t} - \log \sigma_{2,t}) + \phi_1\,\tilde{r}_t + \delta_1\,\varepsilon_t\,, \tag{A.29}$$

$$\log \sigma_{2,t+1} - \theta = \rho_2(\log \sigma_{2,t} - \theta) + \phi_2\,\tilde{r}_t + \delta_2\,\varepsilon_t\,, \tag{A.30}$$

where $\rho_i = 1 - \kappa_i$. Note that these two equations hold at any time t; thus, we can write, for any $k \geq 2$,

$$\log \sigma_{1,t+k} = \log \sigma_{2,t+k} + \rho_1(\log \sigma_{1,t+k-1} - \log \sigma_{2,t+k-1}) + \phi_1\,\tilde{r}_{t+k-1} + \delta_1\,\varepsilon_{t+k-1}\,, \tag{A.31}$$

$$\begin{aligned} \Leftrightarrow \log \sigma_{1,t+k} - \theta &= \log \sigma_{2,t+k} - \theta + \rho_1(\log \sigma_{1,t+k-1} - \log \sigma_{2,t+k-1}) + \phi_1\,\tilde{r}_{t+k-1} + \delta_1\,\varepsilon_{t+k-1}\,, && \text{(A.32)} \\ &= \rho_2(\log \sigma_{2,t+k-1} - \theta) + \rho_1(\log \sigma_{1,t+k-1} - \log \sigma_{2,t+k-1}) \\ &\qquad + \phi_1\,\tilde{r}_{t+k-1} + \delta_1\,\varepsilon_{t+k-1} + \phi_2\,\tilde{r}_{t+k-1} + \delta_2\,\varepsilon_{t+k-1}\,. && \text{(A.33)} \end{aligned}$$

Proceeding recursively, this equation becomes

$$\log \sigma_{t+k} - \theta = \rho_2^{k-1}(\log \sigma_{2,t+1} - \theta) + \rho_1^{k-1}(\log \sigma_{1,t+1} - \log \sigma_{2,t+1}) \\ + \phi_1 \sum_{j=1}^{k-1} \rho_1^{k-1-j} \tilde{r}_{t+j} + \delta_1 \sum_{j=1}^{k-1} \rho_1^{k-1-j} \varepsilon_{t+j} + \phi_2 \sum_{j=1}^{k-1} \rho_2^{k-1-j} \tilde{r}_{t+j} + \delta_2 \sum_{j=1}^{k-1} \rho_2^{k-1-j} \varepsilon_{t+j} \quad \text{(A.34)}$$

Taking the expectation on both sides at t yields

$$\mathbb{E}_t \left[\log \sigma_{1,t+k} - \theta\right] = \rho_2^{k-1}(\log \sigma_{2,t+1} - \theta) + \rho_1^{k-1}(\log \sigma_{1,t+1} - \log \sigma_{2,t+1}) \quad \text{(A.35)}$$

which is such that, when k tends to infinity, $\mathbb{E}_t \left[\log \sigma_{t+k}\right]$ tends to θ. For the covariance, we have

$$\begin{aligned}
&\operatorname{Cov}_t \left(\log \sigma_{1,t+k}, \log \sigma_{1,t+m}\right) \\
&= \operatorname{Cov}_t \left(\frac{\phi_1}{\rho_1} \sum_{j=1}^{k-1} \rho_1^{k-j} \tilde{r}_{t+j} + \frac{\delta_1}{\rho_1} \sum_{j=1}^{k-1} \rho_1^{k-j} \varepsilon_{t+j}\, , + \frac{\phi_2}{\rho_2} \sum_{j=1}^{k-1} \rho_2^{k-j} \tilde{r}_{t+j} + \frac{\delta_2}{\rho_2} \sum_{j=1}^{k-1} \rho_2^{k-j} \varepsilon_{t+j}\, , \right. \\
&\qquad \left. \frac{\phi_1}{\rho_1} \sum_{j=1}^{m-1} \rho_1^{m-j} \tilde{r}_{t+j} + \frac{\delta_1}{\rho_1} \sum_{j=1}^{m-1} \rho_1^{m-j} \varepsilon_{t+j} + \frac{\phi_2}{\rho_2} \sum_{j=1}^{m-1} \rho_2^{m-j} \tilde{r}_{t+j} + \frac{\delta_2}{\rho_2} \sum_{j=1}^{m-1} \rho_2^{m-j} \varepsilon_{t+j} \right). \qquad \text{(A.36)}
\end{aligned}$$

Using the fact that $\tilde{r}_t$ and ε_t are iid standard normal, and assuming $2 \leq k < m$, we have that

$$\begin{aligned}
&\operatorname{Cov}_t \left(\log \sigma_{1,t+k}, \log \sigma_{1,t+m}\right) \\
&= \frac{\phi_1^2}{\rho_1^2} \sum_{j=1}^{k-1} \rho_1^{m+k-2j} + \frac{\phi_2^2}{\rho_2^2} \sum_{j=1}^{k-1} \rho_2^{m+k-2j} + \frac{\phi_1 \phi_2}{\rho_1 \rho_2} \sum_{j=1}^{k-1} \rho_1^{k-j} \rho_2^{m-j} + \rho_1^{m-j} \rho_2^{k-j} \\
&\qquad + \frac{\delta_1^2}{\rho_1^2} \sum_{j=1}^{k-1} \rho_1^{m+k-2j} + \frac{\delta_2^2}{\rho_2^2} \sum_{j=1}^{k-1} \rho_2^{m+k-2j} + \frac{\delta_1 \delta_2}{\rho_1 \rho_2} \sum_{j=1}^{k-1} \rho_1^{k-j} \rho_2^{m-j} + \rho_1^{m-j} \rho_2^{k-j} \qquad \text{(A.37)} \\
&= \left(\frac{\phi_1^2}{\rho_1^2} + \frac{\delta_1^2}{\rho_1^2} \right) \sum_{j=1}^{k-1} \rho_1^{m+k-2j} + \left(\frac{\phi_2^2}{\rho_2^2} + \frac{\delta_2^2}{\rho_2^2} \right) \sum_{j=1}^{k-1} \rho_2^{m+k-2j} \\
&\qquad + \left(\frac{\phi_1 \phi_2}{\rho_1 \rho_2} + \frac{\delta_1 \delta_2}{\rho_1 \rho_2} \right) \sum_{j=1}^{k-1} \rho_1^{k-j} \rho_2^{m-j} + \rho_1^{m-j} \rho_2^{k-j} \qquad \text{(A.38)} \\
&= \left(\phi_1^2 + \delta_1^2\right) \rho_1^{m-k-2} \left(\frac{\rho_1^2 - \rho_1^{2k}}{1 - \rho_1^2} \right) + \left(\phi_2^2 + \delta_2^2\right) \rho_2^{m-k-2} \left(\frac{\rho_2^2 - \rho_2^{2k}}{1 - \rho_2^2} \right) \\
&\qquad + \left(\frac{\phi_1 \phi_2}{\rho_1 \rho_2} + \frac{\delta_1 \delta_2}{\rho_1 \rho_2} \right) \left(\rho_1^k \rho_2^m + \rho_1^m \rho_2^k \right) \sum_{j=1}^{k-1} (\rho_1 \rho_2)^{-j} \qquad \text{(A.39)} \\
&= \left(\phi_1^2 + \delta_1^2\right) \rho_1^{m-k-2} \left(\frac{\rho_1^2 - \rho_1^{2k}}{1 - \rho_1^2} \right) + \left(\phi_2^2 + \delta_2^2\right) \rho_2^{m-k-2} \left(\frac{\rho_2^2 - \rho_2^{2k}}{1 - \rho_2^2} \right) \\
&\qquad + \left(\phi_1 \phi_2 + \delta_1 \delta_2\right) \frac{\rho_1^k \rho_2^m + \rho_1^m \rho_2^k - \rho_1 \rho_2^{m-k+1} - \rho_1^{m-k+1} \rho_2}{\rho_1 \rho_2 (\rho_1 \rho_2 - 1)} \qquad \text{(A.40)} \\
&= \left(\phi_1^2 + \delta_1^2\right) \rho_1^{m-k} \left(\frac{1 - \rho_1^{2(k-1)}}{1 - \rho_1^2} \right) + \left(\phi_2^2 + \delta_2^2\right) \rho_2^{m-k} \left(\frac{1 - \rho_2^{2(k-1)}}{1 - \rho_2^2} \right) \\
&\qquad + \left(\phi_1 \phi_2 + \delta_1 \delta_2\right) \frac{\rho_1^{k-1} \rho_2^{m-1} + \rho_1^{m-1} \rho_2^{k-1} - \rho_2^{m-k} - \rho_1^{m-k}}{\rho_1 \rho_2 - 1} \qquad \text{(A.41)}
\end{aligned}$$

B The Covariance Function of the Ornstein-Uhlenbeck Process

This appendix shows how to derive the covariance function of the Ornstein-Uhlenbeck process from its stochastic differential equation specification. It relies on the Wiener-Khintchine theorem, which states that the power spectral density of a stationary process and its covariance function are Fourier transform pairs. We start by explaining how to calculate the power spectrum.

The Fourier transform and inverse transform of a process $X(t)$ are respectively given by

$$\tilde{X}(s) = \int_{-\infty}^{\infty} X(t)e^{-2\pi i s t}\mathrm{d}t \qquad \text{and} \qquad X(t) = \int_{-\infty}^{\infty} \tilde{X}(s)e^{2\pi i s t}\mathrm{d}s, \tag{B.1}$$

where the integrals, assuming they exist for process $X(t)$, are interpreted in the mean-squared limit. For a stationary process, we compute the power spectral density from the expected second-order moments of the Fourier transform as,

$$\mathbb{E}\left[\tilde{X}(s_1)\tilde{X}^*(s_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{E}\left[X(t_1)X^*(t_2)\right] e^{-2\pi i s_1 t_1} e^{2\pi i s_2 t_2}\mathrm{d}t_1\mathrm{d}t_2, \tag{B.2}$$

where x^* denotes the complex conjugate of x. For a stationary process, we rewrite $\mathbb{E}\left[X(t_1)X^*(t_2)\right]$ as $k(\tau)$ using the change of variables $\tau = t_1 - t_2$, yielding

$$\mathbb{E}\left[\tilde{X}(s_1)\tilde{X}^*(s_2)\right] = \int_{-\infty}^{\infty} k(\tau)e^{-2\pi i s_1 \tau}\mathrm{d}\tau \int_{-\infty}^{\infty} e^{-2\pi i (s_1 - s_2)t_2}\mathrm{d}t_2 \tag{B.3}$$

$$= S(s_1)\delta(s_1 - s_2), \tag{B.4}$$

where $S(s_1)$ is the power spectral density of process $X(t)$, and the Dirac delta is obtained from $\delta(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t}\mathrm{d}t$.

Now consider a first-order Gaussian process $X(t)$ evolving according to

$$\mathrm{d}X(t) + aX(t)\,\mathrm{d}t = b\,\mathrm{d}W(t), \tag{B.5}$$

where $W(t)$ is a Wiener process and the parameters a and b are assumed suitably chosen to ensure stationarity. In the case of the log-volatility process (2.4), it is a simple matter to to find a and b to match the process specification (and without loss of generality, shifting the process to give it a mean of zero).

Denote by $\tilde{X}(s)$ and $\tilde{W}(s)$ respectively the Fourier transform of $X(t)$ and $W(t)$. Note that since complex exponentials are eigenfunctions of the differential operator, we have that $\mathrm{d}X(t) = 2\pi i s\,\tilde{X}(s)$. The Fourier transform of process (B.5) is obtained as

$$(2\pi i s + a)\tilde{X}(s) = b\tilde{W}(s). \tag{B.6}$$

From (B.4), the expectation of the complex conjugate of this expression yields the power spectral density of the process

$$(2\pi i s + a)(2\pi i s + a)^*\,\mathbb{E}\left[\tilde{X}(s_1)\tilde{X}^*(s_2)\right] = b^2\,\mathbb{E}\left[\tilde{W}(s_1)\tilde{W}^*(s_2)\right].$$

Since the power spectral density of white noise is 1, a comparison with eq. (B.4) yields the power spectrum of $X(t)$,

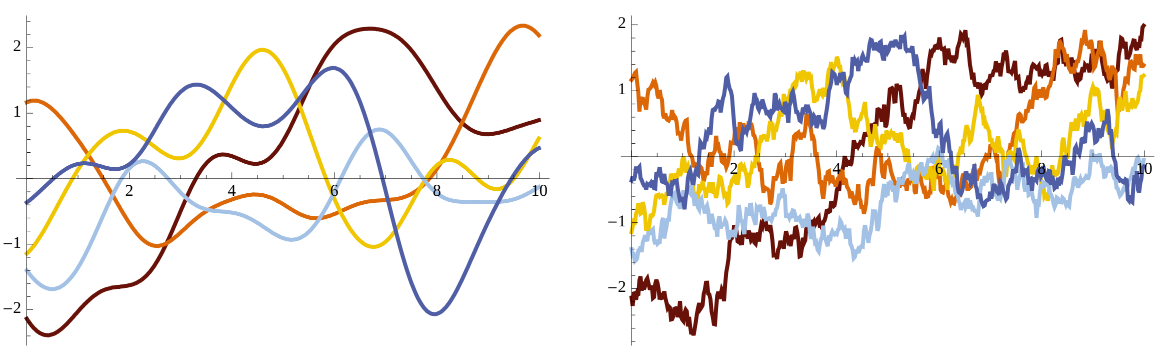
$$S(s) = \frac{b^2}{(4\pi^2 s^2 + a^2)}.$$

Taking the inverse Fourier transform (B.1) yields the covariance function $k(t)$

$$k(t) = \frac{b^2}{2a} e^{-a|t|}. \tag{B.7}$$

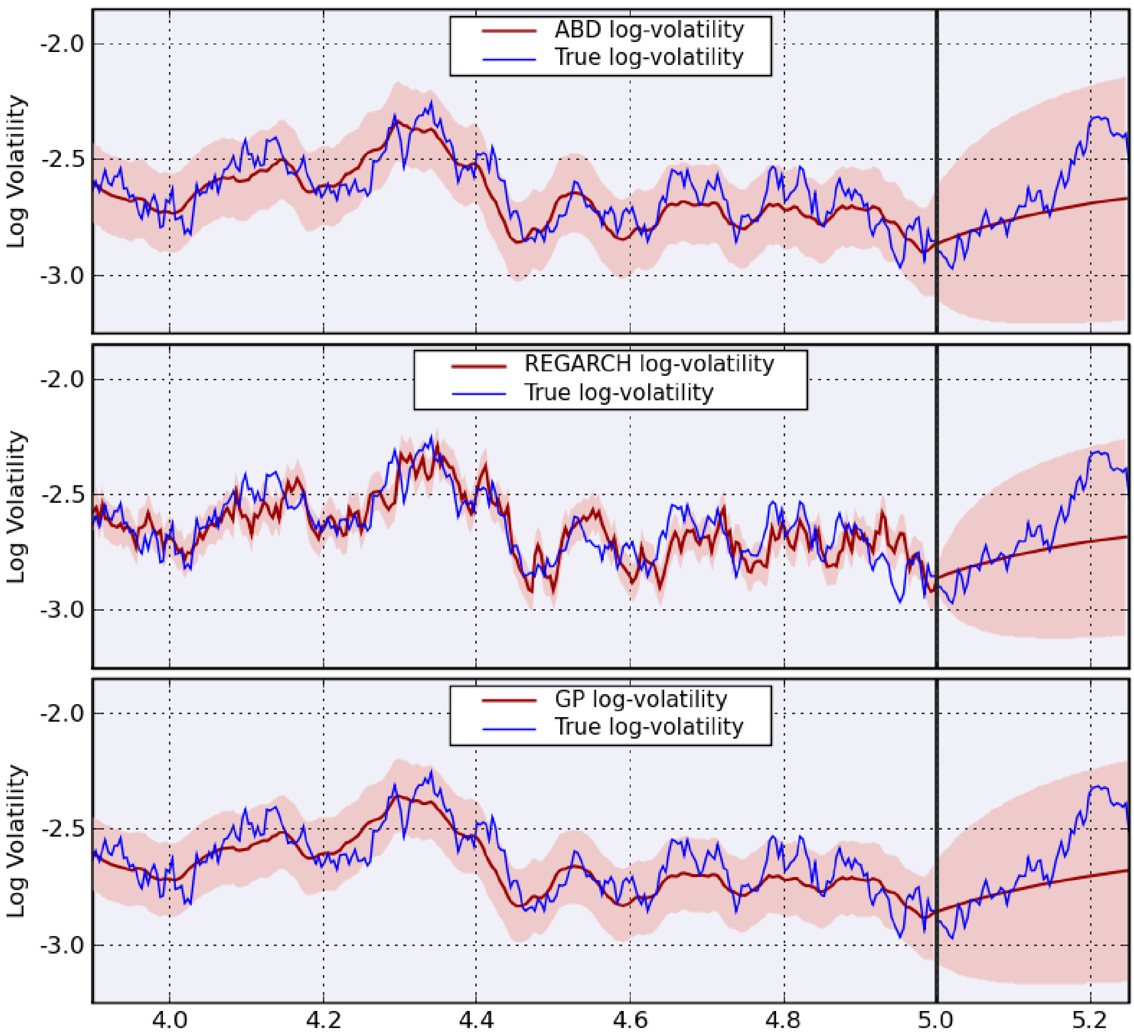
Figures and Tables

Figure 1: *Trajectories drawn from the GP prior with squared exponential and Ornstein–Uhlenbeck covariance functions*



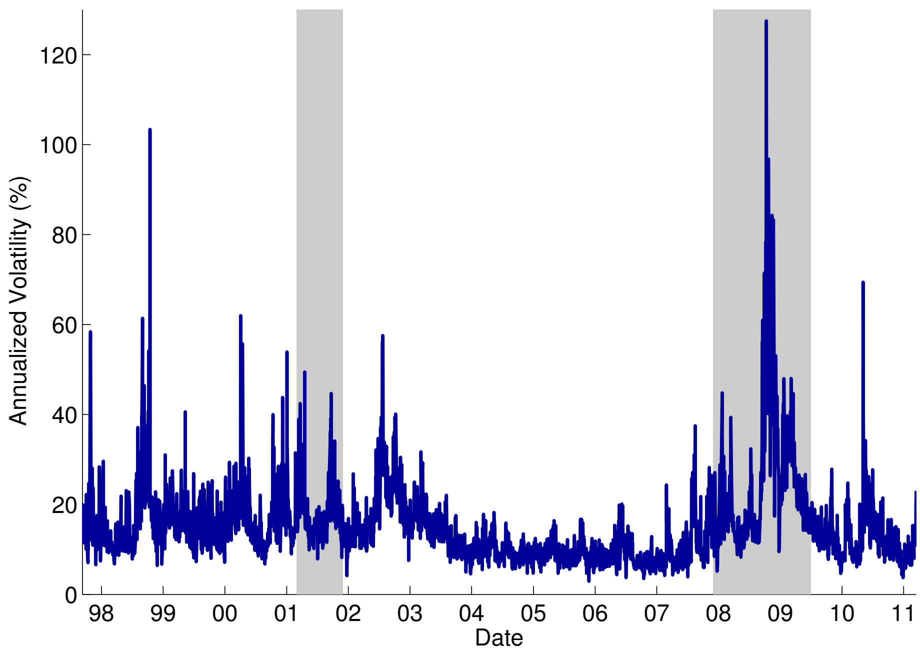
Left: *Random functions drawn from the GP prior with the squared exponential covariance ($\eta_f = \eta_\ell = 1$).* ***Right:*** *"Same functions" under the Ornstein–Uhlenbeck covariance ($\eta_f = 1, \eta_\ell = \frac{1}{2}$)*

Figure 2: *Recovery of a known latent volatility process by the ABD, EGARCH and GP models.*



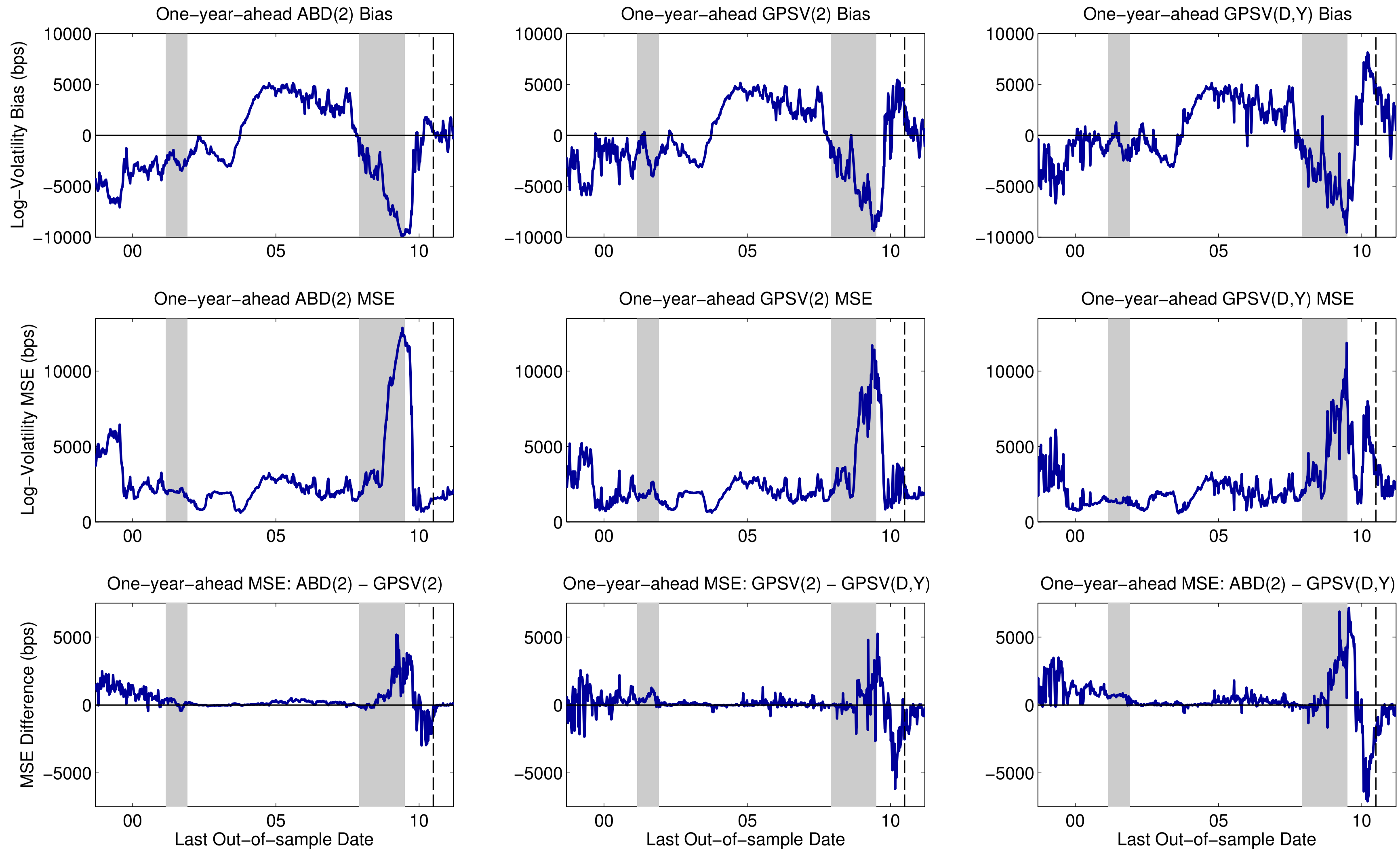
The last one of 5 in-sample years is shown, along with a 3-month forecast (beyond the thick vertical lines). Shaded regions enclose a 95% confidence interval.

Figure 3: *Realized Volatility Series*



This figure plots the realized volatility series used as target in the empirical analysis. Although it is used in the log-volatility domain, the series is graphed in the annualized volatility domain, taking the square root of RV_t *in equation* (5.1). *The realized volatility series is on average 7.3% below the VIX, and both series have a correlation of 82.5%.*

Figure 4: *Time-series of One-year Out-of-sample Forecasting Errors.*



On the first week in the sample, model estimation is performed using 5 years of past observations. Then, the estimation sample is rolled week forward, and the process is repeated. A total of 653 weeks of data are considered. This figure reports one-year out-of-sample forecasting Bias (top row) and MSE (middle row) for the ABD(2), GPSV(2) and GPSV(D,Y) models. The panels in the bottom row report the MSE reduction obtained by using the GPSV(2) model rather than the ABD(2) model (left), GPSV(D,Y) rather than GPSV(2) (middle), and GPSV(D,Y) rather than ABD(2) (right). Each performance measure is reported at the last out-of-sample date; the forecast was thus made one-year earlier. Grey-shaded periods are NBER recessions. The vertical dashed line, at the end of June 2010, highlights the forecast made on the last date of the recession ending in June 2009. The y-axis is in log-volatility basis points. If the volatility is currently 20%, an error of +5,000 (-5,000) bps in the log-domain amounts to an error of +13.0% (-7.9%) in the volatility domain.

Table 1: *Results based on data simulated using the ABD model.*

		Average and Std. Error			Paired t-test vs ABD	
		ABD	**REGARCH**	**GPSV**	**REGARCH**	**GPSV**
In-sample MSE (in bps)		68.73	133.23	68.69	-64.50	0.03
		(0.63)	(1.40)	(0.64)	**(-56.48)**	(0.58)
In-sample Bias (in bps)		11.94	13.72	11.68	-0.09	0.12
		(7.54)	(7.58)	(7.54)	(-0.20)	(0.88)
Out-of-sample NLL	**H 1**	-0.76	1.69	0.15	-2.45	-0.91
		(0.01)	(0.05)	(0.03)	**(-54.89)**	**(-37.89)**
	H 5	-7.22	-0.67	-5.93	-6.55	-1.29
		(0.02)	(0.11)	(0.03)	**(-65.05)**	**(-52.54)**
	H 10	-15.30	-8.54	-13.61	-6.76	-1.69
		(0.02)	(0.11)	(0.04)	**(-66.73)**	**(-64.44)**
	H 21	-33.07	-25.86	-30.51	-7.21	-2.57
		(0.03)	(0.12)	(0.04)	**(-69.82)**	**(-74.94)**
	H 63	-101.01	-92.08	-95.07	-8.93	-5.94
		(0.06)	(0.15)	(0.09)	**(-77.83)**	**(-72.30)**
	H 252	-406.71	-390.04	-385.61	-16.68	-21.10
		(0.13)	(0.28)	(0.34)	**(-79.83)**	**(-65.12)**
Out-of-sample MSE (in bps)	**H 1**	126.79	125.55	125.55	1.24	1.25
		(1.55)	(1.54)	(1.54)	**(15.90)**	**(18.15)**
	H 5	165.06	163.84	163.76	1.22	1.30
		(1.90)	(1.89)	(1.90)	**(14.78)**	**(18.64)**
	H 10	210.13	209.01	208.80	1.12	1.33
		(2.38)	(2.37)	(2.37)	**(11.89)**	**(17.98)**
	H 21	293.67	292.72	292.15	0.96	1.52
		(3.25)	(3.24)	(3.23)	**(7.72)**	**(17.46)**
	H 63	500.85	500.83	498.24	0.02	2.61
		(5.12)	(5.12)	(5.08)	(0.07)	**(17.01)**
	H 252	718.82	722.55	715.65	-3.72	3.17
		(5.10)	(5.16)	(5.07)	**(-8.19)**	**(11.72)**
Out-of-sample Bias (in bps)	**H 1**	243.26	232.37	236.70	4.56	5.05
		(10.99)	(10.96)	(10.95)	**(12.79)**	**(16.41)**
	H 5	242.61	231.02	236.35	4.66	5.24
		(11.90)	(11.87)	(11.86)	**(13.35)**	**(18.09)**
	H 10	234.28	221.85	228.35	3.68	4.66
		(12.99)	(12.97)	(12.95)	**(10.47)**	**(16.80)**
	H 21	227.38	213.23	221.98	2.69	4.21
		(14.77)	(14.76)	(14.73)	**(6.97)**	**(15.12)**
	H 63	185.80	166.00	180.54	-0.04	5.65
		(18.02)	(18.04)	(17.96)	(-0.07)	**(15.16)**
	H 252	67.73	32.66	53.57	-7.25	7.93
		(17.41)	(17.52)	(17.33)	**(-6.99)**	**(11.50)**

We simulate 10,000 6-year-long log-volatility trajectories using equation (2.5). *We estimate the (data-generating) ABD, REGARCH and GPSV models using the first 5 years of each trajectory. Each model is then used to forecast log-volatility from one day to one year ahead (horizon H, in business days). For each forecast, the NLL, MSE and Bias are computed— equations* (4.2) *to* (4.4)*—and their average over the 10,000 trajectories is reported in the three left-hand side columns of the table. The standard error of the mean is reported in parentheses. The last two columns report paired t-tests results comparing the NLL, MSE and magnitude of the bias, |Bias|, trajectory per trajectory, reporting the difference as, e.g.,* $\frac{1}{10{,}000}\sum_{j=1}^{10{,}000} MSE_H^{ABD,j} - MSE_H^{GPSV,j}$. *Positive entries thus indicate that the model outperforms the benchmark; the t-statistics are in parentheses and are bold whenever the difference is significant at the 5% level.*

Table 2: *Results based on data simulated using the two-component ABD(2) model.*

	Average and Standard Errors						Paired t-test vs ABD(2)				
	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)	ABD	REGARCH	REGARCH(2)	GPSV	GPSV(2)
In-sample MSE (in bps)	497.93	472.00	1189.20	1170.55	939.23	688.82	-25.93	-717.20	-698.55	-467.22	-216.82
	(2.20)	(1.83)	(4.66)	(4.62)	(5.98)	(16.43)	**(-21.82)**	**(-174.44)**	**(-173.96)**	**(-87.35)**	**(-13.33)**
In-sample Bias (in bps)	9.41	10.39	11.89	20.25	9.03	9.81	-0.03	-0.35	0.08	0.18	0.04
	(7.61)	(7.60)	(7.64)	(7.73)	(7.60)	(7.60)	(-0.17)	(-0.44)	(0.03)	(0.86)	(0.23)
Out-of-sample NLL H 1	0.40	0.35	56.41	8.12	0.35	0.37	-0.05	-56.07	-7.78	0.00	-0.03
	(0.01)	(0.01)	(0.98)	(0.15)	(0.01)	(0.01)	**(-16.69)**	**(-57.82)**	**(-55.27)**	(-0.88)	**(-11.87)**
H 5	2.01	1.61	348.17	40.94	1.64	1.77	-0.40	-346.56	-39.32	-0.03	-0.16
	(0.02)	(0.02)	(3.26)	(0.37)	(0.02)	(0.01)	**(-47.20)**	**(-106.58)**	**(-109.89)**	**(-8.97)**	**(-31.90)**
H 10	4.06	3.22	742.52	82.33	3.30	3.54	-0.84	-739.30	-79.11	-0.08	-0.32
	(0.03)	(0.02)	(5.55)	(0.57)	(0.02)	(0.02)	**(-63.15)**	**(-133.55)**	**(-141.72)**	**(-16.38)**	**(-41.18)**
H 21	8.58	6.76	1609.34	174.13	6.98	7.44	-1.83	-1602.58	-167.37	-0.22	-0.69
	(0.04)	(0.03)	(9.79)	(0.97)	(0.03)	(0.03)	**(-82.04)**	**(-163.94)**	**(-175.85)**	**(-29.11)**	**(-54.64)**
H 63	26.04	20.27	4926.35	524.04	21.02	22.37	-5.77	-4906.08	-503.77	-0.75	-2.10
	(0.09)	(0.06)	(25.31)	(2.38)	(0.06)	(0.05)	**(-109.76)**	**(-193.97)**	**(-213.59)**	**(-54.46)**	**(-71.61)**
H 252	106.11	81.53	19876.35	2104.45	84.70	89.92	-24.57	-19794.81	-2022.92	-3.16	-8.38
	(0.21)	(0.12)	(93.53)	(8.54)	(0.12)	(0.13)	**(-158.03)**	**(-211.70)**	**(-237.72)**	**(-108.60)**	**(-83.45)**
Out-of-sample MSE H 1	1300.78	1170.24	1172.66	1169.68	1172.86	1167.94	-130.54	-2.42	0.57	-2.62	2.30
(in bps)	(18.62)	(16.89)	(17.00)	(16.89)	(17.01)	(16.87)	**(-16.17)**	(-0.81)	(0.83)	(-0.88)	**(3.45)**
H 5	1482.25	1227.67	1218.45	1227.73	1218.20	1225.95	-254.58	9.22	-0.07	9.47	1.72
	(11.30)	(8.67)	(8.56)	(8.67)	(8.56)	(8.65)	**(-37.22)**	**(7.99)**	(-0.21)	**(8.34)**	**(6.43)**
H 10	1583.53	1275.88	1268.64	1276.14	1268.25	1274.12	-307.65	7.24	-0.26	7.62	1.76
	(10.11)	(6.87)	(6.79)	(6.87)	(6.79)	(6.84)	**(-45.26)**	**(7.98)**	(-0.94)	**(8.58)**	**(7.64)**
H 21	1667.89	1356.72	1357.47	1357.40	1356.74	1355.17	-311.17	-0.75	-0.68	-0.02	1.54
	(9.33)	(5.94)	(5.95)	(5.95)	(5.95)	(5.91)	**(-47.10)**	(-0.87)	**(-2.18)**	(-0.02)	**(6.55)**
H 63	1747.73	1542.20	1555.74	1542.22	1553.63	1539.32	-205.53	-13.55	-0.02	-11.44	2.88
	(8.21)	(5.99)	(6.18)	(5.98)	(6.16)	(5.95)	**(-40.56)**	**(-12.79)**	(-0.05)	**(-10.77)**	**(7.90)**
H 252	1825.47	1766.44	1774.07	1762.17	1768.20	1761.20	-59.03	-7.63	4.27	-1.75	5.24
	(5.96)	(5.61)	(5.68)	(5.57)	(5.61)	(5.55)	**(-29.28)**	**(-12.89)**	**(5.12)**	**(-2.93)**	**(11.57)**
Out-of-sample Bias H 1	-13.67	-95.81	-50.61	-102.20	-38.68	-93.33	-157.72	1.90	0.78	1.67	3.51
(in bps)	(36.07)	(34.20)	(34.24)	(34.19)	(34.25)	(34.16)	**(-14.60)**	(0.45)	(0.76)	(0.39)	**(3.53)**
H 5	159.16	-103.88	-127.89	-92.12	-113.81	-97.83	-393.87	19.46	0.25	20.17	3.32
	(26.49)	(21.71)	(21.44)	(21.71)	(21.44)	(21.67)	**(-32.08)**	**(8.23)**	(0.40)	**(8.71)**	**(5.60)**
H 10	256.29	-85.90	-110.86	-74.65	-94.34	-81.70	-519.69	18.51	-0.36	19.19	3.50
	(25.33)	(19.05)	(18.81)	(19.06)	(18.81)	(19.00)	**(-40.61)**	**(8.34)**	(-0.53)	**(8.88)**	**(6.27)**
H 21	291.39	-75.08	-88.07	-63.29	-66.91	-72.65	-570.72	6.02	-1.64	7.92	2.69
	(24.88)	(18.15)	(18.10)	(18.17)	(18.09)	(18.11)	**(-45.44)**	**(2.66)**	**(-2.04)**	**(3.58)**	**(4.55)**
H 63	242.90	-44.62	-44.28	-28.17	-11.10	-45.94	-368.92	-23.29	-0.84	-18.40	4.25
	(23.42)	(18.97)	(19.29)	(18.98)	(19.24)	(18.91)	**(-34.55)**	**(-9.09)**	(-0.69)	**(-7.21)**	**(5.15)**
H 252	83.51	-35.75	-59.62	-2.67	-8.83	-35.25	-79.59	-15.71	11.08	-2.70	11.36
	(18.78)	(17.82)	(18.00)	(17.73)	(17.85)	(17.69)	**(-15.02)**	**(-10.00)**	**(5.34)**	(-1.72)	**(9.85)**

We simulate 10,000 6-year-long log-volatility trajectories using equations (2.9) *to* (2.11). *We estimate the ABD, (data-generating) ABD(2), REGARCH, REGARCH(2), GPSV and GPSV(2) models using the first 5 years of each trajectory. Each model is then used to forecast log-volatility from one day to one year ahead (horizon H, in business days). For each forecast, the NLL, MSE and Bias are computed—equations* (4.2) *to* (4.4)*—and their average over the 10,000 trajectories is reported in the six left-hand side columns of the table. The standard error of the mean is reported in parentheses. The last five columns report paired t-tests results comparing the NLL, MSE and magnitude of the bias, |Bias|, trajectory per trajectory, reporting the difference as, e.g.,* $\frac{1}{10{,}000}\sum_{j=1}^{10{,}000} MSE_H^{ABD,j} - MSE_H^{GPSV,j}$. *Positive entries thus indicate that the model outperforms the benchmark; the t-statistics are in parentheses and are bold whenever the difference is significant at the 5% level.*

Table 3: *Results based on futures data on the E-mini S&P 500.*

		Average and Standard Errors						Paired t-test vs. GPSV(2)				
		ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV
Out-of-sample NLL	**H 1**	0.24	0.12	20.27	7.92	0.11	0.12	-0.12	0.00	-20.15	-7.81	0.00
		(0.06)	(0.03)	(2.20)	(0.99)	(0.03)	(0.03)	**(-3.21)**	(-0.26)	**(-8.49)**	**(-7.10)**	(0.90)
	H 5	4.93	0.39	101.78	36.01	0.33	0.30	-4.63	-0.09	-101.48	-35.71	-0.03
		(0.56)	(0.09)	(6.16)	(2.67)	(0.09)	(0.09)	**(-5.37)**	**(-5.05)**	**(-12.01)**	**(-9.29)**	**(-2.87)**
	H 10	11.05	0.77	203.80	69.71	0.58	0.53	-10.52	-0.24	-203.27	-69.18	-0.05
		(0.96)	(0.14)	(8.69)	(3.81)	(0.14)	(0.13)	**(-5.50)**	**(-5.45)**	**(-13.15)**	**(-9.82)**	**(-3.15)**
	H 21	25.17	1.68	431.84	145.28	1.20	1.12	-24.05	-0.56	-430.72	-144.16	-0.08
		(1.90)	(0.25)	(14.33)	(6.61)	(0.22)	(0.22)	**(-5.50)**	**(-5.51)**	**(-13.65)**	**(-9.97)**	**(-2.98)**
	H 63	78.63	5.16	1282.65	427.40	3.50	3.29	-75.34	-1.87	-1279.35	-424.10	-0.21
		(5.01)	(0.56)	(32.78)	(16.11)	(0.51)	(0.51)	**(-5.96)**	**(-6.63)**	**(-15.76)**	**(-11.06)**	**(-3.72)**
	H 252	335.13	23.01	5109.22	1699.45	14.94	13.81	-321.32	-9.20	-5095.40	-1685.64	-1.12
		(20.14)	(1.60)	(107.85)	(57.32)	(1.34)	(1.32)	**(-6.22)**	**(-7.49)**	**(-18.33)**	**(-11.89)**	**(-8.41)**
Out-of-sample MSE	**H 1**	782.28	730.24	719.25	732.73	730.30	733.29	-48.99	3.05	14.04	0.56	2.99
(in bps)		(60.06)	(58.57)	(56.73)	(55.76)	(58.35)	(57.29)	(-1.66)	(0.35)	**(2.07)**	(0.07)	(0.50)
	H 5	1191.58	863.82	862.16	910.56	865.99	862.46	-329.12	-1.35	0.31	-48.10	-3.52
		(65.29)	(44.34)	(44.15)	(46.81)	(44.49)	(44.55)	**(-5.01)**	(-0.17)	(0.04)	**(-2.13)**	(-0.48)
	H 10	1496.21	991.83	990.00	1047.95	991.06	984.63	-511.59	-7.20	-5.37	-63.33	-6.44
		(79.75)	(44.83)	(44.69)	(48.12)	(44.64)	(44.78)	**(-4.96)**	(-0.69)	(-0.51)	**(-2.50)**	(-0.66)
	H 21	1931.13	1228.64	1220.54	1286.87	1217.50	1196.78	-734.35	-31.86	-23.76	-90.09	-20.72
		(101.15)	(53.30)	(52.85)	(56.34)	(52.12)	(50.46)	**(-4.88)**	(-1.77)	(-1.36)	**(-3.26)**	(-1.27)
	H 63	2577.92	1802.33	1764.77	1759.54	1756.20	1642.12	-935.80	-160.20	-122.65	-117.42	-114.08
		(125.67)	(88.61)	(86.83)	(78.43)	(85.46)	(71.68)	**(-5.18)**	**(-3.17)**	**(-2.62)**	**(-3.54)**	**(-2.65)**
	H 252	3268.38	2897.58	2824.16	2564.92	2799.01	2511.77	-756.61	-385.81	-312.39	-53.15	-287.25
		(102.52)	(95.28)	(95.29)	(78.34)	(90.39)	(72.75)	**(-6.12)**	**(-4.68)**	**(-3.80)**	(-0.94)	**(-3.95)**
Out-of-sample Bias	**H 1**	72.31	450.88	410.94	559.65	415.22	499.04	-84.54	19.02	31.53	-2.37	21.70
(in bps)		(109.50)	(104.35)	(103.79)	(103.72)	(104.58)	(104.23)	(-1.78)	(1.38)	**(2.89)**	(-0.18)	**(2.29)**
	H 5	-969.97	-307.25	-335.00	-205.04	-311.04	-212.28	-431.35	10.74	5.86	-49.00	-0.54
		(103.27)	(86.25)	(86.12)	(89.91)	(86.73)	(86.95)	**(-5.03)**	(0.67)	(0.37)	(-1.88)	(-0.04)
	H 10	-1193.27	-348.23	-386.76	-217.70	-353.84	-225.61	-659.62	3.58	6.51	-69.66	-3.04
		(112.04)	(86.74)	(86.48)	(91.72)	(86.89)	(87.11)	**(-5.83)**	(0.17)	(0.31)	**(-2.13)**	(-0.15)
	H 21	-1375.37	-416.77	-479.09	-212.00	-430.29	-232.99	-891.38	-26.37	-11.86	-97.83	-11.30
		(127.97)	(94.60)	(93.65)	(99.88)	(93.81)	(93.45)	**(-6.31)**	(-0.79)	(-0.36)	**(-2.61)**	(-0.35)
	H 63	-1452.50	-654.25	-777.20	-205.35	-693.38	-287.38	-1108.07	-167.42	-105.30	-153.22	-107.13
		(148.23)	(113.80)	(110.50)	(113.90)	(110.80)	(106.39)	**(-6.73)**	**(-2.92)**	(-1.80)	**(-2.96)**	(-1.96)
	H 252	-1273.43	-970.95	-1131.89	-79.00	-1017.18	-405.53	-883.12	-422.42	-319.49	36.56	-318.06
		(165.23)	(148.69)	(143.31)	(137.09)	(143.25)	(132.76)	**(-6.20)**	**(-4.47)**	**(-3.24)**	(0.47)	**(-3.48)**

On the first week in the sample, model estimation is performed using 5 years of past observations. Then, the estimation sample is rolled week forward, and the process is repeated. A total of 653 weeks of data are considered. Out-of-sample forecasting NLL, MSE and Bias are computed—equations (4.2) *to* (4.4)*—at various horizons (H, in business days). Their average over the 653 weeks is reported in the six left-hand side columns of the table. The standard error of the mean is reported in parentheses. The last five columns report paired t-tests results comparing the NLL, MSE and magnitude of the bias, |Bias|, trajectory per trajectory, reporting the difference as, e.g.,* $\frac{1}{653}\sum_{j=1}^{653} MSE_H^{ABD(2),j} - MSE_H^{GPSV(2),j}$*. Positive entries thus mean that the model outperforms the benchmark. The t-statistics are reported in parenthesis; they are computed using Newey and West (1987) standard errors, and are bold whenever the difference is significant at the 5% level.*

Table 4: *Results with exogenous variables, based on futures data on the E-mini S&P 500.*

		Average and Standard Errors				Paired t-test vs. GPSV(2)			vs. ABD(2)
		GPSV(2)	GPSV(D)	GPSV(Y)	GPSV(D,Y)	GPSV(D)	GPSV(Y)	GPSV(D,Y)	GPSV(D,Y)
Out-of-sample NLL	**H 1**	0.12	0.10	0.12	0.10	0.01	0.00	0.01	0.01
		(0.03)	(0.03)	(0.03)	(0.03)	**(2.98)**	(-0.46)	**(2.36)**	(1.91)
	H 5	0.30	0.31	0.29	0.29	-0.01	0.01	0.01	0.10
		(0.09)	(0.09)	(0.09)	(0.09)	(-0.74)	(0.75)	(1.00)	**(5.40)**
	H 10	0.53	0.54	0.52	0.51	-0.01	0.01	0.02	0.25
		(0.13)	(0.14)	(0.14)	(0.14)	(-0.69)	(0.53)	(1.27)	**(7.07)**
	H 21	1.12	1.12	1.11	1.06	-0.01	0.01	0.05	0.61
		(0.22)	(0.23)	(0.22)	(0.22)	(-0.21)	(0.48)	**(2.37)**	**(9.09)**
	H 63	3.29	3.33	3.29	3.14	-0.03	0.00	0.15	2.02
		(0.51)	(0.51)	(0.51)	(0.51)	(-0.57)	(0.01)	**(2.60)**	**(10.46)**
	H 252	13.81	13.91	13.88	13.21	-0.10	-0.07	0.61	9.80
		(1.32)	(1.34)	(1.37)	(1.33)	(-0.64)	(-0.39)	**(5.05)**	**(10.46)**
Out-of-sample MSE	**H 1**	733.29	711.78	739.99	718.31	21.51	-6.70	14.98	11.93
(in bps)		(57.29)	(58.06)	(56.34)	(58.95)	**(2.49)**	(-0.81)	(1.80)	(1.13)
	H 5	862.46	858.38	853.70	853.02	4.09	8.77	9.44	10.79
		(44.55)	(44.20)	(42.46)	(43.51)	(0.60)	(1.02)	(1.54)	(1.30)
	H 10	984.63	982.12	973.95	974.42	2.50	10.68	10.20	17.41
		(44.78)	(43.46)	(42.08)	(43.40)	(0.25)	(0.75)	(1.31)	(1.48)
	H 21	1196.78	1202.08	1175.27	1180.26	-5.30	21.51	16.52	48.38
		(50.46)	(50.01)	(45.65)	(47.64)	(-0.36)	(1.12)	(1.08)	**(2.63)**
	H 63	1642.12	1713.12	1582.40	1576.08	-71.00	59.72	66.04	226.25
		(71.68)	(81.26)	(61.39)	(62.33)	**(-2.58)**	(1.92)	(1.96)	**(4.82)**
	H 252	2511.77	2682.97	2506.49	2467.42	-171.20	5.28	44.35	430.16
		(72.75)	(86.55)	(66.49)	(67.61)	**(-3.63)**	(0.06)	(0.53)	**(4.23)**
Out-of-sample Bias	**H 1**	499.04	245.54	592.55	361.43	72.64	-25.50	46.95	27.92
(in bps)		(104.23)	(104.04)	(103.98)	(104.00)	**(4.29)**	**(-2.11)**	**(3.18)**	(1.61)
	H 5	-212.28	-294.75	-99.90	-155.06	7.26	-1.93	5.64	-5.10
		(86.95)	(86.27)	(86.52)	(86.43)	(0.58)	(-0.14)	(0.46)	(-0.30)
	H 10	-225.61	-327.84	-85.55	-152.90	-4.82	5.83	8.08	4.50
		(87.11)	(86.41)	(86.58)	(86.58)	(-0.29)	(0.33)	(0.51)	(0.19)
	H 21	-232.99	-383.17	-44.05	-136.48	-12.48	17.77	5.69	32.07
		(93.45)	(93.16)	(92.19)	(92.65)	(-0.60)	(0.65)	(0.28)	(1.03)
	H 63	-287.38	-581.83	35.89	-108.00	-77.01	41.49	20.91	188.33
		(106.39)	(109.08)	(102.98)	(103.03)	**(-2.30)**	(0.90)	(0.54)	**(3.52)**
	H 252	-405.53	-875.62	282.69	108.74	-153.06	89.66	106.61	529.03
		(132.76)	(138.87)	(132.35)	(131.11)	**(-2.63)**	(0.81)	(1.34)	**(3.11)**

These results are obtained following the same procedure as in Table 3. The GPSV(2) model is augmented to account for day-of-the-week effects (D), or for the bond yield, dividend yield, and earnings yield (Y), or for seasonality and exogenous variables (D,Y). Paired t-test are used to compare the three flavors of the augmented GPSV model to the basic GPSV(2) model; see remarks in Table 3 for computation details. The rightmost column also report paired t-tests comparing the augmented model to the ABD(2) benchmark.

Table 5: *GPSV(D,Y): Out-of-sample forecasting MSE ratios.*

Panel A: Full sample, from 1998/09/11 to 2011/03/14

	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)
H 1	1.09	1.02	1.00	1.02	1.02	1.02
H 5	1.40	1.01	1.01	1.07	1.02	1.01
H 10	1.54	1.02	1.02	1.08	1.02	1.01
H 21	1.64	1.04	1.03	1.09	1.03	1.01
H 63	1.64	1.14	1.12	1.12	1.11	1.04
H 252	1.32	1.17	1.14	1.04	1.13	1.02

Panel B: From 1998/09/11 to 2001/11/27

	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)
H 1	1.22	0.98	0.98	0.97	0.99	0.98
H 5	1.83	1.03	1.03	1.00	1.04	1.01
H 10	2.06	1.04	1.05	1.01	1.05	1.03
H 21	2.19	1.08	1.10	1.04	1.09	1.05
H 63	2.04	1.25	1.27	1.13	1.27	1.12
H 252	1.91	1.60	1.60	1.06	1.62	1.18

Panel C: From 2001/12/03 to 2007/11/28

	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)
H 1	1.07	1.03	1.00	1.03	1.02	1.02
H 5	1.17	1.01	1.00	1.09	1.00	1.00
H 10	1.23	1.01	1.00	1.10	1.00	1.00
H 21	1.27	1.03	1.00	1.09	1.00	1.00
H 63	1.30	1.07	1.00	1.11	1.02	1.02
H 252	1.19	1.10	1.00	1.15	1.04	1.03

Panel D: From 2007/12/05 to 2009/06/30

	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)
H 1	0.85	1.05	1.03	1.09	1.05	1.07
H 5	1.09	1.02	1.02	1.14	1.03	1.03
H 10	1.16	1.01	1.02	1.15	1.03	1.02
H 21	1.22	1.02	1.03	1.16	1.02	1.01
H 63	1.30	1.10	1.10	1.11	1.06	1.01
H 252	1.28	1.26	1.29	1.01	1.18	1.06

Panel E: From 2009/07/07 to 2011/03/14

	ABD	ABD(2)	REGARCH	REGARCH(2)	GPSV	GPSV(2)
H 1	1.10	1.03	1.02	1.03	1.03	1.06
H 5	1.42	0.98	0.98	1.07	0.99	1.01
H 10	1.63	1.00	0.99	1.07	0.99	0.99
H 21	1.87	1.03	1.02	1.10	1.00	0.99
H 63	1.91	1.17	1.13	1.11	1.12	1.02
H 252	1.04	0.79	0.77	0.86	0.76	0.77

This table reports, for each alternate model, the ratio of the its out-of-sample forecasting MSE to that of the GPSV(D,Y) model, at various horizons (H).