1 Introduction

The estimation and prediction of price volatility from market data is an important problem in econometrics and finance [Abramov and Klebaner, 2007], as well as practical risk management [Brandt and Santa-Clara, 2006]. The literature on the subject of volatility estimation is vast. Model-based approaches for a single observable asset begin with the ARCH and GARCH models of Engle [1982] and Bollerslev [1986], moving on to stochastic volatility models (see Shephard [2005], for example).

Multivariate equivalents for each of these model classes exist (see Bauwens et al. [2006] and Asai et al. [2006] for reviews of multivariate GARCH and for multivariate stochastic volatility, respectively). However, the majority of work on the subject uses opening and closing prices as data. This approach invariably disregards information traditionally contained in financial timeseries: the observed high and low price of an asset over the quoted periods. To our current knowledge, only Horst et al. [2012] use the observed maximum and minimum of prices in a likelihood to estimate volatility. They do so, however, in a univariate setting.

Explicit model-based approaches in the multivariate setting which take into account extrema over observational periods are completely lacking in the literature, because deriving an efficient approximation of the corresponding likelihood function has hereto been an open problem. In this paper, we use a result addressing this problem and introduce a *bivariate* stochastic volatility model which takes into account the highest and lowest observed prices of each asset as part of a likelihood-based (Bayesian) estimation procedure.

2 Model

The model we will estimate is a bivaraite, 1-factor stochastic volatility model with leverage:

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} x_{t-\Delta} \\ y_{t-\Delta} \end{pmatrix} + \begin{pmatrix} \mu_x \Delta \\ \mu_y \Delta \end{pmatrix} + \begin{pmatrix} \sqrt{1-\rho_t^2}\sigma_{x,t} & \rho_t \sigma_{x,t} \\ 0 & \sigma_{y,t} \end{pmatrix} \begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{y,t} \end{pmatrix}, \tag{1}$$

$$\inf_{[t-\Delta,t]} x_\tau = a_{x,t} \qquad \sup_{[t-\Delta,t]} x_\tau = b_{x,t} \qquad \inf_{[t-\Delta,t]} y_\tau = a_{y,t} \qquad \sup_{[t-\Delta,t]} y_\tau = b_{y,t}$$

$$\log(\sigma_{x,t+\Delta}) = \alpha_x + \theta_x(\log(\sigma_{x,t}) - \alpha_x) + \tau_x \eta_{x,t}, \tag{2}$$

$$\log(\sigma_{y,t+\Delta}) = \alpha_y + \theta_y(\log(\sigma_{y,t}) - \alpha_y) + \tau_y \eta_{y,t}, \tag{3}$$

$$\operatorname{logit}((\rho_{t+\Delta}+1)/2) = \alpha_\rho + \theta_\rho\left(\operatorname{logit}((\rho_t+1)/2) - \alpha_\rho\right) + \tau_\rho \eta_{\rho,t}. \tag{4}$$

The marginal distribution for all of the innovation terms $\varepsilon_{x,t}, \varepsilon_{y,t}, \eta_{x,t}, \eta_{y,t}, \eta_{\rho,t}$ is the standard Gaussian distribution. The *leverage* terms are defined as $E[\varepsilon_{x,t}\,\eta_{x,t}] = \rho_x$ and $E[\varepsilon_{x,t}\,\eta_{x,t}] = \rho_y$. It should be noted here that we are explicitly allowing the correlation of the process to change over time in a mean-reverting fashion. Finally, we explicitly write down the realized extrema over the periods $[t-\Delta,t]$ to be included as data into the likelihood for the dynamical model. We estimate all parameters and dynamical factors in a fully Bayes framework via the augmented particle filter of Liu and West [2001] which we will describe below.

2.1 Likelihood for the observables

Each period $[t-\Delta,t]$ has six associated observables: opening coordinate $(x_{t-\Delta}, y_{t-\Delta})$, closing coordinate (x_t, y_t), and the observed extrema in each nominal direction $(a_{x,t}, b_{x,t}), (a_{y,t}, b_{y,t})$. Given the evolution model in (1), disregarding the information contained in the extrema yields the usual bivariate Gaussian density in terms of the volatility parameters and the state of the process at time $t-\Delta$:

$$p(x_t, y_t | x_{t-\Delta}, y_{t-\Delta}, \mu_x, \mu_y, \sigma_{x,t}, \sigma_{y,t}, \rho_t) =$$

$$\frac{1}{2\pi\,\Delta\,\sigma_{x,t}\sigma_{y,t}\sqrt{1-\rho_t^2}} \exp\left\{ -\frac{1}{2\Delta(1-\rho_t^2)} \left(\frac{(x_t - x_{t-\Delta})^2}{\sigma_{x,t}^2} - 2\rho_t \frac{(x_t - x_{t-\Delta})(y_t - y_{t-\Delta})}{\sigma_{x,t}\sigma_{y,t}} + \frac{(y_t - y_{t-\Delta})^2}{\sigma_{y,t}^2} \right) \right\}.$$

Incorporating the extreme values over $[t-\Delta, t]$ is accomplished by considering the Fokker-Planck Equation for the forward, continuous-time evolution of the probability density function of (x_t, y_t) and including $(a_{x,t}, b_{x,t}), (a_{y,t}, b_{y,t})$ as boundary conditions where the density is zero. The previous work describes the method by which the full likelihood function is found. However, for the purposes of the particle filter used to estimate the model in this work, we improve upon the computational method by performing a set of normalizing transformations, allowing for more flexibility in the parameters for the basis functions in the Galerkin approximation, and extrapolating over certain low-probability regions of the parameter space.

2.2 Improved likelihood computational method for low probability data

Our method for approximating the likelihood needs to be able to accomodate proposed parameters for given data values in low-likelihood regions. To ensure this property, the previous method used to approximate the likelihood is modified in the following ways:

1. First, when computing the numerical derivative of the Galerkin solution, we use finite difference steps *proportional* to the range of the process in each direction

$$\Delta x = h \cdot L_x, \tag{5}$$
$$\Delta y = h \cdot L_y, \tag{6}$$

   where h fixed. In this way, the discretization step sizes in the normalized problem are constant and equal to h, having the same effect on the order accuracy accross any combination of L_x and L_y. This is in contrast to our previous practice of using h directly on the unnormalized problem. In cases where the observed ranges L_x and L_y are close to unity this approach is acceptable. However, it fails for "fat" or "skinny" computational domains.

2. Given the diffusion parameters (τ_x, τ_y, ρ) for the normalized problem solved by our numerical method, we assume, without loss of generality, that $\tau_x \geq \tau_y$. We introduce another transformation of the time variable:

$$\tilde{t} = t\tau_x^2.$$

   The new normalized diffusion problem then becomes

$$\frac{\partial p}{\partial \tilde{t}} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2} + \rho\left(\frac{\tau_y}{\tau_x}\right)\frac{\partial^2 p}{\partial x \partial y} + \frac{1}{2}\left(\frac{\tau_y}{\tau_x}\right)^2 \frac{\partial^2 p}{\partial y^2}, \tag{7}$$

   where the domain is still the unit square and we find the solution $p(x, y, \tau_x^2)$. This transformation reorganizes the basis expansion so that, conditional on the correlation term ρ, the quality of the solution is goverend by the resolution in the principal $x-$direction, which requires fewer terms because of the greater diffusion coefficient, and the resolution in the $y-$direction, which requires more terms because of the smaller diffusion coefficient. This motivatives the next augmentation of the computational method, which introduces a more general basis element type.

3. We change the form of the basis functions in the Galerkin solution to account for different resolutions in the principal $x-$ and $y-$ directions. The basis elements in (17) of Chapter 2 then become

$$\psi_i(x, y) = \frac{1}{2\pi\tilde{\sigma}_x\tilde{\sigma}_y\sqrt{1-\tilde{\rho}^2}} \exp\left\{-\frac{1}{2(1-\tilde{\rho}^2)}\left(\frac{(x-x_i)^2}{\tilde{\sigma}_x^2} - 2\tilde{\rho}\frac{(x-x_i)(y-y_i)}{\tilde{\sigma}_x\tilde{\sigma}_y} + \frac{(y-y_i)^2}{\tilde{\sigma}_y^2}\right)\right\} \tag{8}$$

   for some parameters $(\tilde{\rho}, \tilde{\sigma}_x, \tilde{\sigma}_y)$ which together control the resolution of the solution in the two principal directions. The scheme used to set the node points over the computational domain remains the same.

Any choice of basis-generating parameters $(\tilde{\rho}, \tilde{\sigma}_x, \tilde{\sigma}_y)$ produces a finite number of basis elements, meaning that there are lower bounds for τ_y/τ_x and $\tilde{t}$ which admit a valid solution. The new normalized problem (7) conveniently provides a computational method for identifying such lower bounds. By generating varying diffusion parameter combinations and observing the joint distribution of the normalized parameters, we can identify extreme combinations of $(\tilde{t}, \tau_y/\tau_x)$ parameter values. To do so, we sample the original diffusion parameters σ_x and σ_y independently from a heavy-tailed distribution in order to capture cases where there is several order magnitude difference between σ_x and σ_y. For our purposes, we sample

$$\log(\sigma_x) \overset{\text{i.i.d.}}{\sim} N(1,1), \qquad \log(\sigma_y) \overset{\text{i.i.d.}}{\sim} N(1,1), \tag{9}$$

$5,000$ times then, for each pair of sampled values, (1) is solved with a forward Euler discretization (step $\Delta = 1 \cdot 10^{-6}$); then the parameters are normalized as in (7). The samples for $\log(\tau_y/\tau_x)$ and $\log(\tilde{t})$ are shown in Figure (3). Across the three regimes of ρ, extreme values for the normalized parameters are defined by having either the smallest τ_y/τ_x or $\tilde{t}$ in the sample produced. These points are denoted by the the red and green points in Figure (3) and are candidates for having a likelihood function which is greatest at those extreme values and may be most problematic to resolve with respect to our Galerkin solution. An intersting observation is that there is a tradeoff between the size of τ_y/τ_x and $\tilde{t}$. In other words, it is very unlikely to encounter data points generated by a diffusion equation having $(\tau_y/\tau_x, \tilde{t})$ in the lower-left corner of the parameter space.

The quality of the Galerkin solution in the $\rho \neq 0$ cases cannot be directly assessed because of a lack of a closed-form analytic solution. However, this is not the case for $\rho = 0$. Figure (1) features both the analytic and Galerkin likelihood surfaces for the small τ_y/τ_x (green) point in Figure (3). The left panel shows the true analytic solution. The middle panel shows the Galerkin solution, where lack of color signifies a negative value produces for the likelihood. This is automatically an inadmissible solution. Finally, the right panel shows the Galerkin solution where likelihood values differing from the true solution by more than 5% have been discarded. We see that the Galerking solution fails in cases where the likelihood function is small or $\tilde{t}$ is small. However, the Galerkin solution can be used to reasonably reconstruct the entire likelihood surface. Figure (2) similarly shows the likelihood surface for the small $\tilde{t}$ (red) point in Figure (3). We see that the choice of basis parameters

$$\tilde{\sigma}_x = 0.3 \qquad \tilde{\sigma}_y = 0.1 \qquad \tilde{\rho} = 0.0, \tag{10}$$

for the $\rho = 0$ case is capable of resolving the main probability mass in the likelihood function for both extreme cases considered. A hard cutoff for the Galerkin solution for $\tilde{t} \geq 0.25$ and $\tau_y/\tau_x \geq 0.4$ is sufficient to capture the main probability mass of the likelihood surface. This was confirmed by computing the surface for all other samples in the middle panel of Figure (3).

Finally, Figure (4) shows the likelihood surfaces for the extreme parameter combinations where $\rho = 0.9$, with parameters used for the basis generation being

$$\tilde{\sigma}_x = 0.3 \qquad \tilde{\sigma}_y = 0.1 \qquad \tilde{\rho} = 0.8. \tag{11}$$

The solution generated has properties consisent with those of the analytic $\rho = 0$ solution: conditional on either τ_y/τ_x or $\tilde{t}$, the solution formed, when valid, exibits unimodality and constant concavity around the mode. As such, we can conclude that we have reasonably resolved the likelihood in its important regions.

So far we have demonstrated that a choice of basis function parameters can reasonably resolve the main probability mass in the likelihood function over the parameter space of the normalized problem. However, the solution does explicitly fail away from the main probability mass of the likelihood. To evaluate the likelihood in such conditions, we introduce an analytic extension of our numerical solution.

Figure 1: .

Figure 2

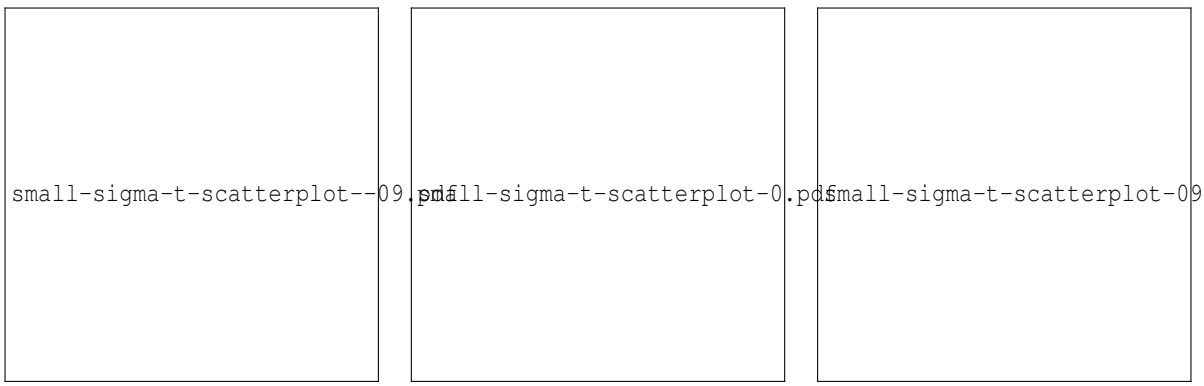

Figure 3: .

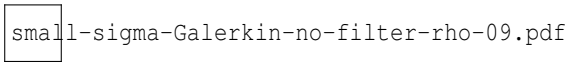

Figure 4: .

4. To motivate the extension of our numerical method, we consider the analytic solution for the problem when $\tilde{t} << 1$. We have already introduced such a small-time solution in the previous chapter and only modify it by considering a system of images defined not by a single reflection, but by a reflection about each of the four boundaries. This produces a system of 16 images of the fundamental solution to the heat equation in (7), which is a bivariate Gaussian with the given diffusion parameters. We assume that the small-time $\tilde{t}_\varepsilon$ is small enough such that the boundary conditions are enforced numerically, which is the case for the previously defined $\tilde{t}_\varepsilon$. This produces a solution to the governing equation explicitly differentiable with respect to the boundaries and having the form

$$p(x,y,\tilde{t}_\varepsilon) = \phi(x,y|x_0,y_0,1,\tau_y/\tau_x,\rho) + \sum_{k'=1}^{K} w_{k'}\phi(x,y|x_{k'},y_{k'},1,\tau_y/\tau_x,\rho), \tag{12}$$

where $\phi(x,y|x_k,y_k,1,\tau_y/\tau_x,\rho)$ is the bivariate Gaussian

$$\phi(x,y|x_k,y_k,1,\tau_y/\tau_x,\rho) = \frac{1}{2\pi\tilde{t}_\varepsilon\tau_y/\tau_x\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)\tilde{t}_\varepsilon}\left(\frac{(x-x_k)^2}{1} - 2\rho\frac{(x-x_k)(y-y_k)}{\tau_y/\tau_x} + \frac{(y-y_k)^2}{\tau_y^2/\tau_x^2}\right)\right\}, \tag{13}$$

$K = 15$, (x_0,y_0) is the initial condition for the problem, and w_k is either 1 or -1. In the images solution, only the location parameters for ϕ are functions of the boundaries, and only one of the images has location parameters (x_k,y_k) as functions of all four boundaries. Thus, upon differentiation, the small-time likelihood becomes

$$\frac{\partial^4}{\partial a_x\partial b_x\partial a_y\partial b_y}p(x,y,\tilde{t}_\varepsilon) =$$
$$\frac{\partial^4}{\partial a_x\partial b_x\partial a_y\partial b_y}\left(\frac{1}{2\pi\tilde{t}_\varepsilon\tau_y/\tau_x\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)\tilde{t}_\varepsilon}\left(\frac{(x-x_k)^2}{1} - 2\rho\frac{(x-x_k)(y-y_k)}{\tau_y/\tau_x} + \frac{(y-y_k)^2}{\tau_y^2/\tau_x^2}\right)\right\}\right). \tag{14}$$

Viewing the likelihood as a function of $\tilde{t}_\varepsilon$ and τ_y^2/τ_x^2 and carrying out the differentiation yields a likelihood proportional to an inverse Gamma-like probability distribution

$$L(\tilde{t}_\varepsilon,\tau_y^2/\tau_x^2|\rho)$$
$$\propto \frac{1}{\tilde{t}_\varepsilon^5}\left(\frac{1}{(\tau_y^2/\tau_x^2)^{9/2}} + C_1\frac{1}{(\tau_y^2/\tau_x^2)^{8/2}} + \cdots + C_8\frac{1}{(\tau_y^2/\tau_x^2)^{1/2}}\right)\exp\left\{-\frac{1}{2(1-\rho^2)\tilde{t}_\varepsilon}\left(\frac{(x-x_k)^2}{1} - 2\rho\frac{(x-x_k)(y-y_k)}{\tau_y/\tau_x} + \frac{(y-y_k)^2}{\tau_y^2/\tau_x^2}\right)\right\}. \tag{15}$$

This further reduces in the $\rho = 0$ case, where

$$L(\tilde{t}_\varepsilon,\tau_y^2/\tau_x^2|\rho=0) \propto \frac{1}{\tilde{t}_\varepsilon^5}\left(\frac{1}{(\tau_y^2/\tau_x^2)^{9/2}}\right)\exp\left\{-\frac{1}{2\tilde{t}_\varepsilon}\left(\frac{(x-x_k)^2}{1} + \frac{(y-y_k)^2}{\tau_y^2/\tau_x^2}\right)\right\}. \tag{16}$$

Since the method of images is valid for $\rho = 0$, the likelihood is proportional to a linear combination of inverse Gamma distributions when considering a general $\tilde{t}$.

Going back to the more general case in (15) as a function of $\tilde{t}_\varepsilon$ alone, we have

$$L(\tilde{t}_\varepsilon|\tau_y^2/\tau_x^2,\rho) \propto \frac{1}{\tilde{t}_\varepsilon^5}\exp\left\{-\frac{1}{2(1-\rho^2)\tilde{t}_\varepsilon}\left(\frac{(x-x_k)^2}{1} - 2\rho\frac{(x-x_k)(y-y_k)}{\tau_y/\tau_x} + \frac{(y-y_k)^2}{\tau_y^2/\tau_x^2}\right)\right\}$$
$$\propto IG\left(\tilde{t}_\varepsilon|\alpha=4,\beta = \frac{1}{2(1-\rho^2)}\left(\frac{(x-x_k)^2}{1} - 2\rho\frac{(x-x_k)(y-y_k)}{\tau_y/\tau_x} + \frac{(y-y_k)^2}{\tau_y^2/\tau_x^2}\right)\right)$$

Since the image k is reflected once about each boundary, both $(x-x_k),(y-y_k) \geq 1$ and combining with the bound $\tau_y/\tau_x \leq 1$,

$$\beta \geq \frac{1}{2(1-\rho^2)}\left(1-2\rho\frac{1}{\tau_y/\tau_x}+\frac{1}{\tau_y^2/\tau_x^2}\right) \geq \frac{3}{2}.$$

Then the mode for the inverse Gamma-like likelihood, conditional on τ_y/τ_x and ρ, is bounded below by

$$\frac{\beta}{\alpha+1} \geq 0.3.$$

The $\tilde{t}$ likelihood approximately follows an inverse Gamma distribution with mode greater than or equal to about $1/3$ within the small-time regions of the parameter space. This functional relationship changes outside the small-time region, but it still dominates the likelihood. Although not true in general, increasing $\tilde{t}$ away from the small-time region can be accomodated with additional reflections of the system of images about the boundaries. The new images add to the likelihood for $\tilde{t}$, but their contribution decays exponentially compared the nearest term to the computational domain (which is image k in (15)) since they scale away linearly as a function of the number of reflections.

A similar analysis can be performed for $(\tau_x/\tau_y)^2$, where $\alpha \leq 7/2$ and

$$\beta \geq \frac{(y-y_k)}{2(1-\rho^2)\tilde{t}_\varepsilon} \geq \frac{1}{2\tilde{t}_\varepsilon}.$$

Hence, the mode for the likelihood as a function of τ_y/τ_x conditional on ρ is bounded by

$$\sqrt{\frac{\beta}{\alpha+1}} \geq \sqrt{\frac{1}{7\tilde{t}_\varepsilon}} \geq 0.33.$$

The above derived bounds for where the main probability mass of the likelihood abides within the parameter space corroborates our empirical observations, and they also further support our choice for basis parameters and their capability of resolving the main parts of the likelihood surface.

However, the parametric approximations derived motivate the inverse Gamma approximation we use to evaulate the likelihood in inadmissible regions of the parameter space. We extend into either the small $\tilde{t}$ or small τ_y/τ_x regions by fitting the function

$$f(x) = Cx^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right)$$

using three valid likelihood points with bigger $\tilde{t}$ or τ_y/τ_x values. For points falling in the small parameter region for both $\tilde{t}$ and τ_y/τ_x, we first approximate likelihood values into one then the other regions. Although more computationally intensive, this correction allows us to validly evaulate likelihoods in the parameter space for the normalized problem that cannot be resolved by our Galerkin solution.

3 Particle Filtering

Bayesian approaches for estimating complex hierarchical models like those in (1) - (4) are common. The primary focus of such inferential algorithms is the generation of dependent samples from the posterior distributions of model parameters and unobservable latent trajectories via Markov Chain Monte Carlo (MCMC). However, the highly non-linear nature of the state-space model considered here prohibits conditional Gaussian representations and thereby limits the application of MCMC methods. Instead, we use a particle filtering approach, which is a particular version of a sequential Monte Carlo algorithm (see **?** for an overview), to generate samples from the posterior distribution of the collection of time-dependent parameters, which we abbreviate to

$$\sigma_t := (\sigma_{x,t}, \sigma_{y,t}, \rho_t),$$

as well as all of the time-constant parameters governing the evolution of the process, which we transform to be on the $(-\infty, \infty)$ scale and denote

$$\phi := (\alpha_x, \alpha_y, \alpha_\rho, \theta_x, \theta_y, \theta_\rho, \log(\tau_x), \log(\tau_y), \log(\tau_\rho), \text{logit}(\rho_x), \text{logit}(\rho_y)).$$

Most sequential Monte Carlo algorithms assume the structural parameters ϕ to be known and fixed, which is not the case for the considered problem. We therefore follow Horst et al. [2012] and use a version of the augmented particle filter of Pitt and Shephard [1999] developed by Liu and West [2001] to sample from the full posterior.

Particle filters use a discrete mixture to represent the posterior distribution $p(\sigma_t, \phi|\mathcal{D}_t)$, where $\mathcal{D}_t$ represents all of the observable information up to time t:

$$\mathcal{D}_t = (x_0, y_0, a_{x,\Delta}, b_{x,\Delta}, a_{y,\Delta}, b_{y,\Delta}, x_\Delta, y_\Delta, \ldots, x_{t-\Delta}, y_{t-\Delta}, a_{x,t}, b_{x,t}, a_{y,t}, b_{y,t}, x_t, y_t).$$

In this mixture approximation, the information about the parameters given the current data is captured in the parameter values and weights associated with each particle:

$$p(\sigma_t, \phi|\mathcal{D}_t) \approx \sum_{k=1}^{K} \delta(\sigma_t^{(k)} - \sigma_t)\delta(\phi^{(k)} - \phi)w_t^{(k)}$$

Given this approximation of $p(\sigma_t, \phi|\mathcal{D}_t)$, additional information at time $t+\Delta$ is incorporated by updating each particle weight and parameter values via an appropriately chosen importance sampling distribution and Bayes' Theorem. In the case of the augmented particle filter of Pitt and Shephard [1999], which treats the structural parameters ϕ are known and fixed, the approximate distribution of the state of the system at time t is

$$p(\sigma_t|\phi, \mathcal{D}_t) \approx \sum_{k=1}^{K} \delta(\sigma_t^{(k)} - \sigma_t)w_t^{(k)};$$

the posterior $p(\sigma_{t+\Delta}|\phi, \mathcal{D}_{t+\Delta})$ is *augmented* with the particle index k and sampled with Bayes' Theorem:

$$p(\sigma_{t+\Delta}, k|\phi, \mathcal{D}_{t+\Delta}) \propto p(x_{t+\Delta}, y_{t+\Delta}, a_{x,t+\Delta}, b_{x,t+\Delta}, a_{y,t+\Delta}, b_{y,t+\Delta}|\sigma_{t+\Delta}, k, \phi, \mathcal{D}_t)p(\sigma_{t+\Delta}, k|\phi, \mathcal{D}_t), \tag{17}$$

$$= p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}, \phi, \mathcal{D}_t)p(\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi)w_t^{(k)}. \tag{18}$$

Pitt and Shephard [1999] sample the joint posterior with a proposal distribution which replaces $\sigma_{t+\Delta}$ with a representative value in the one-step predictive distribution $p(\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi)$, such as the predictive mean. In this way k is sampled then $\sigma_{t+\Delta}$ from the predictive distribution conditional on k. The new weight $w_{t+\Delta}^{(k)}$ is proportional to the ratio of the sampled $\sigma_{t+\Delta}$ and representative value in the likelihood for the data.

Liu and West [2001] extend the augemented particle filter of Pitt and Shephard [1999] to allow for the estimation of the constant structural parameter ϕ. This is done by introducing an artificial evolution for the fixed parameter, which is now indexed by t, and defining the transition density

$$p(\phi_{t+\Delta}|\mathcal{D}_t) \approx \sum_{k=1}^{K} N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t),$$

where a is a discount factor between 0 and 1, $\bar{\phi}_t$ is the average of samples for ϕ at t and V_t is the respective sample covariance at time t. The Gaussian perturbation of ϕ_t is appropriate, as we have transformed each of the structural parameters to $\mathbb{R}$, while the shrinkage kernel approximation preserves the mean and covariance structure of the posterior distribution from t to $t+\Delta$. This limits the injection of entropy into the system that would otherwise occur with simpler, conditionally independent Gaussian perturbations.

The augmented posterior then becomes

$$p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) \propto p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}, \phi_{t+\Delta}, \mathcal{D}_t)p(\sigma_{t+\Delta}|\phi_{t+\Delta}, \sigma_t^{(k)}, \mathcal{D}_t)N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t)w_t^{(k)}. \tag{19}$$

Sampling from the posterior (19) is also done with via a proposal where $(\sigma_{t+\Delta}, \phi_{t+\Delta})$ are replaced with predictive means conditional on k in the likelihood:

$$p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) = p(\mathbf{y}_{t+\Delta}|E[\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi_t^{(k)}], a\phi_t^{(k)} + (1-a)\bar{\phi}_t) \\ \times p(\sigma_{t+\Delta}|\phi_{t+\Delta}^{(k)}, \sigma_t^{(k)}) N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t) w_t^{(k)}. \quad (20)$$

With (20), we can integrate out $(\sigma_{t+\Delta}, \phi_{t+\Delta})$ to propose k, then sample the two remaining parameters respectively. The new weights are computed as the likelihood ratio of the likelihood function. If we denote the predictive means to be

$$m_{t+\Delta}^{(k)} := E[\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi_t^{(k)}], \quad (21)$$

$$\mu_{t+\Delta}^{(k)} := E[\phi_{t+\Delta}|k, \mathcal{D}_t] = a\phi_t^{(k)} + (1-a)\bar{\phi}_t, \quad (22)$$

our proposal distribution can be written as

$$p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) = p(\mathbf{y}_{t+\Delta}|m_{t+\Delta}^{(k)}, \mu_{t+\Delta}^{(k)}) \\ \times p(\sigma_{t+\Delta}|\phi_{t+\Delta}^{(k)}, \sigma_t^{(k)}) N(\phi_{t+\Delta}|\mu_{t+\Delta}^{(k)}, (1-a^2)^2 V_t) w_t^{(k)}. \quad (23)$$

The steps in the particle filtering scheme are explicitly

1) Sample the particle indicator k from the marginal proposal

$$p(k|\mathcal{D}_{t+\Delta}) = \int_{\sigma_{t+\Delta}} \int_{\phi_{t+\Delta}} p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) d\sigma_{t+\Delta} d\phi_{t+\Delta} \propto p(\mathbf{y}_{t+\Delta}|m_{t+\Delta}^{(k)}, \mu_{t+\Delta}^{(k)}) w_t^{(k)}.$$

2) Conditional on k, sample $(\sigma_{t+\Delta}, \phi_{t+\Delta})$ from

$$\phi_{t+\Delta}^{(k)} \sim N(\phi_{t+\Delta}|\mu_{t+\Delta}^{(k)}, (1-a^2)^2 V_t), \\ \sigma_{t+\Delta}^{(k)} \sim p(\sigma_{t+\Delta}|\phi_{t+\Delta}^{(k)}, \sigma_t^{(k)}).$$

3) Compute the new weight as the ratio of likelihoods:

$$w_{t+\Delta}^{(k)} \propto \frac{p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}^{(k)}, \phi_{t+\Delta}^{(k)})}{p(\mathbf{y}_{t+\Delta}|m_{t+\Delta}^{(k)}, \mu_{t+\Delta}^{(k)})}$$

Steps 1) and 3) require a reliable way to compute the likelihood for the observed process. This is particularly true in instances where proposed and predictive values for the parameters are very unlikely given the observation $\mathbf{y}_{t+\Delta}$. The finite precision of the method we developed to compute the likelihood, mainly arising from the truncation of the basis expansion of the solution, may yield negative values for the likelihood. The ad-hoc correction of replacing such values with numerically zero but positive values can yield deleterious results, as what would otherwise be a small value for $w_{t+\Delta}^{(k)}$ may be replaced with unity in the case where the likelihood computation fails when evaluating both $p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}^{(k)}, \phi_{t+\Delta}^{(k)})$ and $p(\mathbf{y}_{t+\Delta}|m_{t+\Delta}^{(k)}, \mu_{t+\Delta}^{(k)})$. To deal with this numerical instability, we introduce a series of improvements to our likelihood computation method which a) increase the resolution of the expansion for the fixed number of basis elements, and b) perform a 1st order approximation in instances where the likelihood is negative. The details are described in Section 2.2.

4 Calibration Study

In this calibration study, we simulate from the model with parameters corresponding to the continuous-time version of the model, where we pick the parameters as we did in the first paper.

5 Application

I will need a bit more guidance here.

References

Vyacheslav M Abramov and Fima C Klebaner. Estimation and prediction of a non-constant volatility. *Asia-Pacific Financial Markets*, 14(1):1–23, 2007.

Manabu Asai, Michael McAleer, and Jun Yu. Multivariate stochastic volatility: a review. *Econometric Reviews*, 25(2-3):145–175, 2006.

Luc Bauwens, Sébastien Laurent, and Jeroen VK Rombouts. Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109, 2006.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3): 307–327, 1986.

Michael W Brandt and Pedro Santa-Clara. Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61(5):2187–2217, 2006.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.

Enrique Ter Horst, Abel Rodriguez, Henryk Gzyl, and German Molina. Stochastic volatility models including open, close, high and low prices. *Quantitative Finance*, 12(2):199–212, 2012.

Jane Liu and Mike West. Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice*, pages 197–223. Springer, 2001.

Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

Neil Shephard. *Stochastic Volatility: Selected Readings*. Oxford University Press, UK, 2005.