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.

   The new normalized problem (7) also provides an easy computational method for identifying tolerances of parameters governing the solution over problem likelihood regions which are intuitively

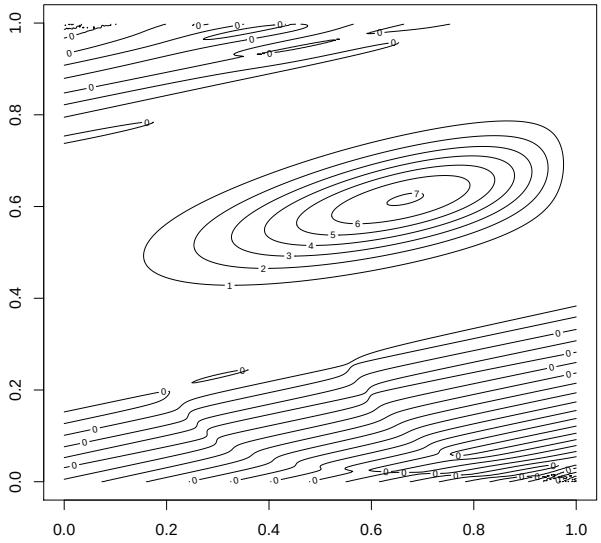

Figure 1: .

associated with $\tau_y/\tau_x << 1$ and $\tilde{t} << 1$. The order magnitude of what constitute small values for these parameters can be explored by sampling the parameters in the following way. Sample the diffusion parameters σ_x and σ_y independently from a heavy-tailed distribution. For our purposes, we sample

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

$100,000$ time then, for each pair of sampled values, (1) is solved with a forward Euler discretization (step $\Delta = 1 \cdot 10^{-6}$) where the extreme values are approximated by the obtained extreme, discrete-time values of the process. Then the parameters are normalized as in (7). The samples for $\log(\tau_y/\tau_x)$ and $\log(\tilde{t})$ are shown in Figure (2). Across the three regimes of ρ, likely values for the normalized parameters are roughly bounded below by $\tilde{t} > 0.08$ and $\tau_y/\tau_x > 0.3$. Hence, if our numerical method can accommodate $\tau_y/\tau_x \approx 0.3$ with $\tilde{t} \approx 0.08$, then it will fail to produce a valid likelihood once in 100,000 times. Here, accommodation is defined by 1) attenuating high-frequency modes of the solution by terminal time, and 2) producing positive likelihood values across the range of initial and final conditions. Firgure (1) shows a countour plot of a solution for a particular initial condition where $\rho = 0.7$, $\tau_y/\tau_x = 0.35$, and $\tilde{t} = 0.08$, which satisfies 1) above. The parameters for the solution are

$$\tilde{\sigma}_x = 0.3 \qquad \tilde{\sigma}_y = 0.08 \qquad \tilde{\rho} = 0.8. \tag{10}$$

4. The finite constraints imposed by the implementation of our numerical solution means that there may be unlikely data points which produce a negative likelihood, even after the above precautions have been placed. When performing posterior inference with particle filters, in such instances it is entirely insufficient to replace negative likelihoods with arbitrary small, but positive, constants.

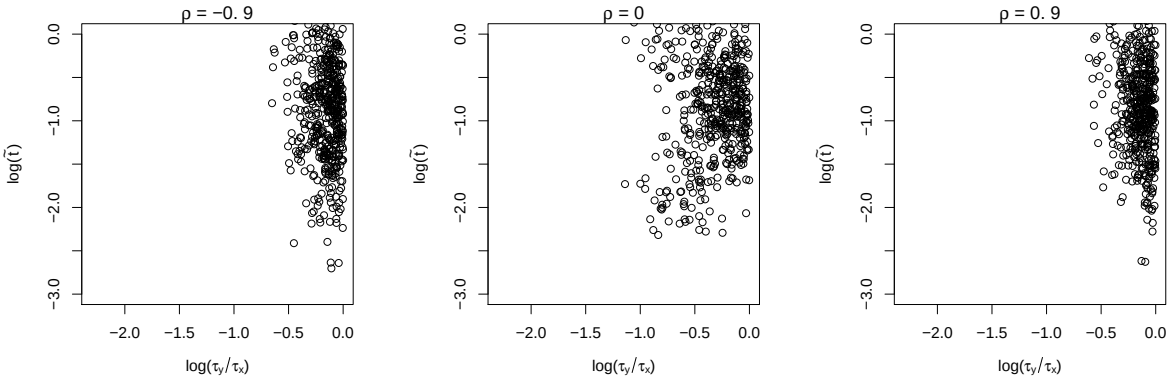

Figure 2: .

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 Doucet et al. [2001] 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:

$$p(\sigma_{t+\Delta}, k|\phi, \mathcal{D}_{t+\Delta}) \propto p(x_{t+\Delta}, y_{t+\Delta}, a_{x,t+\Delta}, b_{x,t+\Delta}, a_{y,t+\Delta}, b_{y,t+\Delta}|\sigma_{t+\Delta}, k, \phi, \mathcal{D}_t) p(\sigma_{t+\Delta}, k|\phi, \mathcal{D}_t), \tag{11}$$

$$= p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}, \phi, \mathcal{D}_t) p(\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi) w_t^{(k)}. \tag{12}$$

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

$$p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) \propto p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}, \phi_{t+\Delta}, \mathcal{D}_t) p(\sigma_{t+\Delta}|\phi_{t+\Delta}, \sigma_t^{(k)}, \mathcal{D}_t) N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t) w_t^{(k)}. \tag{13}$$

Sampling from the posterior (13) 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:

$$\begin{aligned} p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) &= p(\mathbf{y}_{t+\Delta}|E[\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi_t^{(k)}], a\phi_t^{(k)} + (1-a)\bar{\phi}_t) \\ &\quad \times p(\sigma_{t+\Delta}|\phi_{t+\Delta}^{(k)}, \sigma_t^{(k)}) N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t) w_t^{(k)}. \end{aligned} \tag{14}$$

With (14), 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

$$m_{t+\Delta}^{(k)} := E[\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi_t^{(k)}], \tag{15}$$

$$\mu_{t+\Delta}^{(k)} := E[\phi_{t+\Delta}|k, \mathcal{D}_t] = a\phi_t^{(k)} + (1-a)\bar{\phi}_t, \tag{16}$$

our proposal distribution can be written as

$$\begin{aligned} p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) &= p(\mathbf{y}_{t+\Delta}|m_{t+\Delta}^{(k)}, \mu_{t+\Delta}^{(k)}) \\ &\quad \times p(\sigma_{t+\Delta}|\phi_{t+\Delta}^{(k)}, \sigma_t^{(k)}) N(\phi_{t+\Delta}|\mu_{t+\Delta}^{(k)}, (1-a^2)^2 V_t) w_t^{(k)}. \end{aligned} \tag{17}$$

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.

Arnaud Doucet, Nando De Freitas, and NJ Gordon. Sequential monte carlo methods in practice. series statistics for engineering and information science, 2001.

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.