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

The method we use to estimate the model (1) - (4) is essentially comprised of a sequence of importance samples whose weight in the posterior is a function of the ratio of the likelihood function evaluated at different parameter values. For this reason, 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 way:

1. When computing the numerical derivative of the Galerkin solution, perturb the computational boundaries on an already normalized domain. This requires an additional change of paramters for the likelihood.

2. Perform either a $\pi/2$ or -$\pi/2$ rotation so that the greater normalized diffusion parameter corresponds to the same principal direction. Then perform a scaling of the diffusion timescale so that the greater diffusion parameter is always unity.

3. Given the constant diffusion parameter in a single direction, we can use different resolution with the basis functions in each principal direction. In this way we can control separately the slow and fast diffusion phenomena for the process.

4. In the cases where the likelihood is invalid (negative), perform an extrapolation from a parameter combination that does give a valid likelihood approximation.

3 Estimation Methodology

Given the highly non-linear hierarchical model (1) - (4) and the non-Gaussian observational likelihood, we use a particle filter to estimate 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

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

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:

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

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

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

the posterior $p(\sigma_{t+\Delta}|\phi, \mathcal{D}_{t+\Delta})$ is *augmented* with the particle index k and sampled:

$$p(\sigma_{t+\Delta}, k|\phi, \mathcal{D}_{t+\Delta}) \propto p(x_{t+\Delta}, y_{t+\Delta}, a_{x,t+\Delta}, b_{x,t+\Delta}, a_{y,t+\Delta}, b_{y,t+\Delta}|\sigma_{t+\Delta}, k, \phi, \mathcal{D}_t)p(\sigma_{t+\Delta}, k|\phi, \mathcal{D}_t), \tag{5}$$

$$= p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}, \phi, \mathcal{D}_t)p(\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi)\omega_t^{(k)}. \tag{6}$$

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 $\omega_{t+\Delta}^{(k)}$ 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, indexing it 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 augmented posterior becomes

$$p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) \propto p(\mathbf{y}_{t+\Delta}|\sigma_{t+\Delta}, \phi_{t+\Delta}, \mathcal{D}_t)p(\sigma_{t+\Delta}|\phi_{t+\Delta}, \sigma_t^{(k)}, \mathcal{D}_t)N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t)\omega_t^{(k)}. \tag{7}$$

Sampling from the posterior (7) is also done with via a proposal where $(\sigma_{t+\Delta}, \phi_{t+\Delta})$ are replaced with predictive means conditional on k:

$$\begin{aligned} p(\sigma_{t+\Delta}, \phi_{t+\Delta}, k|\mathcal{D}_{t+\Delta}) \propto\ & p(\mathbf{y}_{t+\Delta}|E[\sigma_{t+\Delta}|\sigma_t^{(k)}, \phi_t^{(k)}], a\phi_t^{(k)} + (1-a)\bar{\phi}_t) \\ & \times p(\sigma_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, \sigma_t^{(k)})N(\phi_{t+\Delta}|a\phi_t^{(k)} + (1-a)\bar{\phi}_t, (1-a^2)^2 V_t)\omega_t^{(k)}. \end{aligned} \tag{8}$$

With (8), 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.

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.