

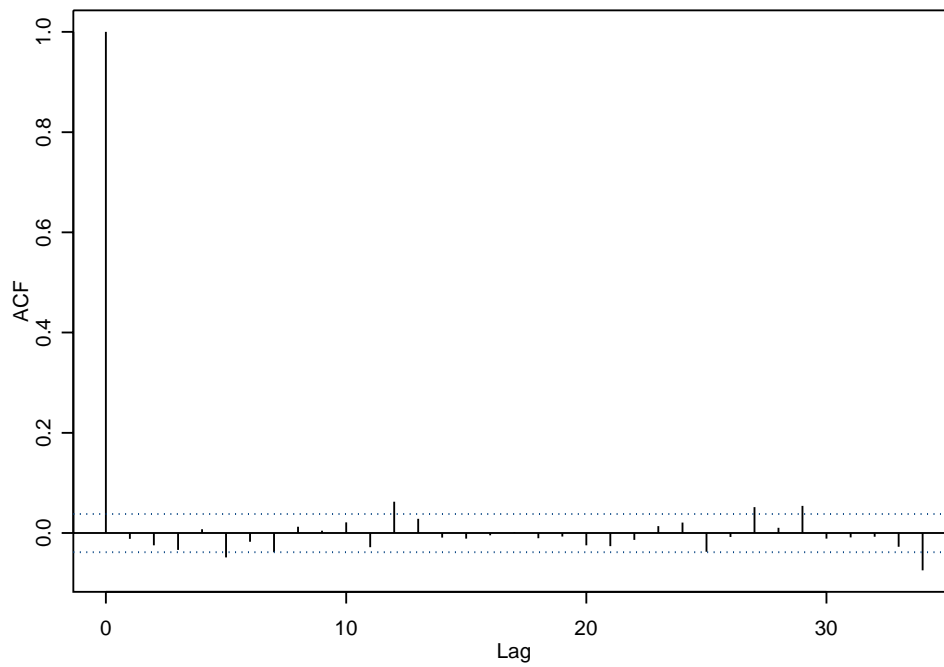
Stochastic Volatility (SV) Models

Jun Yu

1 Motivations

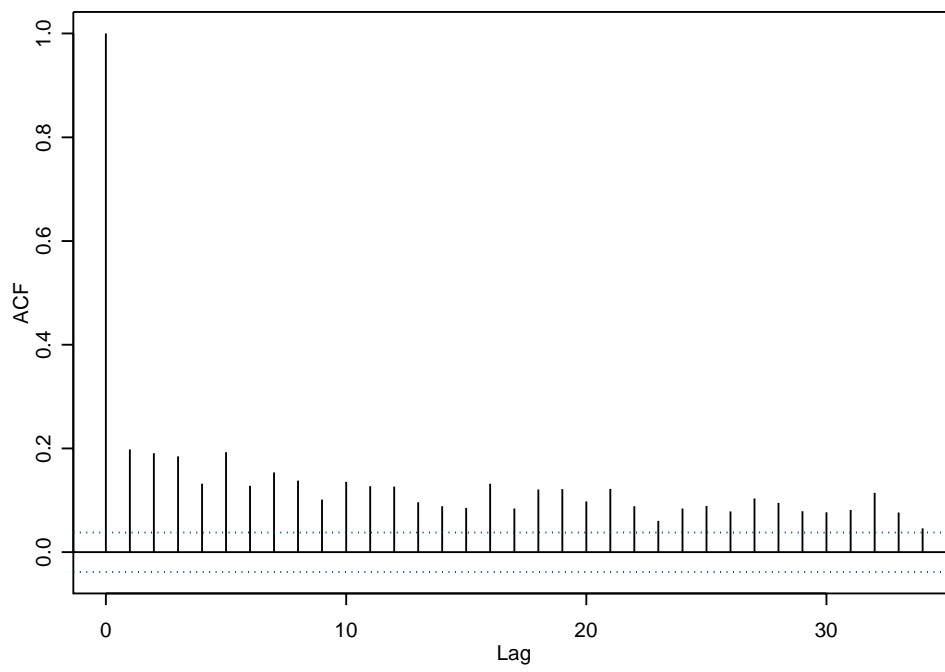
- Some stylised facts about financial asset return distributions:
 1. Distribution is leptokurtic
 2. Volatility clustering
 3. Volatility responds to return news asymmetrically
 4. Returns are cross dependent
 5. Volatilities are cross dependent
 6. Often a lower dimensional factor structure explains most of the correlation
 7. Time varying correlations

Series : sp9303daily



- ACF for typical financial asset returns
- If an ARMA model has to be used, this model suggests an ARMA(0,0).
- The ACF for squared return suggests that ARMA-type models are NOT suitable for financial returns.

Series : sp9303daily^2



2 Univariate SV Models

- MODEL 1: The basic SV model (Taylor, 1982)

– Suppose y_t is the return for a financial asset. The basic SV model is

$$\begin{cases} y_t = \sigma_t \varepsilon_t = \exp(0.5h_t) \varepsilon_t, \varepsilon_t \sim iidN(0, 1) \\ h_t = \lambda + \alpha h_{t-1} + v_t, v_t \sim iidN(0, \sigma_v^2) \end{cases},$$

where $\{\varepsilon_t\}$ and $\{v_t\}$ are uncorrelated.

– Rewriting:

$$\begin{cases} y_t = \sigma \exp(0.5h_t) \varepsilon_t, \varepsilon_t \sim iidN(0, 1) \\ h_t = \alpha h_{t-1} + v_t, v_t \sim iidN(0, \sigma_v^2) \end{cases},$$

– Volatility evolves according to an AR(1) process. Condition on I_{t-1} , h_t is not deterministic any more since v_t is not realized. The assumption could be more realistic than ARCH-type models.

– In the SV model α captures volatility clustering.

– Kurtosis of y_t is $3 \exp(\sigma_v^2 / (1 - \alpha^2))$ which is larger than 3 as long as $\sigma_v \neq 0$. So the unconditional distribution of y_t has fatter tails than the normal distribution. The tail thickness is due to the presence of the second noise.

– $E(y_t y_s) = 0, \forall t \neq s$.

– $E(y_t^2 y_{t-i}^2) = \exp(2\mu_h + \sigma_h^2(1 + \alpha^i))$, where $\mu_h = \frac{\lambda}{1-\alpha}$ and $\sigma_h^2 = \frac{\sigma_v^2}{1-\alpha^2}$.

– We can linearize the model,

$$\begin{cases} x_t = \ln(y_t^2) = h_t + \ln(\varepsilon_t^2) \\ h_t = \lambda + \alpha h_{t-1} + v_t \end{cases}.$$

- MODEL 2: The t-SV model:

$$\begin{cases} y_t = \sigma_t \varepsilon_t = \exp(0.5h_t) \varepsilon_t, \varepsilon_t \sim iidt_{(k)} \\ h_t = \lambda + \alpha h_{t-1} + v_t, v_t \sim iidN(0, \sigma_v^2) \end{cases} .$$

- MODEL 3: Asymmetric SV Model (Yu, 2005):

$$\begin{cases} y_t = \sigma_t \varepsilon_t = \exp(0.5h_t) \varepsilon_t, \varepsilon_t \sim iidN(0, 1) \\ h_t = \lambda + \alpha h_{t-1} + v_t, v_t \sim iidN(0, \sigma_v^2), \end{cases} \quad (2.1)$$

where $\text{corr}(\varepsilon_t, v_{t+1}) = \rho$. Yu (2005) showed that

$$E(h_{t+1}|X_t) = \lambda + \frac{\lambda\phi}{1-\phi^2} + \rho\sigma_v \exp\left(-\frac{\sigma_v^4}{4(1-\phi^2)^2} + \frac{\sigma_v^2\lambda}{(1-\phi^2)(1-\phi)}\right) X_t.$$

This is a linear function in X_t and implies that if $\rho < 0$, and everything else is held constant, a fall in the stock price/return leads to an increase of $E(\ln \sigma_{t+1}^2 | X_t)$.

So this model allows volatility responds asymmetrically to good/bad news

3 Multivariate SV Models

- For the purpose of illustration, we focus on the bivariate multivariate SV models.

Let the observed log-returns at time t be denoted by $\mathbf{y}_t = (y_{1t}, y_{2t})'$ for $t = 1, \dots, T$. Let $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t})'$, $\boldsymbol{\eta}_t = (\eta_{1t}, \eta_{2t})'$, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, $\mathbf{h}_t = (h_{1,t}, h_{2,t})'$, $\Omega_t = \text{diag}(\exp(\mathbf{h}_t/2))$, and

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}, \Sigma_\epsilon = \begin{pmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{pmatrix}, \Sigma_\eta = \begin{pmatrix} \sigma_{\eta_1}^2 & \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} \\ \rho_\eta \sigma_{\eta_1} \sigma_{\eta_2} & \sigma_{\eta_2}^2 \end{pmatrix}.$$

- MODEL 1 (Basic-MSV or **MSV**):

$$\mathbf{y}_t = \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, I),$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. This model is equivalent to stacking two basic univariate SV models together. Clearly, this specification does not allow for correlation across the returns or across the volatilities. However, it does allow for leptokurtic return distributions and volatility clustering.

- MODEL 2 (Constant Correlation-MSV or **CC-MSV**):

$$\mathbf{y}_t = \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon),$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. In this model, the return shocks are allowed to be correlated. As a result, the returns are cross dependent.

- MODEL 3 (MSV with Granger Causality or **GC-MSV**):

$$\mathbf{y}_t = \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon),$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \Phi(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$ and $\phi_{12} = 0$. Since ϕ_{21} can be different from zero, the volatility of the second asset is allowed to be Granger caused by the volatility of the first asset. Consequently, both the returns and volatilities are cross dependent. However, the cross-dependence of volatilities are realized via Granger causality and volatility clustering jointly. Furthermore, when both ϕ_{12} and ϕ_{21} are nonzero, a bilateral Granger causality in volatility between the two assets is allowed.

- MODEL 4 (Generalized CC-MSV or **GCC-MSV**):

$$\mathbf{y}_t = \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\epsilon),$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_\eta),$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. In this model both returns and volatilities are cross dependent. Obviously, both **GC-MSV** and **GCC-MSV** can generate cross dependence in volatilities.

- MODEL 5 (Dynamic Correlation-MSV or **DC-MSV**):

$$\mathbf{y}_t = \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t | \Omega_t \stackrel{iid}{\sim} N(\mathbf{0}, \Sigma_{\epsilon,t}),$$

$$\Sigma_{\epsilon,t} = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix},$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),$$

$$q_{t+1} = \psi_0 + \psi(q_t - \psi_0) + \sigma_\rho v_t, v_t \stackrel{iid}{\sim} N(0, 1), \rho_t = \frac{\exp(q_t) - 1}{\exp(q_t) + 1}.$$

with $\mathbf{h}_0 = \boldsymbol{\mu}, q_0 = \psi_0$. In this model, not only volatilities but also correlation coefficients are time varying. Of course, ρ_t has to be bounded by -1 and 1 for Σ_ϵ to be a well-defined correlation matrix. This constraint is achieved by using the Fisher transformation.

- MODEL 6 (Heavy-tailed MSV or **t-MSV**):

$$\mathbf{y}_t = \Omega_t \boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} t(\mathbf{0}, \Sigma_\epsilon, \nu),$$

$$\mathbf{h}_{t+1} = \boldsymbol{\mu} + \text{diag}(\phi_{11}, \phi_{22})(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \boldsymbol{\eta}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\eta_1}^2, \sigma_{\eta_2}^2)),$$

with $\mathbf{h}_0 = \boldsymbol{\mu}$. In this model, a heavy-tailed multivariate Student t distribution for the return shock is used and hence extra excess kurtosis is allowed.

- MODEL 7 (Additive Factor-MSV or **AFactor-MSV**):

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{D}f_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \stackrel{iid}{\sim} N(\mathbf{0}, \text{diag}(\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2)) \\
f_t &= \exp(h_t/2)u_t, \quad u_t \stackrel{iid}{\sim} N(0, 1), \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \quad \eta_t \stackrel{iid}{\sim} N(0, 1),
\end{aligned}$$

with $h_0 = 0$. The first component in the return equation has a smaller number of factors which capture the information relevant to the pricing of all assets while the second one is idiosyncratic noise which captures the asset specific information. Like the univariate SV model, the **AFactor-MSV** model allows for excess kurtosis and volatility clustering. Clearly, it also allows for cross dependence in both returns and volatilities. Note that in this model that will be introduced below, h_t represents the log-volatility of the common factor, f_t . The conditional correlation coefficient between y_{1t} and y_{2t} is given by:

$$\frac{d \exp(h_t)}{\sqrt{(\exp(h_t) + \sigma_{\epsilon_1}^2)(d^2 \exp(h_t) + \sigma_{\epsilon_2}^2)}} = \frac{d}{\sqrt{(1 + \sigma_{\epsilon_1}^2 \exp(-h_t))(d^2 + \sigma_{\epsilon_2}^2 \exp(-h_t))}}.$$

Unless $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = 0$, the correlation coefficients are time varying but the dynamics of the correlations depend on the dynamics of h_t . Moreover, correlation is an increasing function of h_t , implying that the higher the volatility of the common factor, the higher the correlation in returns.

4 Estimation of SV Models

4.1 Likelihood function

- The ML method is more difficult to use for the SV models than for the ARCH-type models since the likelihood function is numerically more difficult to evaluate for the SV models.
- Let $\theta = (\alpha, \lambda, \sigma_v)$ be the parameters of interest in the basic SV model. We wish to estimate θ from $y = \{y_1, y_2, \dots, y_T\}$. Denote the vector of log-volatilities by $h = \{h_1, h_2, \dots, h_T\}$.
- The likelihood function of the parameter vector θ can be written as

$$L(\theta|y) = pdf(y; \theta) = \int pdf(y, h; \theta) dh \quad (4.2)$$

- The integral cannot be solved analytically. It is of T dimensional. Its dimension cannot be reduced either.

4.2 QMLE with Kalman Filter

- The basic SV model can be represented by a linearized version without losing any information,

$$\begin{cases} y_t = \ln(r_t^2) = h_t + \ln(\varepsilon_t^2) = -1.27 + h_t + \mu_t, \\ h_t = \lambda + \alpha h_{t-1} + v_t \end{cases}, \quad (4.3)$$

with $E(\mu_t) = 0, Var(\mu_t) = \pi^2/2$. If we approximate the distribution of μ_t by a normal distribution with mean 0 and variance $\pi^2/2$, the linearized SV model is approximated by a linear Gaussian State-Space model.

- We here follow the standard notations.

$$\begin{cases} y_t = A'x_t + H'\xi_t + w_t, \\ \xi_{t+1} = F\xi_t + v_{t+1} \end{cases}, \quad (4.4)$$

with $A' = -1.27 + \frac{\lambda}{1-\alpha}$, $x_t = 1$, $H' = 1$, $\xi_t = h_t - \frac{\lambda}{1-\alpha}$, $F = \alpha$, $Q = \sigma_v^2$, $R = \pi^2/2$.

- Based on the linear Gassuain State-Space approximation, Kalman filter can be applied as,

– Initialisation:

$$\begin{cases} \hat{\xi}_{1|0} = 0 \\ \Sigma_{1|0} = \sigma_v^2 / (1 - \alpha^2) \end{cases}, \quad (4.5)$$

– Sequential updating:

$$\begin{cases} \hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + \Sigma_{t|t-1}(\Sigma_{t|t-1} + \pi^2/2)^{-1}(y_t + 1.27 - \frac{\lambda}{1-\alpha} - \hat{\xi}_{t|t-1}) \\ \Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}(\Sigma_{t|t-1} + \pi^2/2)^{-1}\Sigma_{t|t-1} \end{cases}, \quad (4.6)$$

– In-sample sequential prediction:

$$\begin{cases} \hat{\xi}_{t+1|t} = \alpha\hat{\xi}_{t|t-1} + \alpha(1 + \frac{\pi^2}{2\Sigma_{t|t-1}})^{-1}(y_t + 1.27 - \frac{\lambda}{1-\alpha} - \hat{\xi}_{t|t-1}) \\ \Sigma_{t+1|t} = \alpha^2\Sigma_{t|t} + \sigma_v^2 \end{cases}, \quad (4.7)$$

$$\begin{cases} \hat{y}_{t+1|t} = -1.27 + \frac{\lambda}{1-\alpha} + \hat{\xi}_{t+1|t} \\ E[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})'] = \Sigma_{t+1|t} + \pi^2/2 \end{cases}, \quad (4.8)$$

– Out-of-sample forecasting:

$$\begin{cases} \hat{\xi}_{T+h|T} = \alpha^h \hat{E}(\xi_T | I_T) = \alpha^h \hat{\xi}_{T|T} \\ \hat{y}_{T+h|T} = -1.27 + \frac{\lambda}{1-\alpha} + \alpha^h \hat{\xi}_{T|T} \end{cases}, \quad (4.9)$$

– Smoothing:

$$\begin{cases} \hat{\xi}_{t|T} = \hat{\xi}_{t|t} + J_t[\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}] \\ \Sigma_{t+1|T} = \Sigma_{t|t} + J_t(-\Sigma_{t+1|T} + \Sigma_{t+1|t})J_t' \\ J_t = \Sigma_{t|t}\alpha\Sigma_{t+1|t}^{-1} \end{cases}, \quad (4.10)$$

with $t = T - 1, T - 2, \dots, 1$.

- The quasi-likelihood is computed by

$$\ln L(\alpha, \lambda, \sigma_v^2) = -\frac{1}{2} \sum \log(\Sigma_{t|t-1} + \pi^2/2) - \frac{1}{2} \sum \frac{(y_t + 1.27 - \frac{\lambda}{1-\alpha} - \hat{\xi}_{t|t-1})^2}{\Sigma_{t|t-1} + \pi^2/2}.$$

- The h-day ahead forecast is computed by (4.9) with the QML estimates plugged in.

4.3 Simulated Maximum Likelihood

- QML estimators are inefficient
- To do the full ML estimation, we need to calculate the likelihood function, ie, to evaluate the multi-dimensional integral numerically. One way to do this is to using a Monte Carlo method.
- Here we introduce an class of Monte Carlo methods via importance sampling technique.
- We rewrite the multi-dimensional integral by

$$L(\theta|y) = \int pdf(y, h; \theta)dh = \int \frac{pdf(y, h; \theta)}{q(h)}q(h)dh = \int \frac{pdf(y, h; \theta)}{q(h)}dQ(h) \quad (4.11)$$

where $q(h)$ is an importance density and $Q(h)$ is an importance distribution function.

- The idea of the simulated ML method is to draw sample $h^{(1)}, \dots, h^{(S)}$ from q so that we can approximate $L(\theta|y)$ by $\frac{1}{S} \sum_{i=1}^S \frac{pdf(y, h^{(i)}; \theta)}{q(h^{(i)})}$.

- The key of the simulated ML method is to match $pdf(y, h; \theta)$ and $q(h)$ as closely as possible while ensuring that it is easy to simulate from q . To do that, we propose to base the importance sampler on the Laplace approximation to $pdf(y, h; \theta)$. In particular, we choose the mean of q to be the mode of $\ln pdf(y, h; \theta)$ with respect of h , and the variance of q to be the negative of inverse of the second derivative of $\ln pdf(y, h; \theta)$ with respect of h evaluated at the mode. That is $h^{(s)} \sim N(h^*, -\Omega^{-1})$ where

$$h^* = \arg \max_h \ln pdf(y, h; \theta) \quad (4.12)$$

and

$$\Omega = \frac{\partial^2 \ln pdf(y, h^*; \theta)}{\partial h \partial h'} \quad (4.13)$$

4.4 MCMC

- Bayesian inference is then based on the posterior distribution of the unobservables given the data. In the sequel, we will denote the probability density function of a random variable θ by $p(\theta)$. By successive conditioning, the joint prior density is $p(\sigma^2, \alpha, \sigma_v^2, h_0, h_1, \dots, h_T) = p(\sigma^2, \alpha, \sigma_v^2)p(h_0|\sigma_v^2) \prod_{t=1}^n p(h_t|h_{t-1}, \alpha, \sigma_v^2)$.
- We assume $h_0 \sim N(0, \sigma_v^2/(1 - \alpha^2))$, prior independence of the parameters σ^2 , α , and σ_v^2 , and use the standard noninformative priors, i.e. $p(\sigma^2) \propto \frac{1}{\sigma^2}$, $\alpha \propto 1$, and $p(\sigma_v^2) \propto \frac{1}{\sigma_v^2}$. $p(h_t|h_{t-1}, \alpha, \sigma_v^2)$ is defined through the variance equation. The likelihood $p(y_1, \dots, y_T|\sigma^2, \alpha, \sigma_v^2, h_0, \dots, h_T)$ is defined through the return equation and the conditional independence assumption:

$$p(y_1, \dots, y_T|\sigma^2, \alpha, \sigma_v^2, h_0, \dots, h_T) = \prod_{t=1}^T p(y_t|h_t, \sigma^2). \quad (4.14)$$

Then, by Bayes' theorem, the joint posterior distribution of the unobservables given the data is proportional to the prior times likelihood, i.e.

$$p(\sigma^2, \alpha, \sigma_v^2, h_0, \dots, h_T|y_1, \dots, y_T) \propto \frac{p(\sigma^2)p(\alpha)p(\tau^2)p(h_0|\sigma_v^2) \prod_{t=1}^T p(h_t|h_{t-1}, \alpha, \sigma_v^2) \times \prod_{t=1}^T p(y_t|h_t, \sigma^2)}{\prod_{t=1}^T p(y_t|h_t, \sigma^2)}. \quad (4.15)$$