VII. Time-varying variances

A. Introduction to ARCH models
B. Extensions
C. Markov-switching GARCH
D. Stochastic volatility

GARCH family:

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t$$
$$u_t = \sqrt{h_t}\, v_t$$
$$v_t \sim \text{i.i.d. } (0,1) \text{ (e.g. } N(0,1))$$
$$h_t = h(u_{t-1}, u_{t-2}, \ldots)$$

Implication:

the difference between the realized value y_t and its conditional expectation $\mathbf{x}_t'\boldsymbol{\beta}$ is the only information useful for forecasting the variance h_t

Stochastic volatility:

Some latent variables in addition to u_{t-j} contribute to h_t

Example:

$$y_t = \exp(h_t/2)v_t$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t$$
$$\begin{bmatrix} v_t \\ \eta_t \end{bmatrix} \sim \text{ i.i.d. } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

argument in favor of stochastic vol:

more natural and flexible

argument in favor of GARCH:

ultimately our forecast

$E(u_t^2|\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots)$

will be some function of

$(\mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, y_{t-2}, \ldots)$

so why not take this function as a primitive of the model?

Note sv model above implies

$$y_t^2 = \exp(h_t)v_t^2$$
$$\log y_t^2 = h_t + \log v_t^2$$
$$\log y_t^2 = \mu + (h_t - \mu) + \log v_t^2$$

For $\xi_t = h_t - \mu$ this is a state-space model of the form

$$\xi_t = \phi\xi_{t-1} + \sigma\eta_t$$
$$\log y_t^2 = \mu + \xi_t + \log v_t^2$$

problem: $\log v_t^2$ is not Normally distributed

Gibbs sampler blocks:

(1) $\sigma^2|\phi, \mu, \mathbf{h}, \mathbf{y}$

gamma prior for σ^{-2} implies gamma posterior

(2) $\phi|\sigma^2, \mu, \mathbf{h}, \mathbf{y}$

easy way:

Normal prior for $\phi|\sigma^2$ implies Normal posterior

However: authors want to restrict $\phi \in (-1, 1)$ and use unconditional distribution of first observation

prior:

$\phi^* \sim$ Beta(α_1, α_2)

$\Rightarrow p(\phi^*) \propto (\phi^*)^{\alpha_1 - 1}(1 - \phi^*)^{\alpha_2 - 1}$

$\phi = 2\phi^* - 1 \Rightarrow \phi \in (-1, 1)$

$[\phi^* = (1 + \phi)/2]$

$$\Rightarrow p(\phi) \propto \begin{cases} \left[\frac{1+\phi}{2}\right]^{a_1-1}\left[\frac{1-\phi}{2}\right]^{\alpha_2-1} & \phi \in (-1,1) \\ 0 & \text{otherwise} \end{cases}$$

$$h_1|\mu,\phi,\sigma^2 \sim N(\mu,\sigma^2/(1-\phi^2))$$

$$h_t|h_{t-1},h_{t-2},\ldots,h_1,\mu,\phi,\sigma^2)$$

$$\sim N(\mu+\phi(h_{t-1}-\mu),\sigma^2)$$

$$p(\mathbf{h}|\mu,\phi,\sigma^2) \propto \frac{1}{[\sigma^2/(1-\phi^2)]^{1/2}} \times$$

$$\exp\left[-\frac{(h_1-\mu)^2}{2\sigma^2/(1-\phi^2)} - \frac{\sum_{t=2}^{T}[h_t-\mu-\phi(h_{t-1}-\mu)]^2}{2\sigma^2}\right]$$

$$p(\phi|\sigma^2,\mu,\mathbf{h},\mathbf{y}) = p(\phi|\sigma^2,\mu,\mathbf{h})$$

$$\propto p(\phi)p(\mathbf{h}|\mu,\phi,\sigma^2)$$

can sample with importance
or Metropolis-Hastings

(3) $\mu|\phi, \sigma^2, \mathbf{h}, \mathbf{y}$

prior:

$\mu|\sigma^2 \sim N(m_\mu, \sigma^2 M_\mu)$

diffuse:

$M_\mu^{-1} = 0$

posterior:

$\mu|\sigma^2, \phi, \mathbf{h} \sim N(m_\mu^*, \sigma^2 M_\mu^*)$

(correct for different variance of h_1 from $h_2, h_3, \ldots, h_T$)

(4) $\mathbf{h}|\boldsymbol{\theta}, \mathbf{y}$

for $\boldsymbol{\theta} = (\phi, \mu, \sigma^2)'$

(i) solution? (doesn't work well)

Let $\mathbf{h}_{-t} = (h_1, h_2, \ldots, h_{t-1}, h_{t+1}, \ldots, h_T)'$

$p(h_t|\mathbf{h}_{-t}, \boldsymbol{\theta}, \mathbf{y}) = p(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}, y_t)$

$\qquad \propto p(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta}) p(y_t|h_t, \boldsymbol{\theta})$

To get $p(h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta})$, calculate the joint distribution $p(h_{t-1}, h_t, h_{t+1}|\boldsymbol{\theta})$ and use formula for conditional Normal:

$h_t|h_{t-1}, h_{t+1}, \boldsymbol{\theta} \sim N(m_{h_t}, \sigma^2 M_{h_t})$

$m_{h_t} = \mu + \frac{\phi[(h_{t-1}-\mu)+(h_{t+1}-\mu)]}{(1+\phi^2)}$

$M_{h_t} = \frac{1}{1+\phi^2}$

Why doesn't this work well?
Answer: $h_t, h_{t+1} | \mathbf{y}, \boldsymbol{\theta}$ are too highly correlated

(ii) solution? (a little better)
approximate non-Gaussian state-space model with mixture of Gaussian state space

Recall
$z_t = \log v_t^2 \sim \log[\chi^2(1)]$
claim: can approximate this density arbitrarily well with a mixture of Normals:

$$p(z_t) = \sum_{i=1}^{K} \frac{\pi_i}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{(z_t - \mu_i)}{2\sigma_i^2} \right]$$

$K = 7$ gives excellent approximation
values of π_i, σ_i, μ_i are numerically known
(not a function of data, function of
$\log[\chi^2(1)]$ distribution)

Claim 2: we could generate a
value for z_t from this distribution
in two steps

step 1: generate $s_t \in \{1, 2, \ldots, K\}$

Prob$(s_t = i) = \pi_i$

step 2: generate $z_t | s_t \sim N(\mu_{s_t}, \sigma_{s_t}^2)$

Gibbs sampler blocks:

(1) $\sigma^2 | \phi, \mu, \mathbf{h}, \mathbf{y}, \mathbf{s}$

(2) $\phi | \sigma^2, \mu, \mathbf{h}, \mathbf{y}, \mathbf{s}$

(3) $\mu | \phi, \sigma^2, \mathbf{h}, \mathbf{y}, \mathbf{s}$

(1)-(3) are done same as before

(4) $\mathbf{s}|\boldsymbol{\theta},\mathbf{h},\mathbf{y}$

Notice that conditional on y_t and h_t, the value of $z_t = \log(y_t^2) - h_t$ is known (though authors replace $\log(y_t^2)$ with $y_t^* = \log(y_t^2 + 0.001)$ for robustness)

$$p(s_t = j|z_t) = \frac{(\pi_j/\sigma_j)\exp\left[-\frac{(z_t-\mu_j)^2}{2\sigma_j^2}\right]}{\sum_{i=1}^{K}(\pi_i/\sigma_i)\exp\left[-\frac{(z_t-\mu_i)^2}{2\sigma_i^2}\right]}$$

Recall our original state-space description:

$$\log y_t^2 = \mu + (h_t - \mu) + \log v_t^2$$

$$h_t - \mu = \phi(h_{t-1} - \mu) + \sigma\eta_t$$

or for

$$z_t = \log v_t^2 \sim N(\mu_{s_t}, \sigma_{s_t}^2)$$

$$\xi_t = h_t - \mu$$

$$y_t^* = \log(y_t^2 + 0.001)$$

this becomes

$$y_t^* = \mu + \xi_t + z_t$$

$$\xi_t = \phi\xi_{t-1} + \sigma\eta_t$$

$$z_t \sim N(\mu_{s_t}, \sigma_{s_t}^2)$$

So we can generate

$\xi_1,\ldots,\xi_T|\boldsymbol{\theta},\mathbf{y},\mathbf{s}$

from Kalman smoother

(and thus $h_t = \xi_t + \mu$)

problem: $\mathbf{h}$ is highly correlated with $\boldsymbol{\theta} = (\phi, \mu, \sigma^2)'$

(iii) solution? (almost there)
goal: sample from $p(\boldsymbol{\theta}, \mathbf{h} | \mathbf{y}^*, \mathbf{s})$ rather than $p(\boldsymbol{\theta} | \mathbf{h}, \mathbf{y}^*, \mathbf{s})$ and $p(\mathbf{h} | \boldsymbol{\theta}, \mathbf{y}^*, \mathbf{s})$ separately

$$p(\phi, \sigma^2, \mu, \mathbf{h} | \mathbf{y}^*, \mathbf{s}) = p(\phi, \sigma^2 | \mathbf{y}^*, \mathbf{s}) p(\mu, \mathbf{h} | \phi, \sigma^2, \mathbf{y}^*, \mathbf{s})$$

Tool to accomplish: augmented Kalman filter

$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1} \quad \mathbf{v}_{t+1} \sim N(\mathbf{0}, \mathbf{Q})$

$y_t = \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{h}'\boldsymbol{\xi}_t + w_t \quad w_t \sim N(0, R)$

[could generalize to vector $\mathbf{y}_t$ using earlier vectorization tricks]

Kalman filter:

$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{K}_t(y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\hat{\boldsymbol{\xi}}_{t|t-1})$

$\mathbf{K}_t = \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{h}(\mathbf{h}'\mathbf{P}_{t|t-1}\mathbf{h} + R)^{-1}$

note: $\{\mathbf{K}_t\}_{t=1}^{T}$ does not depend on $\boldsymbol{\beta}$ or $\{\mathbf{x}_t, y_t\}_{t=1}^{T}$

$y_t|\boldsymbol{\theta}, \Omega_{t-1} \sim N(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{h}'\hat{\boldsymbol{\xi}}_{t|t-1}, c_{t|t-1}^2)$

$c_{t|t-1}^2 = \mathbf{h}'\mathbf{P}_{t|t-1}\mathbf{h} + R$

so usual KF could tell us $p(\mathbf{y}|\boldsymbol{\theta})$ for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\lambda}')'$ and $\boldsymbol{\lambda}$ elements of parameter vector other than $\boldsymbol{\beta}$

goal: find $\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}$

Consider running through the KF for two special cases:

(1) set $\boldsymbol{\beta} = \mathbf{0}$ and variable in observation equation to y_t:

$$\xi_{t+1}^{(1)} = \mathbf{F}\xi_t^{(1)} + \mathbf{K}_t[y_t - \mathbf{h}'\xi_t^{(1)}]$$

starting from $\xi_0^{(1)} = \hat{\xi}_{0|0}$ for $\hat{\xi}_{0|0}$ the standard KF starting value with nonzero $\boldsymbol{\beta}$

(2i) set $\boldsymbol{\beta} = \mathbf{0}$ and variable in observation equation to $-x_{it}$ (for $i = 1, 2, \ldots, k$)

$$\xi_{t+1}^{(2i)} = \mathbf{F}\xi_t^{(2i)} + \mathbf{K}_t[-x_{it} - \mathbf{h}'\xi_t^{(2i)}]$$

starting from $\xi_0^{(2i)} = \mathbf{0}$

Define matrix

$$\boldsymbol{\xi}_t^{(2)} = \begin{bmatrix} \boldsymbol{\xi}_t^{(2,1)} & \boldsymbol{\xi}_t^{(2,2)} & \cdots & \boldsymbol{\xi}_t^{(2,k)} \end{bmatrix}$$

Then $\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta} = \sum_{i=1}^{k} \boldsymbol{\xi}_t^{(2i)}\beta_i$

$$\boldsymbol{\xi}_{t+1}^{(2i)} = \mathbf{F}\boldsymbol{\xi}_t^{(2i)} + \mathbf{K}_t[-x_{it} - \mathbf{h}'\boldsymbol{\xi}_t^{(2i)}]$$

$$\boldsymbol{\xi}_{t+1}^{(2)}\boldsymbol{\beta} = \mathbf{F}\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta} + \mathbf{K}_t\left[-\mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}\right]$$

$$\boldsymbol{\xi}_{t+1}^{(1)} = \mathbf{F}\boldsymbol{\xi}_t^{(1)} + \mathbf{K}_t[y_t - \mathbf{h}'\boldsymbol{\xi}_t^{(1)}]$$

$$\boldsymbol{\xi}_{t+1}^{(2)}\boldsymbol{\beta} = \mathbf{F}\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta} + \mathbf{K}_t\left[-\mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}\right]$$

$$\boldsymbol{\xi}_{t+1}^{(1)} + \boldsymbol{\xi}_{t+1}^{(2)}\boldsymbol{\beta} = \mathbf{F}[\boldsymbol{\xi}_t^{(1)} + \boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}] + \mathbf{K}_t\left[y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\boldsymbol{\xi}_t^{(1)} - \mathbf{h}'\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}\right]$$

starting from $\boldsymbol{\xi}_0^{(1)} + \boldsymbol{\xi}_0^{(2)}\boldsymbol{\beta} = \hat{\boldsymbol{\xi}}_{0|0}$

$$\boldsymbol{\xi}_{t+1}^{(1)} + \boldsymbol{\xi}_{t+1}^{(2)}\boldsymbol{\beta} = \mathbf{F}[\boldsymbol{\xi}_t^{(1)} + \boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}] + \mathbf{K}_t\left[y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\boldsymbol{\xi}_t^{(1)} - \mathbf{h}'\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}\right]$$

starting from $\boldsymbol{\xi}_0^{(1)} + \boldsymbol{\xi}_0^{(2)}\boldsymbol{\beta} = \hat{\boldsymbol{\xi}}_{0|0}$

Compare with standard KF:

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\hat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{K}_t\left[y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\hat{\boldsymbol{\xi}}_{t|t-1}\right]$$

starting from $\hat{\boldsymbol{\xi}}_{0|0}$

conclusion: $\hat{\boldsymbol{\xi}}_{t|t-1} = \boldsymbol{\xi}_t^{(1)} + \boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}$

Since

$\hat{\boldsymbol{\xi}}_{t|t-1} = \boldsymbol{\xi}_t^{(1)} + \boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}$ and

$y_t|\boldsymbol{\theta}, \Omega_{t-1} \sim N(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{h}'\hat{\boldsymbol{\xi}}_{t|t-1}, c_{t|t-1}^2)$

it follows that

$y_t|\boldsymbol{\theta}, \Omega_{t-1} \sim N(\mathbf{x}_t'\boldsymbol{\beta} + \mathbf{h}'\boldsymbol{\xi}_t^{(1)} + \mathbf{h}'\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta}, c_{t|t-1}^2)$

$v_t^{(1)}|\boldsymbol{\theta}, \Omega_{t-1} \sim N(\tilde{\mathbf{x}}_t'\boldsymbol{\beta}, c_{t|t-1}^2)$

for

$v_t^{(1)} = y_t - \mathbf{h}'\boldsymbol{\xi}_t^{(1)}$

$\tilde{\mathbf{x}}_t' = \mathbf{x}_t' + \mathbf{h}'\boldsymbol{\xi}_t^{(2)}$

Notice $\boldsymbol{\beta}$ was not used to form $v_t^{(1)}$, $\tilde{\mathbf{x}}_t$, or $c_{t|t-1}^2$

Conclusion: the distribution $\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\lambda}$ can be found from Bayesian posterior distribution of $\boldsymbol{\beta}$ for the regression

$$v_t^{(1)} = \tilde{\mathbf{x}}_t'\boldsymbol{\beta} + u_t^{(1)}$$

$$u_t^{(1)} \sim N(0, c_{t|t-1}^2)$$

that is, with diffuse prior,

$\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\lambda} \sim N(\mathbf{m}_{\boldsymbol{\beta}}^*, \mathbf{M}_{\boldsymbol{\beta}}^*)$

$\mathbf{M}_{\boldsymbol{\beta}}^* = \left(\sum_{t=1}^{T} \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t' / c_{t|t-1}^2\right)^{-1}$

$\mathbf{m}_{\boldsymbol{\beta}}^* = \mathbf{M}_{\boldsymbol{\beta}}^* \left(\sum_{t=1}^{T} \tilde{\mathbf{x}}_t v_t^{(1)} / c_{t|t-1}^2\right)$

What have we accomplished?
Original formulation Gibbs sampled $p(\boldsymbol{\beta}|\boldsymbol{\xi},\mathbf{y},\boldsymbol{\lambda})$ and $p(\boldsymbol{\xi}|\boldsymbol{\beta},\mathbf{y},\boldsymbol{\lambda})$
problem was they were too correlated

Augmented Kalman filter allowed us instead to sample from $p(\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\lambda})$ and then with usual Kalman filter we can draw from $p(\boldsymbol{\xi}|\boldsymbol{\beta},\mathbf{y},\boldsymbol{\lambda})$ to get $p(\boldsymbol{\xi},\boldsymbol{\beta}|\mathbf{y},\boldsymbol{\lambda})$

Note this also allows us to calculate the distribution of $\mathbf{y}$ without conditioning on β or ξ:

$$p(\mathbf{y}|\boldsymbol{\lambda}) = \frac{p(\mathbf{y}|\boldsymbol{\beta} = \mathbf{0}, \boldsymbol{\lambda})p(\boldsymbol{\beta} = \mathbf{0})}{p(\boldsymbol{\beta} = \mathbf{0}|\mathbf{y}, \boldsymbol{\lambda})}$$

The first term we know from the basic Kalman filter:

$$p(\mathbf{y}|\boldsymbol{\beta} = \mathbf{0}, \boldsymbol{\lambda}) = \prod_{t=1}^{T} (2\pi c_{t|t-1}^2)^{-1/2} \times$$

$$\exp\left\{-\frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{h}'\boldsymbol{\xi}_t^{(1)} - \mathbf{h}'\boldsymbol{\xi}_t^{(2)}\boldsymbol{\beta})^2}{2c_{t|t-1}^2}\right\}_{\boldsymbol{\beta}=\mathbf{0}}$$

$$= \prod_{t=1}^{T} (2\pi c_{t|t-1}^2)^{-1/2} \exp\left\{-\frac{(y_t - \mathbf{h}'\boldsymbol{\xi}_t^{(1)})^2}{2c_{t|t-1}^2}\right\}$$

The second term we know by evaluating our prior at $\boldsymbol{\beta} = \mathbf{0}$:

$$p(\boldsymbol{\beta} = \mathbf{0}) = (2\pi)^{-k/2}|\mathbf{M}_{\beta}|^{-1/2} \times$$

$$\exp\left\{-(1/2)(\boldsymbol{\beta} - \mathbf{m}_{\beta})'\mathbf{M}_{\beta}^{-1}(\boldsymbol{\beta} - \mathbf{m}_{\beta})\right\}_{\boldsymbol{\beta}=\mathbf{0}}$$

$$= (2\pi)^{-k/2}|\mathbf{M}_{\beta}|^{-1/2} \exp\left\{-(1/2)\mathbf{m}_{\beta}'\mathbf{M}_{\beta}^{-1}\mathbf{m}_{\beta}\right\}$$

The third term we know by evaluating our posterior at $\boldsymbol{\beta} = \mathbf{0}$:

$$p(\boldsymbol{\beta} = \mathbf{0}|\mathbf{y},\boldsymbol{\lambda}) = (2\pi)^{-k/2}|\mathbf{M}_{\boldsymbol{\beta}}^{*}|^{-1/2} \times \exp\left\{-(1/2)(\boldsymbol{\beta} - \mathbf{m}_{\boldsymbol{\beta}}^{*})'\mathbf{M}_{\boldsymbol{\beta}}^{*-1}(\boldsymbol{\beta} - \mathbf{m}_{\boldsymbol{\beta}}^{*})\right\}_{\boldsymbol{\beta}=\mathbf{0}}$$

$$= (2\pi)^{-k/2}|\mathbf{M}_{\boldsymbol{\beta}}^{*}|^{-1/2} \exp\left\{-(1/2)\mathbf{m}_{\boldsymbol{\beta}}^{*\prime}\, \mathbf{M}_{\boldsymbol{\beta}}^{*-1}\mathbf{m}_{\boldsymbol{\beta}}^{*}\right\}$$

Thus for $\mathbf{m}_{\boldsymbol{\beta}} = \mathbf{0}$ we have

$$p(\mathbf{y}|\boldsymbol{\lambda}) = \frac{p(\mathbf{y}|\boldsymbol{\beta}=\mathbf{0},\boldsymbol{\lambda})p(\boldsymbol{\beta}=\mathbf{0})}{p(\boldsymbol{\beta}=\mathbf{0}|\mathbf{y},\boldsymbol{\lambda})}$$

$$\propto \prod_{t=1}^{T}(2\pi c_{t|t-1}^{2})^{-1/2} \exp\left\{-\frac{(y_t-\mathbf{h}'\boldsymbol{\xi}_t^{(1)})^2}{2c_{t|t-1}^{2}}\right\} \times |\mathbf{M}_{\boldsymbol{\beta}}|^{-1/2}|\mathbf{M}_{\boldsymbol{\beta}}^{*}|^{1/2} \exp\left\{(1/2)\mathbf{m}_{\boldsymbol{\beta}}^{*\prime}\, \mathbf{M}_{\boldsymbol{\beta}}^{*-1}\mathbf{m}_{\boldsymbol{\beta}}^{*}\right\}$$

For our stochastic volatility example, we can use this to calculate $p(\mathbf{y}^{*}|\mathbf{s},\phi,\sigma^2)$ and thus to draw from $p(\phi,\sigma^2|\mathbf{y}^{*},\mathbf{s}) \propto p(\mathbf{y}^{*}|\mathbf{s},\phi,\sigma^2)p(\phi)p(\sigma^{-2})$ using same numerical methods as before.

Finally, we can draw from $p(\boldsymbol{\beta}|\mathbf{y}^*, \mathbf{s}, \phi, \sigma^2)$ using the $N(\mathbf{m}^*_{\boldsymbol{\beta}}, \mathbf{M}^*_{\boldsymbol{\beta}})$ distribution, and finally from $p(\boldsymbol{\xi}|\boldsymbol{\beta}, \mathbf{y}^*, \mathbf{s}, \phi, \sigma^2)$ using the standard Kalman filter.

So why isn't this the authors final solution? Answer: above calculations replaced

$$\log y_t^2 = h_t + \log v_t^2 \qquad \log v_t^2 \sim \chi^2(1)$$

with the approximation

$$y_t^* = \log(y_t^2 + 0.001) \simeq h_t + z_t$$

$$p(z_t) = \sum_{i=1}^{7} \pi_i (2\pi\sigma_i^2)^{-1/2} \times \exp\left[-\frac{(z_t-\mu_i)^2}{2\sigma_i^2}\right]$$

(iv) solution!
reweight to correct for approximation bias

Note that for any y_t and generated h_t we can calculate the true density

$$p(y_t|h_t) = (2\pi \exp h_t)^{-1/2} \exp\left\{-\frac{y_t^2}{2\exp(h_t)}\right\}$$

and the value of

$$\kappa(y_t^*|h_t) = \sum_{i=1}^{7} \pi_i (2\pi\sigma_i^2)^{-1/2} \times \exp\left[-\frac{(y_t^* - h_t - \mu_i)^2}{2\sigma_i^2}\right]$$

KSC then form the weight $w(\mathbf{h}) = \exp\left\{\sum_{t=1}^{T} [\log p(y_t|h_t) - \log \kappa(y_t^*|h_t)]\right\}$ and weighting the draws from $\kappa(\boldsymbol{\theta}, \mathbf{h}|\mathbf{y}^*)$ as in importance sampling, i.e.,

$$E(\boldsymbol{\theta}|\mathbf{y}) = \frac{\sum_{m=1}^{M} \boldsymbol{\theta}^{(m)} w(\mathbf{h}^{(m)})}{\sum_{m=1}^{M} w(\mathbf{h}^{(m)})}$$

for $m = 1, \ldots, M$ the separate simulated draws