Financial Econometrics and Volatility Models
Estimation of Stochastic Volatility Models

Eric Zivot

April 26, 2010

Outline

- Likehood of SV Models
- Survey of Estimation Techniques for SV Models
- GMM Estimation of SV Models
- State Space Estimation of SV Models
- Forecasting Volatility from SV Models

Reading

- FMUND, chapter 4 (section 7)
- MFTS, chapter 14 (section 4.3), 21 (section 7.4), 23 (section 4.2)
- APDVP, chapter 11

Estimation Techniques for SV Models

- Likelihood function

- Estimation methods

Standard SV Model

$$
\begin{gathered}
r_t = \mu + \sigma_t u_t \\
\ln(\sigma_t) - \alpha = \phi(\ln(\sigma_{t-1} - \alpha) + \eta_t, |\phi| < 1 \\
\begin{pmatrix} u_t \\ \eta_t \end{pmatrix} \sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}\right)
\end{gathered}
$$

or

$$
\begin{aligned}
r_t &= \mu + \exp(w_t/2)u_t,\ w_t = \ln(\sigma_t^2) = 2\ln(\sigma_t) \\
w_t - \alpha_w &= \phi(w_{t-1} - \alpha_w) + \eta_{w,t},\ \eta_{w,t} \sim iid\ N(0, \sigma_{\eta_w}^2) \\
\begin{pmatrix} u_t \\ \eta_{w,t} \end{pmatrix} &\sim iid\ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{\eta_w}^2 \end{pmatrix}\right)
\end{aligned}
$$

Typicall, μ is first estimated using

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} r_t$$

and the SV model is fit to the demeaned returns

$$y_t = r_t - \hat{\mu}$$

The parameters to be estimated are then

$$\begin{aligned}
\boldsymbol{\theta} &= (\alpha, \phi, \sigma_\eta^2)' \text{ or } (\alpha, \phi, \beta^2)',\ \beta^2 = \sigma_\eta^2(1-\phi^2) \\
\boldsymbol{\theta}_w &= (\alpha_w, \phi, \sigma_{\eta_w}^2)' \text{ or } (\alpha_w, \phi, \beta_w^2)',\ \beta^2 = \sigma_{\eta_w}^2(1-\phi^2)
\end{aligned}$$

Likelihood Function

$\{r_1, \ldots, r_n\}$ = sample of observed returns from SV model

The likelihood function for $\{r_1, \ldots, r_n\}$ given $\boldsymbol{\theta}$ is

$$L(r_1, \ldots, r_n|\theta) = \int \cdots \int f(r_1, \ldots, r_n|\sigma_1, \ldots, \sigma_n; \theta) f(\sigma_1, \ldots, \sigma_n) d\sigma_1 \cdots d\sigma_n$$

cannot be evaluated analytically. However,

$$\begin{aligned} f(r_1, \ldots, r_n|\sigma_1, \ldots, \sigma_n; \theta) &= \Pi_{i=1}^{n} f(r_i|\sigma_i; \theta) \\ f(r_i|\sigma_i) &= N(\mu, \sigma_i^2) \end{aligned}$$

and $f(\sigma_1, \ldots, \sigma_n)$ can be evaluated using the fact that $\ln \sigma_t$ follows a Gaussian AR(1) model.

- Method of moments

- Generalized method of moments (GMM)

- Quasi-MLE from state space model

- Simulated MLE

- Indirect Inference

- Bayesian MCMC

Method of Moments

Idea: Estimate parameters by matching population moments to sample moments

Population parameters: α, β^2, ϕ

Population moments

$$\begin{aligned}
E[|r_t - \mu|] &= E[a_t] = \sqrt{2/\pi}\left(\alpha + \frac{1}{2}\beta^2\right) \\
\mathrm{var}(r_t) &= \exp(2\alpha + 2\beta^2) \\
\mathrm{kurt}(r_t) &= 3\exp(4\beta^2)
\end{aligned}$$

Method of moment estimators for α and β using $\mathsf{var}(r_t)$ and $\mathsf{kurt}(r_t)$:

$$\begin{aligned}\hat{\beta}^2 &= \frac{1}{4}\ln\left(\frac{\widehat{\mathsf{kurt}}(r_t)}{3}\right)\\ \hat{\alpha} &= \frac{\ln(\widehat{\mathsf{var}}(r_t))}{2} - \hat{\beta}^2\end{aligned}$$

Problem: $\widehat{\mathsf{kurt}}(r_t)$ is very sensitive to outliers

Method moment estimators for α and β using $E[|r_t - \mu|]$ and $\mathsf{var}(r_t)$:

$$
\begin{aligned}
\hat{\alpha} &= \ln\left(\frac{\pi \bar{a}^2}{2\sqrt{\widehat{\mathsf{var}}(r_t)}}\right) \\
\hat{\beta} &= \ln\left(\frac{2\widehat{\mathsf{var}}(r_t)}{\pi \bar{a}^2}\right)
\end{aligned}
$$

where

$$
\bar{a} = \frac{1}{n}\sum_{t=1}^{n} |r_t - \bar{r}|, \; r = \frac{1}{n}\sum_{t=1}^{n} r_t
$$

- No simple method of moments estimator for ϕ

- Note: Method of moment estimators are not unique

Review of GMM Estimation

Idea: GMM optimally combines moment conditions to estimate population parameters

Let $\{\mathbf{w}_t\}$ be a covariance stationary and ergodic vector process representing the underlying data. Let the $p \times 1$ vector $\boldsymbol{\theta}$ denote the population parameters. The moment conditions $\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$ are $K \geq p$ possibly nonlinear functions satisfying

$$E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] = \mathbf{0}$$

where $\boldsymbol{\theta}_0$ represent the true parameter vector.

Global identification of $\boldsymbol{\theta}_0$ requires that

$$\begin{aligned} E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)] &= \mathbf{0} \\ E[\mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})] &\neq \mathbf{0} \text{ for } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{aligned}$$

Local Identification requires that the $K \times p$ matrix

$$\mathbf{G} = E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$$

has full column rank p.

The sample moment condition for an arbitrary $\boldsymbol{\theta}$ is

$$\mathbf{g}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta})$$

If $K = p$, then $\boldsymbol{\theta}_0$ is apparently just identified and the GMM objective function is

$$J(\boldsymbol{\theta}) = n\mathbf{g}_n(\boldsymbol{\theta})'\mathbf{g}_n(\boldsymbol{\theta})$$

which does not depend on a weight matrix.

The corresponding GMM estimator is then

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \; J(\boldsymbol{\theta})$$

and satisfies $\mathbf{g}_n(\hat{\boldsymbol{\theta}}) = 0$.

If $K > p$, then $\boldsymbol{\theta}_0$ is apparently overidentified.

Let $\hat{\mathbf{W}}$ denote a $K \times K$ symmetric and p.d. weight matrix, possibly dependent on the data, such that $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ as $n \to \infty$ with $\mathbf{W}$ symmetric and p.d.

The GMM estimator of $\boldsymbol{\theta}_0$, denoted $\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})$, is defined as

$$\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) = \arg\min_{\boldsymbol{\theta}} \; J(\boldsymbol{\theta}, \hat{\mathbf{W}}) = n\mathbf{g}_n(\boldsymbol{\theta})'\hat{\mathbf{W}}\mathbf{g}_n(\boldsymbol{\theta})$$

The first order conditions are

$$\begin{aligned} \frac{\partial J(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}), \hat{\mathbf{W}})}{\partial \boldsymbol{\theta}} &= 2\mathbf{G}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))'\hat{\mathbf{W}}\mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) = \mathbf{0} \\ \mathbf{G}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) &= \frac{\partial \mathbf{g}_n(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))}{\partial \boldsymbol{\theta}'} \end{aligned}$$

Asymptotic Properties of Nonlinear GMM

Under standard regularity conditions, it can be shown that

$$\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) \xrightarrow{p} \boldsymbol{\theta}_0$$

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}) - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \mathsf{avar}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})))$$

where

$$\begin{aligned} \mathsf{avar}(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}})) &= (\mathbf{G}'\mathbf{W}\mathbf{G})^{-1}\mathbf{G}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{G}(\mathbf{G}'\mathbf{W}\mathbf{G})^{-1} \\ \mathbf{G} &= E\left[\frac{\partial \mathbf{g}(\mathbf{w}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right], \ \mathbf{S} = \mathsf{avar}(\sqrt{n}\mathbf{g}_n(\boldsymbol{\theta}_0)) \end{aligned}$$

The efficient GMM estimator uses a weight matrix $\mathbf{W}$ that minimizes avar$(\hat{\boldsymbol{\theta}}(\hat{\mathbf{W}}))$. Hansen (1982) showed that the optimal weight matrix is $\mathbf{W} = \mathbf{S}^{-1}$. In this case,

$$\text{avar}(\hat{\boldsymbol{\theta}}(\mathbf{S}^{-1})) = (\mathbf{G}'\mathbf{S}^{-1}\mathbf{G})^{-1}$$

If $\{\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ is an ergodic stationary MDS then

$$\mathbf{S} = E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)'].$$

A consistent estimator of $\mathbf{S}$ takes the form

$$\hat{\mathbf{S}}_{\text{HC}} = n^{-1}\sum_{t=1}^{n}\mathbf{g}_t(\mathbf{w}_t, \hat{\boldsymbol{\theta}})\mathbf{g}_t(\mathbf{w}_t, \hat{\boldsymbol{\theta}})',\ \ \hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$$

If $\{\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)\}$ is a mean-zero serially correlated ergodic-stationary process then

$$\begin{aligned} \mathbf{S} &= \mathsf{LRV} = \boldsymbol{\Gamma}_0 + 2\sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}_j') \\ \boldsymbol{\Gamma}_j &= E[\mathbf{g}_t(\mathbf{w}_t, \boldsymbol{\theta}_0)\mathbf{g}_t(\mathbf{w}_{t-j}, \boldsymbol{\theta}_0)'] \end{aligned}$$

and a consistent estimator has the form

$$\begin{aligned} \hat{\mathbf{S}}_{\mathsf{HAC}} &= \hat{\boldsymbol{\Gamma}}_0(\hat{\boldsymbol{\theta}}) + \sum_{j=1}^{n-1} k\left(\frac{j}{q(n)}\right)(\hat{\boldsymbol{\Gamma}}_j(\hat{\boldsymbol{\theta}}) + \hat{\boldsymbol{\Gamma}}_j'(\hat{\boldsymbol{\theta}})) \\ \hat{\boldsymbol{\Gamma}}_j(\hat{\boldsymbol{\theta}}) &= \frac{1}{n}\sum_{t=j+1}^{n} \mathbf{g}_t(\mathbf{w}_t, \hat{\boldsymbol{\theta}})\mathbf{g}_{t-j}(\mathbf{w}_{t-j}\hat{\boldsymbol{\theta}})' \\ & \hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0 \end{aligned}$$

The efficient GMM estimator may be computed using

- two-step procedure
- iterated procedure
- continuous updating procedure.

Application to log-normal SV model

Consider the alternative parameterization of the simple log-normal stochastic volatility (SV) model assuming $\mu = 0$:

$$
\begin{gathered}
r_t = \sigma_t u_t = \exp(w_t/2)u_t,\ t = 1, \ldots, n \\
w_t = \ln \sigma_t^2 = \omega + \phi w_{t-1} + \eta_{w,t} \\
(u_t, \eta_{w,t})' \sim \text{iid } N(\mathbf{0}, \operatorname{diag}(1, \sigma_{\eta_w}^2)) \\
\boldsymbol{\theta}_w = (\omega, \phi, \sigma_{\eta_w})'
\end{gathered}
$$

For $0 < \phi < 1$ and $\sigma_{\eta_w} \geq 0$, the series r_t is strictly stationary and ergodic, and unconditional moments of all orders exist.

The GMM estimation of the SV model is surveyed in Andersen and Sorensen (1996).

They recommended using moment conditions for GMM estimation based on lower-order moments of r_t, since higher-order moments tend to exhibit erratic finite sample behavior.

They considered a GMM estimation based on (subsets) of 24 moments considered by Jacquier, Polson, and Rossi (1994). To describe these moment conditions, first define

$$\alpha_w = \frac{\omega}{1-\phi},\ \beta_w^2 = \frac{\sigma_{\eta_w}^2}{1-\phi^2}$$

The moment conditions, which follow from properties of the log-normal distribution and the Gaussian AR(1) model, are expressed as

$$
\begin{aligned}
E[|r_t|] &= (2/\pi)^{1/2} E[\sigma_t] \\
E[r_t^2] &= E[\sigma_t^2] \\
E[|r_t^3|] &= 2\sqrt{2/\pi} E[\sigma_t^3] \\
E[r_t^4] &= 3E[\sigma_t^4] \\
E[|r_t r_{t-j}|] &= (2/\pi) E[\sigma_t \sigma_{t-j}],\ j = 1, \ldots, 10 \\
E[r_t^2 r_{t-j}^2] &= E[\sigma_t^2 \sigma_{t-j}^2],\ j = 1, \ldots, 10
\end{aligned}
$$

where for any positive integer j and positive constants p and s,

$$
\begin{aligned}
E[\sigma_t^p] &= \exp\left(\frac{p\alpha_w}{2} + \frac{p^2 \beta_w^2}{8}\right) \\
E[\sigma_t^p \sigma_{t-j}^s] &= E[\sigma_t^p] E[\sigma_t^s] \exp\left(\frac{ps\phi^j \beta_w^2}{4}\right)
\end{aligned}
$$

Let

$$\mathbf{w}_t = (|r_t|, r_t^2, |r_t^3|, r_t^4, |r_t r_{t-1}|, \ldots, \\ |r_t r_{t-10}|, r_t^2 r_{t-1}^2, \ldots, r_t^2 r_{t-10}^2)'$$

and define the 24×1 vector

$$g(\mathbf{w}_t, \boldsymbol{\theta}_w) = \begin{pmatrix} |r_t| - (2/\pi)^{1/2} \exp\left(\frac{\alpha_w}{2} + \frac{\beta_w^2}{8}\right) \\ r_t^2 - \exp\left(\alpha_w + \frac{\beta_w^2}{2}\right) \\ \vdots \\ r_t^2 r_{t-10}^2 - \exp\left(\alpha_w + \frac{\beta_w^2}{2}\right)^2 \exp\left(\phi^{10}\beta_w^2\right) \end{pmatrix}$$

Then, $E[g(\mathbf{w}_t, \boldsymbol{\theta}_w)] = \mathbf{0}$ is the population moment condition used for the GMM estimation of the model parameters $\boldsymbol{\theta}_w = (\alpha_w, \phi, \beta_w^2)'$.

Since the elements of $\mathbf{w}_t$ are serially correlated, the efficient weight matrix $\mathbf{S} =$ avar$(\bar{\mathbf{g}})$ must be estimated using an HAC estimator.

Review of Linear Gaussian State Space Models

State Space Models

Defn: A *state space model* for an $N-$dimensional time series $\mathbf{y}_t$ consists of a *measurement equation* relating the observed data to an $m-$ dimensional state vector $\boldsymbol{\alpha}_t$, and a Markovian transition equation that describes the evolution of the state vector over time.

The *measurement equation* has the form

$$\begin{aligned}
\underset{N\times 1}{\mathbf{y}_t} &= \underset{(N\times m)}{\mathbf{Z}_t}\underset{(m\times 1)}{\boldsymbol{\alpha}_t} + \underset{N\times 1}{\mathbf{d}_t} + \underset{N\times 1}{\boldsymbol{\varepsilon}_t}, \; t = 1,\ldots,T \\
\boldsymbol{\varepsilon}_t &\sim \text{iid } N(\mathbf{0}, \mathbf{H}_t)
\end{aligned}$$

The *transition equation* for the state vector $\boldsymbol{\alpha}_t$ is the first order Markov process

$$\begin{aligned}
\underset{m\times 1}{\boldsymbol{\alpha}_t} &= \underset{(m\times m)}{\mathbf{T}_t}\underset{(m\times 1)}{\boldsymbol{\alpha}_{t-1}} + \underset{m\times 1}{\mathbf{c}_t} + \underset{(m\times g)}{\mathbf{R}_t}\underset{(g\times 1)}{\boldsymbol{\eta}_t} \; t = 1,\ldots,T \\
\boldsymbol{\eta}_t &\sim \text{iid } N(\mathbf{0}, \mathbf{Q}_t)
\end{aligned}$$

For most applications, it is assumed that the measurement equation errors ε_t are independent of the transition equation errors

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\eta}_s'] = \mathbf{0} \text{ for all } s, t = 1, \ldots, T$$

The state space representation is completed by specifying the behavior of the initial state

$$\begin{aligned} \boldsymbol{\alpha}_0 &\sim N(\mathbf{a}_0, \mathbf{P}_0) \\ E[\boldsymbol{\varepsilon}_t \mathbf{a}_0'] &= \mathbf{0},\ E[\boldsymbol{\eta}_t \mathbf{a}_0'] = \mathbf{0} \text{ for } t = 1, \ldots, T \end{aligned}$$

The matrices $\mathbf{Z}_t, \mathbf{d}_t, \mathbf{H}_t, \mathbf{T}_t, \mathbf{c}_t, \mathbf{R}_t$ and $\mathbf{Q}_t$ are called the *system matrices*, and contain non-random elements. If these matrices do not depend deterministically on t the state space system is called *time invariant*.

Note: If $\mathbf{y}_t$ is covariance stationary, then the state space system will be time invariant.

Initial State Distribution for Covariance Stationary Models

If the state space model is covariance stationary, then the state vector $\boldsymbol{\alpha}_t$ is covariance stationary. The unconditional mean of $\boldsymbol{\alpha}_t$, $\mathbf{a}_0$, may be determined using

$$E[\boldsymbol{\alpha}_t] = \mathbf{T}E[\boldsymbol{\alpha}_{t-1}] + \mathbf{c} = \mathbf{T}E[\boldsymbol{\alpha}_t] + \mathbf{c}$$

Solving for $E[\boldsymbol{\alpha}_t]$, assuming $\mathbf{T}$ is invertible, gives

$$\mathbf{a}_0 = E[\boldsymbol{\alpha}_t] = (\mathbf{I}_m - \mathbf{T})^{-1}\mathbf{c}$$

Similarly, $var(\boldsymbol{\alpha}_0)$ may be determined using

$$\begin{aligned}
\mathbf{P}_0 &= \mathsf{var}(\boldsymbol{\alpha}_t) = \mathbf{T}\mathsf{var}(\boldsymbol{\alpha}_t)\mathbf{T}' + \mathbf{R}\mathsf{var}(\boldsymbol{\eta}_t)\mathbf{R}' \\
&= \mathbf{T}\mathbf{P}_0\mathbf{T}' + \mathbf{R}\mathbf{Q}\mathbf{R}'
\end{aligned}$$

Then,

$$\begin{aligned}\text{vec}(\mathbf{P}_0) &= \text{vec}(\mathbf{T}\mathbf{P}_0\mathbf{T}') + \text{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}') \\ &= (\mathbf{T} \otimes \mathbf{T})\text{vec}(\mathbf{P}_0) + \text{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}')\end{aligned}$$

which implies that

$$\text{vec}(\mathbf{P}_0) = (\mathbf{I}_{m^2} - \mathbf{T} \otimes \mathbf{T})^{-1}\text{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}')$$

Example: Unobserved component AR(2) model

$$\begin{aligned} y_t &= \mu + c_t \\ c_t &= \phi_1 c_{t-1} + \phi_2 c_{t-2} + \eta_t, \eta_t \sim \sigma^2 \end{aligned}$$

The state vector is $\boldsymbol{\alpha}_t = (c_t, c_{t-1})'$, which is unobservable, and the transition equation is

$$\begin{pmatrix} c_t \\ c_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \eta_t$$

so that

$$\mathbf{T} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \ \mathbf{R} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \mathbf{Q} = \sigma^2$$

This representation has measurement equation matrices

$$\mathbf{Z}_t = (1, 0), \ d_t = \mu, \varepsilon_t = 0, \mathbf{H}_t = 0$$

Distribution of initial state

$$\boldsymbol{\alpha}_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$$

Here $\boldsymbol{\alpha}_t = (c_t, c_{t-1})'$ is stationary and has mean zero

$$\mathbf{a}_0 = E[\boldsymbol{\alpha}_t] = \mathbf{0}$$

For the state variance, solve

$$\text{vec}(\mathbf{P}_0) = (\mathbf{I}_4 - \mathbf{T} \otimes \mathbf{T})^{-1}\text{vec}(\mathbf{RQR}')$$

Simple algebra gives

$$\mathbf{I}_4 - \mathbf{T} \otimes \mathbf{T} = \begin{pmatrix} 1-\phi_2^2 & -\phi_1\phi_2 & -\phi_1\phi_2 & -\phi_2^2 \\ -\phi_1 & 1 & -\phi_2 & 0 \\ -\phi_1 & -\phi_2 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$\text{vec}(\mathbf{RQR}') = \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and so

$$\text{vec}(\mathbf{P}_0) = \begin{pmatrix} 1-\phi_2^2 & -\phi_1\phi_2 & -\phi_1\phi_2 & -\phi_2^2 \\ -\phi_1 & 1 & -\phi_2 & 0 \\ -\phi_1 & -\phi_2 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The Kalman Filter

The *Kalman filter* is a set of recursion equations for determining the optimal estimates of the state vector $\boldsymbol{\alpha}_t$ given information available at time t, I_t. The filter consists of two sets of equations:

1. *Prediction equations*

2. *Updating equations*

To describe the filter, let

$$\begin{aligned}\mathbf{a}_t &= E[\boldsymbol{\alpha}_t|I_t] = \text{ optimal estimator of } \boldsymbol{\alpha}_t \text{ based on } I_t \\ \mathbf{P}_t &= E[(\boldsymbol{\alpha}_t - \mathbf{a}_t)(\boldsymbol{\alpha}_t - \mathbf{a}_t)'|I_t] = \text{ MSE matrix of } \mathbf{a}_t\end{aligned}$$

Prediction Equations

Given $\mathbf{a}_{t-1}$ and $\mathbf{P}_{t-1}$ at time $t-1$, the optimal predictor of $\boldsymbol{\alpha}_t$ and its associated MSE matrix are

$$\begin{aligned} \mathbf{a}_{t|t-1} &= E[\boldsymbol{\alpha}_t|I_{t-1}] = \mathbf{T}_t\mathbf{a}_{t-1} + \mathbf{c}_t \\ \mathbf{P}_{t|t-1} &= E[(\boldsymbol{\alpha}_t - \mathbf{a}_{t-1})(\boldsymbol{\alpha}_t - \mathbf{a}_{t-1})'|I_{t-1}] \\ &= \mathbf{T}_t\mathbf{P}_{t-1}\mathbf{T}'_{t-1} + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}'_t \end{aligned}$$

The corresponding optimal predictor of $\mathbf{y}_t$ give information at $t-1$ is

$$\mathbf{y}_{t|t-1} = \mathbf{Z}_t\mathbf{a}_{t|t-1} + \mathbf{d}_t$$

The *prediction error* and its MSE matrix are

$$\begin{aligned} \mathbf{v}_t &= \mathbf{y}_t - \mathbf{y}_{t|t-1} = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1} - \mathbf{d}_t \\ &= \mathbf{Z}_t(\boldsymbol{\alpha}_t - \mathbf{a}_{t|t-1}) + \boldsymbol{\varepsilon}_t \\ E[\mathbf{v}_t\mathbf{v}'_t] &= \mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}'_t + \mathbf{H}_t \end{aligned}$$

These are the components that are required to form the prediction error decomposition of the log-likelihood function.

Updating Equations

When new observations $\mathbf{y}_t$ become available, the optimal predictor $\mathbf{a}_{t|t-1}$ and its MSE matrix are updated using

$$\begin{aligned}\mathbf{a}_t &= \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}(\mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1} - \mathbf{d}_t) \\ &= \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{v}_t \\ \mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}_t\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1}\end{aligned}$$

The value $\mathbf{a}_t$ is referred to as the *filtered estimate* of $\boldsymbol{\alpha}_t$ and $\mathbf{P}_t$ is the MSE matrix of this estimate. It is the optimal estimate of $\boldsymbol{\alpha}_t$ given information available at time t.

Prediction Error Decomposition

Let $\boldsymbol{\theta}$ denote the parameters of the state space model. These parameters are embedded in the system matrices. For the state space model with a fixed value of $\boldsymbol{\theta}$, the Kalman Filter produces the prediction errors, $\mathbf{v}_t(\boldsymbol{\theta})$, and the prediction error variances, $\mathbf{F}_t(\boldsymbol{\theta})$, from the prediction equations. The *prediction error decomposition* of the log-likelihood function follows immediately:

$$\begin{aligned} \ln L(\boldsymbol{\theta}|\mathbf{y}) &= -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|\mathbf{F}_t(\boldsymbol{\theta})| \\ &\quad -\frac{1}{2}\sum_{t=1}^{T}\mathbf{v}_t'(\boldsymbol{\theta})\mathbf{F}_t^{-1}(\boldsymbol{\theta})\mathbf{v}_t(\boldsymbol{\theta}) \end{aligned}$$

State Space Representation of Log-Normal AR(1) SV model

Parameterization 1: let $y_t = r_t - \mu$.

$$\begin{aligned} y_t &= \sigma_t u_t \\ \ln \sigma_t &= \omega + \phi \ln \sigma_{t-1} + \eta_t \end{aligned}$$

Then

$$\begin{aligned} \ln |y_t| &= \ln \sigma_t + \ln |u_t| \\ E[\ln |u_t|] &= -0.63518, \ \text{var}(\ln |u_t|) = \pi^2/8 \end{aligned}$$

The measurement equation is

$$\ln |y_t| = -0.6358 + \ln \sigma_t + \xi_t, \ \xi_t \sim iid\ (0, \pi^2/8)$$

The transition equation is

$$\ln \sigma_t = \omega + \phi \ln \sigma_{t-1} + \eta_t, \ \eta_t \sim iid\ N(0, \sigma_\eta^2)$$

State space parameters

$$
\begin{aligned}
\alpha_t &= \ln \sigma_t = \text{state variable} \\
Z_t &= 1,\ d_t = -0.63518, T = \phi \\
a_0 &= \omega/(1-\phi),\ P_0 = \sigma_\eta^2/(1-\phi^2)
\end{aligned}
$$

Parameterization 2:

$$r_t = \exp(w_t/2)u_t,$$
$$w_t = \omega + \phi w_{t-1} + \eta_{w,t}$$

Then

$$\begin{aligned} \ln r_t^2 &= w_t + \ln u_t^2 \\ E[\ln u_t^2] &= -1.27, \ \mathrm{var}(\ln u_t^2) = \pi^2/2 \end{aligned}$$

The measurement equation is

$$\ln r_t^2 = -1.27 + w_t + \zeta_t, \ \zeta_t \sim iid \ (0, \pi^2/2)$$

The transition equation is

$$w_t = \omega + \phi w_{t-1} + \eta_{w,t}, \ \eta_{w,t} \sim iid \ N(0, \sigma^2_{\eta_w})$$

State space parameters

$$\begin{aligned}
\alpha_t &= \ln \sigma_t = \text{state variable} \\
Z_t &= 1,\ d_t = -1.27, T = \phi \\
a_0 &= \omega/(1-\phi),\ P_0 = \sigma^2_{\eta_w}/(1-\phi^2)
\end{aligned}$$

Complications

- ξ_t and ζ_t are not Gaussian random variables

- Kalman filter only provided minimum MSE linear estimators of state space parameters and state variables

- Quasi-MLE can be performed from prediction error decomposition of Gaussian log-likelihood

- Must use "sandwich asymptotic variance" for valid standard errors

- $r_t \approx 0 \Rightarrow \ln|r_t|$ and $\ln r_t^2 \approx -\infty$ which causes numerical problems

Transformation to improve numerical stability (Breidt and Carriquiry, 1996)

$$\begin{aligned} x_t &= \ln(r_t^2 + s) - \frac{s}{r_t^2 + s} \\ s &= \widehat{\text{var}}(r_t) \times 0.02 \end{aligned}$$

Log-Normal AR(1) SV Model with Student-t Errors

$$y_t = \sigma_t u_t = \sigma_t,\ u_t \sim iid\ St(v),\ v > 2$$
$$\ln(\sigma_t) - \alpha = \phi(\ln(\sigma_{t-1} - \alpha) + \eta_t,\ \eta_t \sim iid\ N(0, \sigma_\eta^2)$$
$$u_t \text{ is independent of } \eta_t \text{ for all } t$$

Since

$$\begin{aligned} u_t &= \sqrt{\omega_t} v_t \\ & v_t \sim iid\ N(0,1) \\ & (v-2)\omega_t^{-1} \sim \chi_v^2 \end{aligned}$$

It follows that

$$\begin{aligned} \ln|y_t| &= \ln\sigma_t + \frac{1}{2}\ln\omega_t + \ln|v_t| = \mu_\xi(v) + \ln\sigma_t + \xi_t^* \\ \xi_t^* &= \xi_t - \mu_\xi(v), \\ E[\xi_t] &= \mu_\xi(v) = \\ var(\xi_t^*) &= \sigma_\xi^2(v) = \end{aligned}$$