1 Model and Likelihood

Consider a linear K-factor model for D assets of the form

$$y_{it} = \alpha_d + \mathbf{f}_t' \boldsymbol{\beta}_d + \varepsilon_{it}$$

where $d = 1, \ldots, D$ and $t = 1, \ldots, T$ and $\mathbf{f}_t' = (f_{t1}, \ldots, f_{tK})$ is a $K \times 1$ vector. This is a special case of the seemingly unrelated regression (SUR) model in which the regressors are *identical* across equations. Stacking observations for a given time period across assets, define $\mathbf{y}_t' = (y_{1t}, \ldots, y_{Dt})$ and analogously $\boldsymbol{\varepsilon}_t' = (\varepsilon_{t1}, \ldots \varepsilon_{tD})$. Now let $\mathbf{x}_t' = (1, \mathbf{f}_t)$ and $\boldsymbol{\gamma}_d' = (\alpha_d, \boldsymbol{\beta}_d')$ so we have

$$\mathbf{y}_t = X_t \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_t$$

where $X_t = I_D \otimes \mathbf{x}_t'$ and $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}_1', \ldots, \boldsymbol{\gamma}_D')$. Now, suppose that

$$\boldsymbol{\varepsilon}_t | \mathbf{x}_t \sim \text{ iid } \mathcal{N}_D(0, \Sigma)$$

It follows that

$$L(\beta, \Sigma) \propto |\Sigma|^{-T/2} \exp\left[-\frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t - X_t \boldsymbol{\gamma})' \Sigma^{-1} (\mathbf{y}_t - X_t \boldsymbol{\gamma}) \right]$$

We parameterize this problem in terms of the $D \times D$ precision matrix Σ and the $D(K+1) \times 1$ vector of regression coefficients $\boldsymbol{\gamma}$.

2 Prior and Posterior Distribution

To complete the model we specify a prior distribution on the full collection of parameters $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \Omega^{-1})$, namely

$$\pi(\boldsymbol{\theta}) = \mathcal{N}_{d+p} (\gamma | \gamma_0, G_0) \, \mathcal{W}_d \left(\Omega^{-1} | \rho_0, R_0 \right)$$

This prior is conditionally conjugate with the normal likelihood yielding the posterior distribution $Y_n = (\{y_t\}, \{f_t\})$

$$\pi(\theta | Y_n) \propto \mathcal{N}_{d+p} (\gamma | \gamma_0, G_0) \, \mathcal{W}_d \left(\Omega^{-1} | \rho_0, R_0 \right) \prod_{t=1}^{n} \mathcal{N}_d(y_t | X_t \gamma, \Omega)$$

3 MCMC

We can summarize this posterior distribution by the MCMC algorithm of Chib and Greenberg (1995). It requires the repeated recursive sampling of the following two conditional distributions:

$$\gamma|Y_n, \Omega^{-1} \sim \mathcal{N}_{d+p}\left(\hat{\gamma}, G_n\right)$$

where

$$\hat{\gamma} = G_n\left(G_0^{-1}\gamma_0 + \sum_{t=1}^{n} X_t'\Omega^{-1}y_t\right)$$

$$G_n = \left(G_0^{-1} + \sum_{t=1}^{n} X_t'\Omega^{-1}X_t\right)^{-1}$$

and

$$\Omega^{-1}|Y_n, \gamma \sim \mathcal{W}_d\left(\rho_0 + n, \left(R_0^{-1} + \sum_{t=1}^{n}\left(y_t - X_t\gamma\right)\left(y_t - X_t\gamma\right)'\right)^{-1}\right)$$

4 Marginal likelihood

The marginal likelihood is available by the method of Chib (1995). From the Chib (1995) identity, we have

$$\log m\left(\{y_t\}\right) = \log \pi\left(\gamma^*\right) + \log \pi\left(\Omega^{-1*}\right) + \sum_{t=1}^{n} \log p\left(y_t|X_t\gamma^*, \Omega^*\right) - \log \pi\left(\gamma^*, \Omega^{-1*}|Y_n\right)$$

where the last term is calculated as

$$\pi\left(\Omega^{-1*}|Y_n\right) \times \pi\left(\gamma^*|Y_n, \Omega^{-1*}\right)$$

in which the first term is estimated by averaging the full-conditional Wishart density over the draws $\left\{\gamma^{(g)}\right\}_{g=1}^{G}$ from the main MCMC run

$$\pi\left(\Omega^{-1*}|Y_n\right) = \frac{1}{G}\sum_{g=1}^{G}\mathcal{W}_d\left(\Omega^{-1*}|\rho_0 + n, \left(R_0^{-1} + \sum_{t=1}^{n}\left(y_t - X_t\gamma^{(g)}\right)\left(y_t - X_t\gamma^{(g)}\right)'\right)^{-1}\right)$$

and the second term $\pi\left(\gamma^*|Y_n, \Omega^{-1*}\right)$ is available directly as

$$\pi(\gamma^*|Y_n, \Omega^{-1*}) = \mathcal{N}_{d+p}\left(\gamma^*|\hat{\gamma}^*, G_n^*\right)$$

where

$$\hat{\gamma}^* = G_n^*\left(G_0^{-1}\gamma_0 + \sum_{t=1}^{n} X_t'\Omega^{-1*}y_t\right)$$

$$G_n^* = \left(G_0^{-1} + \sum_{t=1}^{n} X_t'\Omega^{-1*}X_t\right)^{-1}$$

5   Prediction

Suppose we are interested in predicting the cross-section of returns y_{n+1} at time $(n+1)$. The Bayes prediction density of these returns, conditioned on the data Y_{n+1} and the factors f_{n+1}, is given by

$$p(y_{n+1}|Y_n, f_{n+1}) = \int_{\gamma, \Omega^{-1}} \mathcal{N}_d\left(y_{n+1}|X_{n+1}\gamma, \Omega\right) d\pi\left(\gamma, \Omega^{-1}|Y_n\right)$$

which is estimated by the ergodic Monte Carlo average

$$p(y_{n+1}|Y_n, f_{n+1}) = \frac{1}{G}\sum_{g=1}^{G} \mathcal{N}_d\left(y_{n+1}|X_{n+1}\gamma^{(g)}, \Omega^{(g)}\right)$$

with the MCMC draws $\left\{\gamma^{(g)}, \Omega^{(g)}\right\}$ from the posterior distribution.

6   Student-t errors

Suppose now that the errors are distributed as multivariate-t

$$\varepsilon_t \sim t_{d,\nu}\left(0, \Omega\right)$$

so that

$$E(\varepsilon_t) = 0\,, \ \nu > 1$$

$$Var\left(\varepsilon_t\right) = \frac{\nu}{\nu - 2}\Omega, \ \nu > 2$$

The analysis of this model utilizes the hierarchical reprentation

$$\varepsilon_t|\lambda_t \sim N\left(0, \lambda_t^{-1}\Omega\right)$$

$$\lambda_t \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

which means that conditioned on $(\nu, \{\lambda_t\})$, the results presented for the Gaussian model can be applied with minor modifications. The MCMC sampling is completed with the sampling of $(\nu, \{\lambda_t\})$.

Following Albert and Chib (1993), let us assume that the support of ν is the set of values $\{\nu_j\}_{j=1}^J$, for example, $\{4, 6, 8, 10, 12, 14, 16\}$ and that a priori

$$\Pr\left(\nu = \nu_j\right) = q_j$$

Then, simple calculations show that

$$\gamma|Y_n, \Omega^{-1}, \nu, \{\lambda_t\} \sim \mathcal{N}_{d+p}\left(\hat{\gamma}_\lambda, G_{n,\lambda}\right)$$

where

$$\hat{\gamma}_\lambda = G_{n,\lambda}\left(G_0^{-1}\gamma_0 + \sum_{t=1}^n \lambda_t X_t'\Omega^{-1}y_t\right)$$

$$G_{n,\lambda} = \left(G_0^{-1} + \sum_{t=1}^n \lambda_t X_t'\Omega^{-1}X_t\right)^{-1}$$

and

$$\Omega^{-1}|Y_n, \gamma, \nu, \{\lambda_t\} \sim \mathcal{W}_d\left(\rho_0 + n, \left(R_0^{-1} + \sum_{t=1}^n \lambda_t\left(y_t - X_t\gamma\right)\left(y_t - X_t\gamma\right)'\right)^{-1}\right)$$

Moreover,

$$\Pr\left(\nu = \nu_j|Y_n, \gamma, \Omega^{-1}\right) \propto q_j \prod_{t=1}^n t_{d,\nu_j}\left(y_t|X_t\gamma, \Omega\right)$$

and

$$\lambda_t|Y_n, \gamma, \nu \sim G\left(\frac{\nu + d}{2}, \frac{\nu + \left(y_t - X_t\gamma\right)'\left(y_t - X_t\gamma\right)}{2}\right)$$

One sweep of the MCMC sampling is completed by sampling these four distributions in this order.

4

6.1 Marginal likelihood

The Chib (1995) method can again be applied to find the log marginal likelihood as

$$\log \Pr(\nu^*) + \log \pi(\gamma^*) + \log \pi\left(\Omega^{-1*}\right) + \sum_{t=1}^{n} \log t_{d,\nu^*}\left(y_t | X_t \gamma^*, \Omega^*\right) - \log \pi\left(\nu^*, \gamma^*, \Omega^{-1*} | Y_n\right)$$

where ν^* is the posterior mode (which is easily computed from the sampled values), the last term is calculated as

$$\Pr(\nu^* | Y_n) \times \pi\left(\Omega^{-1*} | Y_n, \nu^*\right) \times \pi\left(\gamma^* | Y_n, \Omega^{-1*}, \nu^*\right)$$

in which the first term is obtained from the posterior frequency distribution of ν, the second term is obtained from a reduced run in which ν is fixed at ν^* and the remaining three distributions are sampled and the draws

$$\left\{\gamma^{(g)}, \lambda_t^{(g)}\right\}_{g=1}^{G}$$

from this reduced MCMC run are used to calculate $\pi\left(\Omega^{-1*} | Y_n, \nu^*\right)$ as

$$\frac{1}{G} \sum_{g=1}^{G} \mathcal{W}_d\left(\Omega^{-1*} | \rho_0 + n, \left(R_0^{-1} + \sum_{t=1}^{n} \lambda_t^{(g)}\left(y_t - X_t\gamma^{(g)}\right)\left(y_t - X_t\gamma^{(g)}\right)'\right)^{-1}\right)$$

and the final term $\pi\left(\gamma^* | Y_n, \Omega^{-1*}\right)$ is obtained from a second reduced run in which ν is fixed at ν^* and Ω^{-1} is fixed at Ω^{-1*} and the draws

$$\left\{\lambda_t^{(g)}\right\}$$

from this reduced run are used to give

$$\pi(\gamma^* | Y_n, \Omega^{-1*}, \nu^*) = \frac{1}{G} \sum_{g=1}^{G} \mathcal{N}_{d+p}\left(\gamma^* | \hat{\gamma}_{\lambda^{(g)}}^*, G_{n,\lambda^{(g)}}^*\right)$$

where

$$\hat{\gamma}_{\lambda^{(g)}}^* = G_{n,\lambda^{(g)}}^*\left(G_0^{-1}\gamma_0 + \sum_{t=1}^{n} \lambda_t^{(g)} X_t' \Omega^{-1*} y_t\right)$$

$$G_{n,\lambda^{(g)}}^* = \left(G_0^{-1} + \sum_{t=1}^{n} \lambda_t^{(g)} X_t' \Omega^{-1*} X_t\right)^{-1}$$