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

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

Let Y_T denote the full data sample, i.e. $\{\mathbf{y}_t, \mathbf{x}_t\}_{t=1}^T$. Then the likelihood is

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

where we parameterize this problem in terms of the $D \times D$ *precision* matrix Ω^{-1} and the $p \times 1$ vector of regression coefficients $\boldsymbol{\gamma}$, where $p = D(K+1)$.

2 Prior and Posterior Distribution

To complete the model we specify the following prior distribution

$$\pi(\boldsymbol{\gamma}, \Omega^{-1}) = \mathcal{N}_p(\boldsymbol{\gamma} | \boldsymbol{\gamma}_{\mathbf{0}}, G_0) \, \mathcal{W}_D\left(\Omega^{-1} | \rho_0, R_0\right)$$

This prior is conditionally conjugate with the normal likelihood. In particular, we have $\boldsymbol{\gamma} | \Omega^{-1}, Y_T \sim \mathcal{N}_p(\bar{\boldsymbol{\gamma}}, G_T)$ where

$$\begin{aligned} G_T &= \left[G_0^{-1} + \sum_{t=1}^{T} X_t' \Omega^{-1} X_t \right]^{-1} \\ \bar{\boldsymbol{\gamma}} &= G_T \left[G_0^{-1} \boldsymbol{\gamma}_0 + \sum_{t=1}^{T} X_t' \Omega^{-1} \mathbf{y}_t \right] \end{aligned}$$

and $\Omega^{-1}|Y_T \sim \mathcal{W}_D\left(\rho_0 + T, R_T\right)$ where

$$R_T = \left[R_0^{-1} + \sum_{t=1}^{T} \left(\mathbf{y}_t - X_t\boldsymbol{\gamma}\right)\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}\right)'\right]^{-1}$$

3 MCMC

Using the full set of conditional posteriors, given in the preceding section, we can simulate from the joint posterior for this model using a Gibbs sampler:

1. Select a starting value $\Omega^{-1(0)}$ for the precision matrix.
2. Draw $\boldsymbol{\gamma}^{(1)} \sim \mathcal{N}\left(\bar{\boldsymbol{\gamma}}^{(1)}, G_T^{(1)}\right)$ where

$$\begin{aligned} G_T^{(1)} &= \left[G_0^{-1} + \sum_{t=1}^{T} X_t'\Omega^{-1(0)}X_t\right]^{-1} \\ \bar{\boldsymbol{\gamma}}^{(1)} &= G_T^{(1)}\left[G_0^{-1}\boldsymbol{\gamma}_0 + \sum_{t=1}^{T} X_t'\Omega^{-1(0)}\mathbf{y}_t\right] \end{aligned}$$

3. Draw $\Omega^{-1(1)} \sim \mathcal{W}_D\left(\rho_T, R_T^{(1)}\right)$ where

$$R_T^{(1)} = \left[R_0^{-1} + \sum_{t=1}^{T} \left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^{(1)}\right)\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^{(1)}\right)'\right]^{-1}$$

4. Repeat the preceding two steps a total of G times. In the gth iteration:

   (i) Draw $\boldsymbol{\gamma}^{(g)} \sim \mathcal{N}\left(\bar{\boldsymbol{\gamma}}^{(g)}, G_T^{(g)}\right)$ where

$$\begin{aligned} G_T^{(g)} &= \left[G_0^{-1} + \sum_{t=1}^{T} X_t'\Omega^{-1(g-1)}X_t\right]^{-1} \\ \bar{\boldsymbol{\gamma}}^{(g)} &= G_T^{(g)}\left[G_0^{-1}\boldsymbol{\gamma}_0 + \sum_{t=1}^{T} X_t'\Omega^{-1(g-1)}\mathbf{y}_t\right] \end{aligned}$$

(ii) Draw $\Omega^{-1(g)} \sim \mathcal{W}_D\left(\rho_T, R_T^{(g)}\right)$ where

$$R_T^{(g)} = \left[R_0^{-1} + \sum_{t=1}^{T} \left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^{(g)}\right)\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^{(g)}\right)'\right]^{-1}$$

5. Discard the first B draws.

Note that in iteration g, $G_T^{(g)}$ and $\bar{\boldsymbol{\gamma}}^{(g)}$ are calculated using $\Omega^{-1(g-1)}$ while $R_T^{(g)}$ is calculated using $\boldsymbol{\gamma}^{(0)}$. This is because we choose to initialize the sample with a starting value $\Omega^{-1(0)}$ for the precision matrix rather than for the vector of regression coefficients.

4 Computational Details

4.1 Evaluating the MV Normal Density

As one of the steps in the calculation of the marginal likelihood (see below) we will need to repeatedly evaluate the log of a multivariate normal density at a fixed set of parameter values. Let Z be a $p \times n$ matrix, each of whose columns is a point $\mathbf{z}$ at which we wish to evaluate $\log \mathcal{N}_p\left(\mathbf{z}|\mu, \Sigma\right)$ where μ is the mean vector and Σ the covariance matrix of a multivariate normal. Because our problem is parameterized in terms of the *precision* matrix rather than the covariance matrix, the calculations given here assume that we are given Σ^{-1} rather than Σ. In terms of the precision matrix, the log of the MV normal density is given by

$$\log \mathcal{N}_p\left(\mathbf{z}|\mu, \Sigma^{-1}\right) = -\frac{p}{2}\log\left(2\pi\right) + \frac{1}{2}\log\left|\Sigma^{-1}\right| - \frac{1}{2}\left(\mathbf{z}-\mu\right)'\Sigma^{-1}\left(\mathbf{z}-\mu\right)$$

Now let R be the Cholesky factor of Σ^{-1} so that $\Sigma^{-1} = R'R$ and define $\tilde{\mathbf{z}} = \mathbf{z} - \mu$ and $\mathbf{v} = R\tilde{\mathbf{z}}$. Using these definitions,

$$\left(\mathbf{z}-\mu\right)'\Sigma^{-1}\left(\mathbf{z}-\mu\right) = (R\tilde{\mathbf{z}})'(R\tilde{\mathbf{z}}) = \mathbf{v}'\mathbf{v}$$

and, letting R_{ii} denote the ith diagonal element of R,

$$\begin{aligned}\frac{1}{2}\log|\Sigma^{-1}| &= \frac{1}{2}\log|R'R| = \frac{1}{2}\log\left(|R'|\cdot|R|\right) = \frac{1}{2}\left(\log|R'| + \log|R|\right) \\ &= \frac{1}{2}\left(2\log|R|\right) = \sum_{i=1}^{p}\log R_{ii}\end{aligned}$$

since $|A| = |A'|$, $|AB| = |A|\cdot|B|$ and the determinant of a triangular matrix equals the product of its diagonal elements. Thus, we have

$$\log\mathcal{N}_p\left(\mathbf{z}|\mu,\Sigma^{-1}\right) = -\frac{p}{2}\log\left(2\pi\right) + \text{trace}\left[\log\left(\text{diag}\left\{R\right\}\right)\right] - \frac{1}{2}\mathbf{v}'\mathbf{v}$$

The only term in the preceding expression that depends on $\mathbf{z}$ is $\mathbf{v}'\mathbf{v}$. We can calculate this term simultaneously for all columns of Z as follows. First let $\widetilde{Z}$ denote the result subtracting of subtracting the vector μ from each column of Z, i.e. $\widetilde{Z} = Z - \mu\mathbf{1}_n'$. To calculate $\mathbf{v}'\mathbf{v}$ for each column of Z we simply square the elements of $R\widetilde{Z}$ and take the column sums of the resulting matrix.

4.2 Efficient Calculation of R_T

In the second step of each iteration we compute $\left(R_0^{-1} + \sum_{t=1}^{T}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right)^{-1}$ where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - X_t\boldsymbol{\gamma}$. Since R_0 is simply the prior scale matrix for Ω^{-1} and hence remains unchanged during the iterations, we can pre-compute it and store the result before starting the sampler. Since X_t is a sparse matrix, there is a much more efficient and compact way to compute the sum of outer products of residuals. Define:

$$\widetilde{Y} = \begin{bmatrix}\mathbf{y}_1' \\ \vdots \\ \mathbf{y}_T'\end{bmatrix}, \quad \widetilde{X} = \begin{bmatrix}\mathbf{x}_1' \\ \vdots \\ \mathbf{x}_T'\end{bmatrix}, \quad \Gamma = \begin{bmatrix}\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\gamma}_D\end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix}\boldsymbol{\varepsilon}_1' \\ \vdots \\ \boldsymbol{\varepsilon}_T'\end{bmatrix}$$

so that $\boldsymbol{\varepsilon} = \widetilde{Y} - \widetilde{X}\Gamma$. Note that the vector of regression coefficients $\boldsymbol{\gamma}$ is the vec of the *matrix* of regression coefficients Γ. Thus, expressed in terms of dense matrix operations

$$R_T^{-1} = R_0^{-1} + \left(\widetilde{Y} - \widetilde{X}\Gamma\right)'\left(\widetilde{Y} - \widetilde{X}\Gamma\right)$$

The final step is to invert this sum (which is positive definite) to calculate R_T. Note that the Matrix Inversion Lemma (Sherman-Morrison-Woodbury Formula) does *not* simplify this calculation unless $D > T$.

4.3 Efficient Calculation of G_T

Because we parameterize our multivariate normal sampler in terms of the *precision* matrix rather than the covariance matrix, we work with the *inverse* of G_T, namely

$$G_T^{-1} = G_0^{-1} + \sum_{t=1}^{T} X_t' \Omega^{-1} X_t$$

Since it is simply the prior precision matrix for the vector $\boldsymbol{\gamma}$ of regression coefficients we can pre-compute G_0^{-1} (assuming that we elicit a prior in terms of the covariance matrix). Now, the sum over $X_t' \Omega X_t$ can in fact be simplified using the properties of the Kronecker product.[1] Recall that $X_t = I_D \otimes \mathbf{x}_t'$. Since $(A \otimes B)' = A' \otimes B'$,

$$X_t' \Omega^{-1} X_t = (I_D \otimes \mathbf{x}_t) \, \Omega^{-1} X_t$$

Since $\Omega^{-1} X_t = (\Omega^{-1} X_t) \otimes 1$, $\Omega^{-1} = \Omega^{-1} \otimes 1$, and $(A \otimes B)(C \otimes D) = AC \otimes BD$, provided that everything is conformable, we have

$$\begin{aligned}
(I_D \otimes \mathbf{x}_t) \, \Omega^{-1} X_t &= (I_D \otimes \mathbf{x}_t) \, (\Omega^{-1} X_t \otimes 1) \\
&= \Omega^{-1} X_t \otimes \mathbf{x}_t = \left[\Omega^{-1} (I_D \otimes \mathbf{x}_t')\right] \otimes \mathbf{x}_t \\
&= \left[(\Omega^{-1} \otimes 1)(I_D \otimes \mathbf{x}_t')\right] \otimes \mathbf{x}_t \\
&= \Omega^{-1} \otimes \mathbf{x}_t' \otimes \mathbf{x}_t
\end{aligned}$$

Finally, since $A \otimes (B + C) = A \otimes B + A \otimes C$,

$$\sum_{t=1}^{T} X_t' \Omega^{-1} X_t = \sum_{t=1}^{T} \Omega^{-1} \otimes \mathbf{x}_t' \otimes \mathbf{x}_t = \Omega^{-1} \otimes \left(\sum_{t=1}^{T} \mathbf{x}_t' \otimes \mathbf{x}_t \right)$$

[1] See, e.g., Horn and Johnson (1994) Chapter 4.2.

This is an extremely useful simplification: because $\sum_{t=1}^{T} \mathbf{x}_t' \otimes \mathbf{x}_t$ involves neither Ω^{-1} nor $\boldsymbol{\gamma}$, only the data, we can pre-compute this quantity. In fact, there is one final simplification that makes this quantity even simpler. By writing out the definition of the Kronecker Product, we see that $\mathbf{x}_t' \otimes \mathbf{x}_t = \mathbf{x}_t \mathbf{x}_t'$ and hence

$$\sum_{t=1}^{T} X_t' \Omega^{-1} X_t = \Omega^{-1} \otimes \left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right) = \Omega^{-1} \otimes \widetilde{X}' \widetilde{X}$$

where $\widetilde{X}' = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_T \end{bmatrix}$. Thus, we have $G_T^{-1} = G_0^{-1} + \Omega^{-1} \otimes \widetilde{X}' \widetilde{X}$.

4.4 Efficient Calculation of $\bar{\gamma}$

The vector $\bar{\boldsymbol{\gamma}}$ is constructed from several pieces. The first is $G_0^{-1}\gamma_0$, the solution to the linear system $G_0 \mathbf{v} = \gamma_0$. Since this piece depends only on the prior, we can pre-compute it. The next piece is the sum $\sum_{t=1}^{T} X_t' \Omega^{-1} \mathbf{y}_t$. We we noted above, X_t is sparse so there is a more efficient way to compute this quantity. Indeed, while this is far from obvious at first glance, it is possible to *factor* Ω^{-1} outside of the sum using some clever matrix operations, allowing us to drastically reduce the computational complexity of the sampler. To accomplish this simplification we combine the definition of X_t as $I_D \otimes \mathbf{x}_t'$ with two properties of the Kronecker Product, namely:

$$(A \otimes B)(C \otimes D) = AC \otimes BD$$

which holds provided that the respective matrices are conformable and

$$\operatorname{vec}(AB) = (B' \otimes I_k) \operatorname{vec}(A)$$

where A is $k \times \ell$ and B is $\ell \times m$. Applying the first property twice in succession followed by the second property, we find that

$$\begin{aligned} X_t' \Omega^{-1} \mathbf{y}_t &= \left(I_D \otimes \mathbf{x}_t'\right)' \Omega^{-1} \mathbf{y}_t = \left(I_D \otimes \mathbf{x}_t\right) \Omega^{-1} \mathbf{y}_t \\ &= \left(I_D \otimes \mathbf{x}_t\right) \left(\Omega^{-1} \mathbf{y}_t \otimes 1\right) = I_D \Omega^{-1} \mathbf{y}_t \otimes \mathbf{x}_t 1 \\ &= \Omega^{-1} \mathbf{y}_t \otimes \mathbf{x}_t = \left[\left(\Omega^{-1} \mathbf{y}_t\right) 1\right] \otimes \left[I_{K+1} \mathbf{x}_t\right] \\ &= \left(\Omega^{-1} \mathbf{y}_t \otimes I_{K+1}\right) \left(1 \otimes \mathbf{x}_t\right) \\ &= \left(\left[\Omega^{-1} \mathbf{y}_t\right] \otimes I_{K+1}\right) \mathbf{x}_t = \left(\left[\mathbf{y}_t' \Omega^{-1}\right]' \otimes I_{K+1}\right) \mathrm{vec}(\mathbf{x}_t) \\ &= \mathrm{vec}\left(\mathbf{x}_t \mathbf{y}_t' \Omega^{-1}\right) \end{aligned}$$

where we have used the fact that $\mathrm{vec}\left(\mathbf{x}_t\right) = \mathbf{x}_t$. Finally, since we can interchange the vec summation operations,

$$\begin{aligned} \sum_{t=1}^{T} \mathrm{vec}\left(\mathbf{x}_t \mathbf{y}_t' \Omega^{-1}\right) &= \mathrm{vec}\left[\sum_{t=1}^{T} \mathbf{x}_t \mathbf{y}_t' \Omega^{-1}\right] = \mathrm{vec}\left[\left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{y}_t'\right) \Omega^{-1}\right] \\ &= \mathrm{vec}\left(\widetilde{X}' \widetilde{Y} \Omega^{-1}\right) \end{aligned}$$

where, as above,

$$\widetilde{Y} = \begin{bmatrix} \mathbf{y}_1' \\ \vdots \\ \mathbf{y}_T' \end{bmatrix}, \quad \widetilde{X} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_T' \end{bmatrix}$$

Thus we see that

$$\bar{\boldsymbol{\gamma}} = G_T \left[G_0^{-1} \boldsymbol{\gamma}_0 + \mathrm{vec}\left(\widetilde{X}' \widetilde{Y} \Omega^{-1}\right)\right]$$

Because it does not change between iterations, we can pre-compute the product $\widetilde{X}' \widetilde{Y}$. The only term that remains to be addressed is G_T. Because our normal sampler is parameterized in terms of the precision matrix rather than the covariance matrix we calculated G_T^{-1} rather than G_T above. Rather than inverting it in this step, which is a very bad idea given its size, we notice that our expression for $\bar{\boldsymbol{\gamma}}$ takes the form $\mathbf{v} = A^{-1}\mathbf{b}$. Therefore,

$$\bar{\boldsymbol{\gamma}} = \mathrm{solve}\left[G_T^{-1},\ G_0^{-1} \boldsymbol{\gamma}_0 + \mathrm{vec}\left(\widetilde{X}' \widetilde{Y} \Omega^{-1}\right)\right]$$

5 Calculating the Marginal likelihood

We calculate the marginal likelihood using the method of Chib (1995). Let θ denote the full collection of parameters. By Bayes' Rule

$$\pi\left(\theta|Y_T\right) = \frac{\pi(\theta)f(Y_T|\theta)}{f(Y_T)}$$

where $f(Y_T)$ is the marginal likelihood, aka the marginal data density, aka the evidence. This identity holds true for *any* value of θ. In particular it holds at the posterior mean θ^*. Solving for $f(Y_T)$ and evaluating the result at θ^*, we have

$$f\left(Y_T\right) = \frac{\pi(\theta^*)f(Y_T|\theta^*)}{\pi(\theta^*|Y_T)}$$

Thus, we can express the *log* marginal likelihood as

$$\log f(Y_T) = \log \pi(\theta^*) + \log f\left(Y_T|\theta^*\right) - \log \pi\left(\theta^*|Y_T\right)$$

Specializing this to the SUR model considered above,

$$\log f(Y_T) = \log \pi(\boldsymbol{\gamma}^*) + \log \pi\left(\Omega^{-1*}\right) + \log f\left(Y_T|\boldsymbol{\gamma}^*, \Omega^{-1*}\right) - \log \pi\left(\boldsymbol{\gamma}^*, \Omega^{-1*}|Y_T\right)$$

since our priors over $\boldsymbol{\gamma}$ and Ω^{-1} are independent. The Chib (1995) method approximates $\log f(Y_T)$ by evaluating each of the terms on the right-hand-side of the preceding expression using the output of the Gibbs sampler.

The Contribution of the Prior Evaluating the first two terms, $\log \pi(\boldsymbol{\gamma}^*)$ and $\log \pi\left(\Omega^{-1*}\right)$, is easy: these are simply the priors for $\boldsymbol{\gamma}$ and Ω^{-1} evaluated at the posterior means. We take the sample average of the Gibbs draws to approximate $\boldsymbol{\gamma}^*$ and Ω^{-1*} and evaluate the Normal and Wishart distributions at these points, with parameters given by the prior:

$$\begin{aligned} \pi\left(\boldsymbol{\gamma}^*\right) &= \mathcal{N}_p\left(\boldsymbol{\gamma}^*|\boldsymbol{\gamma}_0, G_0\right) \\ \pi\left(\Omega^{-1*}\right) &= \mathcal{W}_D\left(\Omega^{-1*}|\rho_0, R_0\right) \end{aligned}$$

The Contribution of the Likelihood Above we assumed a normal distribution for the regression errors, specifically, $\boldsymbol{\varepsilon}_t|\mathbf{x}_t \sim$ iid $\mathcal{N}_D(0,\Omega)$. From the regression specification it follows that $\mathbf{y}_t \sim$ iid $\mathcal{N}_D\left(X_t\boldsymbol{\gamma},\Omega\right)$ and thus the log likelihood evaluated at the posterior mean is

$$\log f(Y_T|\boldsymbol{\gamma}^*,\Omega^{-1*}) = \sum_{t=1}^{T} \log \mathcal{N}_D\left(\mathbf{y}_t|X_t\boldsymbol{\gamma}^*,\Omega^{-1*}\right)$$

parameterized in terms of the precision matrix rather than the covariance matrix. Equivalently, but more conveniently, we may write

$$\log f(Y_T|\boldsymbol{\gamma}^*,\Omega^{-1*}) = \sum_{t=1}^{T} \log \mathcal{N}_D\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^*|\mathbf{0},\Omega^{-1*}\right)$$

The advantage of this version of the likelihood is that the parameters of the normal density are constant over t, allowing us to exploit the efficient algorithm for repeatedly evaluating a MV normal density with fixed parameters, described above. Note that we can simultaneously calculate all of the arguments for the normal density as follows:

$$(\widetilde{Y} - \widetilde{X}\Gamma^*)' = (\boldsymbol{\varepsilon}^*)' = \left[\begin{array}{ccc} \boldsymbol{\varepsilon}_1^* & \ldots & \boldsymbol{\varepsilon}_T^* \end{array}\right]$$

where $\boldsymbol{\varepsilon}_t^* = \mathbf{y}_t - X_t\boldsymbol{\gamma}^*$ and $\Gamma^* = (\boldsymbol{\gamma}_1^*,\ldots,\boldsymbol{\gamma}_D^*)$.

The Contribution of the Posterior To evaluate the third term, we factorize the joint posterior as the product of a conditional and marginal, namely:

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

so that we have

$$\log \pi\left(\boldsymbol{\gamma}^*,\Omega^{-1*}|Y_T\right) = \log \pi\left(\boldsymbol{\gamma}^*|\Omega^{-1*},Y_T\right) + \log \pi\left(\Omega^{-1*}|Y_T\right)$$

Because we have *analytical expressions* for the conditional posteriors in this model we can evaluate the first term in the product immediately. We have

$\boldsymbol{\gamma}|\Omega^{-1} \sim \mathcal{N}_p\left(\boldsymbol{\gamma}|\bar{\boldsymbol{\gamma}}, G_T\right)$ where G_T and $\bar{\gamma}$ depend only on the prior, the data, and Ω^{-1}. To perform the required calculation, we simply evaluate the normal density at $\boldsymbol{\gamma}^*$ and evaluate G_T and $\bar{\gamma}$ at Ω^{-1*}, that is:

$$\pi\left(\boldsymbol{\gamma}^*|\Omega^{-1*}, Y_T\right) = \mathcal{N}_p\left(\boldsymbol{\gamma}^*|\bar{\gamma}^*, G_T^{-1*}\right)$$

where

$$\begin{aligned} G_T^{-1*} &= \left[G_0^{-1} + \sum_{t=1}^{T} X_t' \Omega^{-1*} X_t\right] = \left[G_0^{-1} + \Omega^{-1*} \otimes \widetilde{X}'\widetilde{X}\right] \\ \bar{\boldsymbol{\gamma}}^* &= G_T^* \left[G_0^{-1}\boldsymbol{\gamma}_0 + \sum_{t=1}^{T} X_t' \Omega^{-1*} \mathbf{y}_t\right] = \text{solve}\left[G_T^{-1*},\ G_0^{-1}\gamma_0 + \text{vec}\left(\widetilde{X}'\widetilde{Y}\Omega^{-1*}\right)\right] \end{aligned}$$

The evaluation of the second term in the product that gives the contribution of the posterior to the marginal likelihood is a bit more involved. We write

$$\begin{aligned} \pi\left(\Omega^{-1*}|Y_T\right) &= \int \pi\left(\boldsymbol{\gamma}, \Omega^{-1*}|Y_T\right)\ d\boldsymbol{\gamma} \\ &= \int \pi\left(\Omega^{-1*}|\boldsymbol{\gamma}, Y_T\right) \pi\left(\boldsymbol{\gamma}|Y_T\right)\ d\boldsymbol{\gamma} \end{aligned}$$

and approximate the second integral using the draws from the Gibbs sampler:

$$\begin{aligned} \pi\left(\Omega^{-1*}|Y_T\right) &\approx \frac{1}{G}\sum_{g=1}^{G} \pi\left(\Omega^{-1*}|\boldsymbol{\gamma}^{(g)}, Y_T\right) \\ &= \frac{1}{G}\sum_{g=1}^{G} \mathcal{W}_D\left(\Omega^{-1*}\left|\rho_0 + T, R_T^{(g)}\right.\right) \end{aligned}$$

where

$$\begin{aligned} R_T^{(g)} &= \left[R_0^{-1} + \sum_{t=1}^{T}\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^{(g)}\right)\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}^{(g)}\right)'\right]^{-1} \\ &= \left[R_0^{-1} + \left(\widetilde{Y} - \widetilde{X}\Gamma^{(g)}\right)'\left(\widetilde{Y} - \widetilde{X}\Gamma^{(g)}\right)\right]^{-1} \end{aligned}$$

6 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.

7 Gibbs Sampler with Student-t Errors

7.1 A Hierarchical Representation

Suppose now that the errors follow a multivariate Student-t distribution rather than a normal distribution:

$$\boldsymbol{\varepsilon}_t \sim t_{D,\nu}\left(0, \Omega\right)$$

where ν denotes the degrees of freedom of the distribution, the location parameter is zero and the scale matrix is Ω. If $\nu > 1$ then $E(\boldsymbol{\varepsilon}) = 0$. If $\nu > 2$ then $Var\left(\boldsymbol{\varepsilon}\right) = \nu\sigma/(n-2)$. Replacing the normal likelihood from above with the Student-t likelihood, however, breaks the conditional conjugacy that we exploited above to construct an MCMC algorithm based on the Gibbs sampler. The solution to this problem is to work with a hierarchical representation in which the Student-t likelihood is introduced as a scale mixtrure of normal distributions, in particular

$$\boldsymbol{\varepsilon}_t|\lambda_t \sim N\left(0, \lambda_t^{-1}\Omega\right)$$
$$\lambda_t \sim G\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

where $G(\alpha, \beta)$ denotes the Gamma distribution with shape parameter α and rate parameter β. (See below for more discussion on the parameterization of the gamma distribution.) Using this representation, after conditioning on $(\nu, \boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots \lambda_T)'$, we are essentially back in the familiar normal case from above. In any inference that we carry out, as well as in the calculation of the marginal likelihood, we will marginalize over $\boldsymbol{\lambda}$ by simply ignoring these draws. We will continue to place a normal prior on $\boldsymbol{\gamma}$ and a Wishart prior on Ω^{-1}, exactly as we did when working with the normal likelihood above.

7.2 The Sampler

The sampler proceeds by fixing the degrees of freedom parameter ν. If ν is to be chosen from the data, this can be accomplished using the marginal likelihood, as described below. Holding ν fixed, the full set of conditional posteriors is as follows:

Regression Coefficients: $\boldsymbol{\gamma}|\Omega^{-1}, Y_T \sim \mathcal{N}_p\left(\bar{\boldsymbol{\gamma}}_\lambda, G_{T,\lambda}\right)$

$$\begin{aligned} G_{T,\lambda} &= \left[G_0^{-1} + \sum_{t=1}^{T} \lambda_t X_t' \Omega^{-1} X_t\right]^{-1} \\ \bar{\boldsymbol{\gamma}}_\lambda &= G_{T,\lambda}\left[G_0^{-1}\boldsymbol{\gamma}_0 + \sum_{t=1}^{T} \lambda_t X_t' \Omega^{-1} \mathbf{y}_t\right] \end{aligned}$$

Inverse Scale Matrix: $\Omega^{-1}|Y_T \sim \mathcal{W}_D\left(\rho_0 + T, R_{T,\lambda}\right)$

$$R_{T,\lambda} = \left[R_0^{-1} + \sum_{t=1}^{T} \lambda_t \left(\mathbf{y}_t - X_t\boldsymbol{\gamma}\right)\left(\mathbf{y}_t - X_t\boldsymbol{\gamma}\right)'\right]^{-1}$$

Auxiliary Parameter: $\lambda_t|\boldsymbol{\gamma}, \nu, Y_T \sim G\left(\frac{\nu + D}{2}, \frac{\nu + \boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t}{2}\right), \boldsymbol{\varepsilon}_t = \mathbf{y}_t - X_t\boldsymbol{\gamma}$

To implement the Gibbs sampler, we simply need to draw sequentially from these distributions, in the order given above. We will require, however,

starting values for both Ω^{-1} and each of the λ_t parameters. A reasonable starting value for λ_t is one, which makes the initial draws for the regression coefficients and the inverse scale matrix the same as if were work withing with the normal model.

8 Computational Details for Student-t Model

8.1 Parameterizing the Gamma Distribution

The Gamma distribution is can be parameterized in two different ways. The parameterization upon which the algorithms described above are based uses $G(\alpha, \beta)$ to denote the density

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where α is the shape parameter and β is the rate parameter. The base R function for drawing from this distribution is parameterized as follows:

```
rgamma(n, shape, rate = 1, scale = 1/rate)
```

so that we have a choice of specifying *either* the rate parameter *or* its reciprocal, which is called the *scale parameter*. If we let $s = 1/\beta$ denote the scale parameter, an alternative parameterization of the density is given by

$$f(x|\alpha, s) = \frac{1}{s^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/s}$$

The C function that underlies `rgamma` (see `Rmath.h`) is parameterized according to

```
double rgamma(double a, double scl)
```

so we must specify the *scale* parameter if we want to call this from C++.

8.2 Evaluating the Multivariate Student-t Density

As one of the steps in the calculation of the marginal likelihood we need to repeatedly evaluate the log of a multivariate Student-t density at a fixed set of parameter values. Let Z be a $p \times n$ matrix, each of whose columns is a point $\mathbf{z}$ at which we wish to evaluate the log density. The expression for the density itself is

$$t_p\left(\mathbf{z}|\nu,\mu,\Sigma\right) = \frac{\Gamma\left[(\nu+p)/2\right]}{|\Sigma|^{1/2}\left(\nu\pi\right)^{p/2}\Gamma\left(\nu/2\right)}\left[1+\frac{1}{\nu}\left(\mathbf{z}-\mu\right)'\Sigma^{-1}\left(\mathbf{z}-\mu\right)\right]^{-(\nu+p)/2}$$

where scalar parameter ν is the degrees of freedom of the distribution while the $p \times 1$ vector μ is the location parameter and the positive definite $p \times p$ matrix Σ is the scale matrix. If $\nu > 1$ then $E(\mathbf{z}) = \mu$. If $\nu > 2$ then $Var(\mathbf{z}) = \nu\Sigma/(\nu-2)$. For our problem, it makes sense to work in terms of the *inverse* of the scale matrix, Σ^{-1}. Parameterized in this way, the log of the multivariate Student-t density is given by

$$\begin{aligned}\log t_p\left(\mathbf{z}|\nu,\mu,\Sigma^{-1}\right) &= \log\Gamma\left[\left(\nu+p\right)/2\right] - \log\Gamma\left(\nu/2\right) - \frac{p}{2}\log\left(\nu\pi\right) \\ &\quad + \frac{1}{2}\log|\Sigma^{-1}| - \frac{1}{2}(\nu+p)\log\left[1+\frac{1}{\nu}\left(\mathbf{z}-\mu\right)'\Sigma^{-1}\left(\mathbf{z}-\mu\right)\right]\end{aligned}$$

where we have used the fact that $|\Sigma|^{-1} = |\Sigma|^{-1}$ to write a positive $\frac{1}{2}\log|\Sigma^{-1}|$ term in place of a negative $\frac{1}{2}\log|\Sigma|$ term. Now, let R be the Cholesky factor of Σ^{-1} so that $\Sigma^{-1} = R'R$ and define $\tilde{\mathbf{z}} = \mathbf{z} - \mu$ and $\mathbf{v} = R\tilde{\mathbf{z}}$. Using these definitions,

$$\left(\mathbf{z}-\mu\right)'\Sigma^{-1}\left(\mathbf{z}-\mu\right) = (R\tilde{\mathbf{z}})'(R\tilde{\mathbf{z}}) = \mathbf{v}'\mathbf{v}$$

and, letting R_{ii} denote the ith diagonal element of R,

$$\begin{aligned}\frac{1}{2}\log|\Sigma^{-1}| &= \frac{1}{2}\log|R'R| = \frac{1}{2}\log\left(|R'|\cdot|R|\right) = \frac{1}{2}\left(\log|R'| + \log|R|\right) \\ &= \frac{1}{2}\left(2\log|R|\right) = \sum_{i=1}^{p}\log R_{ii}\end{aligned}$$

since $|A| = |A'|$, $|AB| = |A| \cdot |B|$ and the determinant of a triangular matrix equals the product of its diagonal elements. Thus, we have

$$\log t_p\left(\mathbf{z}|\nu,\mu,\Sigma^{-1}\right) = \log\Gamma\left[(\nu+p)/2\right] - \log\Gamma\left(\nu/2\right) - \frac{p}{2}\log\left(\nu\pi\right) + \text{trace}\left[\log\left(\text{diag}\left\{R\right\}\right)\right] - \frac{1}{2}(\nu+p)\log\left[1+\frac{1}{\nu}\mathbf{v}'\mathbf{v}\right]$$

Note that, since $\mathbf{v}'\mathbf{v}/\nu$ could be very close to zero, it is best to implement `log1p`$(\mathbf{v}'\mathbf{v}/\nu)$. An explanation of `log1p` is given here.

Note that the only term in the log density expression that depends on $\mathbf{z}$ is $\mathbf{v}'\mathbf{v}$. We can calculate this term simultaneously for *all columns* of Z as follows. First let $\widetilde{Z}$ denote the result subtracting of subtracting the vector μ from each column of Z, i.e. $\widetilde{Z} = Z - \mu\mathbf{1}_n'$. To calculate $\mathbf{v}'\mathbf{v}$ for each column of Z we simply square the elements of $R\widetilde{Z}$ and take the column sums of the resulting matrix.

8.3 Efficient Calculation of $R_{T,\lambda}$

In the normal model from above we described the efficient calculation of

$$R_T = \left[R_0^{-1} + \sum_{t=1}^{T}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right]^{-1}$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - X_t\boldsymbol{\gamma}$. In the Student-t model we instead need to calculate

$$R_{T,\lambda} = \left[R_0^{-1} + \sum_{t=1}^{T}\lambda_t\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right]^{-1}$$

which differs only in that each term of the sum is multiplied by the scalar λ_t. Recall that we defined

$$\widetilde{Y} = \begin{bmatrix}\mathbf{y}_1'\\ \vdots \\ \mathbf{y}_T'\end{bmatrix}, \quad \widetilde{X} = \begin{bmatrix}\mathbf{x}_1'\\ \vdots \\ \mathbf{x}_T'\end{bmatrix}, \quad \Gamma = \begin{bmatrix}\boldsymbol{\gamma}_1 & \cdots & \boldsymbol{\gamma}_D\end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix}\boldsymbol{\varepsilon}_1'\\ \vdots \\ \boldsymbol{\varepsilon}_T'\end{bmatrix}$$

so that $\boldsymbol{\varepsilon} = \widetilde{Y} - \widetilde{X}\Gamma$ allowing us to express the sum in the expression for R_T more compactly in matrix notation as $\left(\widetilde{Y} - \widetilde{X}\Gamma\right)' \left(\widetilde{Y} - \widetilde{X}\Gamma\right)$. To calculate the sum in the expression for $R_{T,\lambda}$, however, we need to take into account λ_t. To this end, define $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$ and $\Lambda = \text{diag}(\boldsymbol{\lambda})$. Then we see that

$$R_{T,\lambda} = \left[R_0^{-1} + \left(\widetilde{Y} - \widetilde{X}\Gamma\right)' \Lambda \left(\widetilde{Y} - \widetilde{X}\Gamma\right)\right]^{-1}$$

Everything else is the same as above, for example can pre-compute R_0^{-1} and there is no gain from the Sherman-Morris-Woodbury formula unless $D > T$.

8.4 Efficient Calculation of $G_{T,\lambda}$

In the case of the normal model, described above, we used clever matrix algebra to pre-compute quantities involved in the expression[2]

$$G_T^{-1} = G_0^{-1} + \sum_{t=1}^{T} X_t' \Omega^{-1} X_t$$

We won't end up with such nice results as in the normal case here, but we can still express the calculations in a more convenient form by exploiting some of the same properties of the Kronecker product. Recall that

$$X_t' \Omega^{-1} X_t = \Omega^{-1} \otimes (\mathbf{x}_t' \otimes \mathbf{x}_t') = \Omega^{-1} \otimes (\mathbf{x}_t \mathbf{x}_t')$$

as we showed above in the discussion for the normal model. One further property of the Kronecker product that we did not use in our earlier calculations is the following:

$$(kA) \otimes B = A \otimes (kB) = k(A \otimes B)$$

if k is a scalar and A, B are matrices. Thus, since λ_t is a scalar,

$$\lambda_t X_t' \Omega^{-1} X_t = \lambda_t \left[\Omega^{-1} \otimes (\mathbf{x}_t \mathbf{x}_t')\right] = \Omega^{-1} \otimes (\lambda_t \mathbf{x}_t \mathbf{x}_t')$$

[2]Since the MV normal sample we have implement is in terms of the precision matrix rather than covariance matrix we work with G_T^{-1} rather than its inverse G_T.

and thus, since the Kronecker product is bilinear,

$$\sum_{t=1}^{T} \lambda_t X_t' \Omega^{-1} X_t = \Omega^{-1} \otimes \left(\sum_{t=1}^{T} \lambda_t \mathbf{x}_t \mathbf{x}_t' \right) = \Omega^{-1} \otimes \left(\widetilde{X}' \Lambda \widetilde{X} \right)$$

where $\widetilde{X}$ and Λ are defined in the preceding subsection on calculating $R_{T,\lambda}$. Thus, we have

$$G_{T,\lambda}^{-1} = G_0^{-1} + \Omega^{-1} \otimes \left(\widetilde{X}' \Lambda \widetilde{X} \right)$$

Unlike the case of the normal model, we cannot pre-compute the second term in the Kronecker product since it depends on Λ, which is updated in each iteration of the sampler. We could, however, pre-compute the matrices $\mathbf{x}_t \mathbf{x}_t'$ and then take the λ-weighted sum in each step of the sampler. It is unclear whether this efficiency gain is worth the trouble, but we could try it if things are running too slowly.

8.5 Efficient Calculation of $\bar{\boldsymbol{\gamma}}_\lambda$

Recall from our calculations for the normal model given above that, using the properties of the Kronecker product and the vec operator,

$$X_t' \Omega^{-1} \mathbf{y}_t = \text{vec}(\mathbf{x}_t \mathbf{y}_t' \Omega^{-1})$$

Replacing Ω^{-1} with $\lambda_t \Omega^{-1}$ and using the fact that λ_t is a scalar, and hence commutes, along with the fact that $\text{vec}(kA) = k\text{vec}(A)$ for any scalar k and matrix A, it follows that

$$\lambda_t X_t' \Omega^{-1} \mathbf{y}_t = \text{vec}(\lambda_t \mathbf{x}_t \mathbf{y}_t' \Omega^{-1}) = \lambda_t \text{vec}(\mathbf{x}_t \mathbf{y}_t' \Omega^{-1})$$

From here it follows that

$$\sum_{t=1}^{T} \lambda_t X_t' \Omega^{-1} \mathbf{y}_t = \sum_{t=1}^{T} \text{vec}(\lambda_t \mathbf{x}_t \mathbf{y}_t' \Omega^{-1}) = \text{vec} \left[\left(\sum_{t=1}^{T} \lambda_t \mathbf{x}_t \mathbf{y}_t' \right) \Omega^{-1} \right] = \text{vec} \left(\widetilde{X}' \Lambda \widetilde{Y} \Omega^{-1} \right)$$

where $\widetilde{X}, \widetilde{Y}$ and Λ are as defined in the subsection on efficient calculation of $R_{T,\lambda}$ from above. The rest of the calculation is essentially the same as that for $\bar{\boldsymbol{\gamma}}$ in the case of the normal model, described above.

8.6 Drawing λ_t

This step in the sampler has no counterpart in the normal model:

$$\lambda_t|\boldsymbol{\gamma}, \nu, Y_T \sim G\left(\frac{\nu + D}{2}, \frac{\nu + \boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_t}{2}\right)$$

where $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - X_t\boldsymbol{\gamma}$. Notice that we have one auxiliary parameter, λ_t, per observation. Each of these comes from a *different* Gamma distribution: while the shape parameter is common, the rate parameter differs according to the residual $\boldsymbol{\varepsilon}$ associated with each observation. Because we have a C++ function for drawing gamma variates that is vectorized with respect to its second argument, the rate parameter, we need an expression for the while *vector* of rate parameters. I don't think there's any straightforward way to write this out using matrix operations, so we'll just have to use a loop to construct the desired vector.

9 Marginal Likelihood for Student-t Model

Because the Gibbs sampler for the Student-t model has three blocks rather than two, we need to use a more complicated version of the Chib (1995) method to calculate the marginal likelihood for this model.

General Three-Block Algorithm First we describe the method for an arbitrary three-block Gibbs sampler. By re-arranging Bayes' Rule we have the identity

$$f(y) = \frac{f(y|\theta_1^*, \theta_2^*, \theta_3^*)\pi\left(\theta_1^*, \theta_2^*, \theta_3^*\right)}{\pi\left(\theta_1^*, \theta_2^*, \theta_3^*|y\right)}$$

for any specified values $(\theta_1^*, \theta_2^*, \theta_3^*)$ of the parameters. In particular this holds at the *posterior mean* which is where we will evaluate the expression. Hence, the *log* marginal likelihood is given by

$$\log f(y) = \log \pi\left(\theta_1^*, \theta_2^*, \theta_3^*\right) + \log f(y|\theta_1^*, \theta_2^*, \theta_3^*) - \pi\left(\theta_1^*, \theta_2^*, \theta_3^*|y\right)$$

The first two terms are easy: we simply evaluate the log of the prior and likelihood at the posterior mean for the three parameters. The third one, however, is more complicated. To calculate it we use the factorization:

$$\pi\left(\theta_1^*, \theta_2^*, \theta_3^* | y\right) = \pi(\theta_1^* | y)\pi\left(\theta_2^* | \theta_1^*, y\right)\pi\left(\theta_3^* | \theta_2^*, \theta_1^*, y\right)$$

This leaves us with three a product of new terms that we need to calculate. The last of the terms in the product, $\pi(\theta_3^* | \theta_2^*, \theta_1^*, y)$ is immediately available: this conditional density is known since we used it as a step in the Gibbs sampler. All we need to do is substitute in the appropriate values for θ_2^*, θ_1^* and evaluate the density at θ_3^*. To calculate the first term in the product we need to marginalize over θ_2, θ_3. A Monte-Carlo approximation to the appropriate integral can be computed directly from the Gibbs sampler output:

$$\widehat{\pi}\left(\theta_1^* | y\right) = \frac{1}{G}\sum_{g=1}^{G} \pi\left(\theta_1^* | \theta_2^{(g)}, \theta_3^{(g)}, y\right)$$

To do this we rely on the fact that $\pi(\theta_1 | \theta_2, \theta_3, y)$ is a known density – we use it in the Gibbs sampler. The middle term in the product is the most difficult one to calculate. To begin, notice that

$$\pi\left(\theta_2^* | \theta_1^*, y\right) = \int \pi(\theta_2^*, \theta_3 | \theta_1^*, y)\, d\theta_3 = \int \pi(\theta_2^* | \theta_1^*, \theta_3^*, y)\pi(\theta_3 | \theta_1^*, y)\, d\theta_3$$

The idea is to construct a Monte-Carlo approximation of the integral on the right-hand-side of the preceding expression. The approximation we use is

$$\hat{\pi}\left(\theta_2^* | \theta_1^*, y\right) = \frac{1}{G}\sum_{g=1}^{G} \pi(\theta_2^* | \theta_1^*, \theta_3^{(g)}, y)$$

but the draws $\left\{\theta_3^{(g)}\right\}$ come *not* from the original run of the Gibbs sampler but from a so-called "reduced run" in which we sample $\theta_2^{(g)}$ and $\theta_3^{(g)}$ from $\pi(\theta_2 | \theta_1^*, \theta_3, y)$ and $\pi(\theta_3 | \theta_1^*, \theta_2, y)$. In other words, the reduced run holds θ_1 *fixed* at θ_1^*, the posterior mean calculated from the draws of the *usual* Gibbs sampler. We can carry out the reduced run using the exact same algorithm as we use for the full Gibbs sampler: we just need to keep θ_1^* fixed and make sure that we store the draws $\theta_3^{(g)}$ that we will need to calculate $\hat{\pi}(\theta_2^* | \theta_1^*, y)$.

Specializing to the Student-t Model

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

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

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

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

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

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