This version: April 24, 2023

1 The generative model

1.1 Setup and Notation

- Consider a sequence of length K factor returns f_t where $t \in 1:T$. The factor returns can either include a constant for the intercept, or the data may already be centered. Factor returns are measured on a monthly basis.

- An unobserved vector of latent returns for asset $i \in 1:N$ is a function of these factors such that

$$x_{it} = f_t' \beta_i + \varepsilon_{it}^u$$

  - Here, β_i is a $K \times 1$ vector of exposures.
  - For notational parsimony, and because other assets only enter the model through priors, drop the i subscript.

- The econometrician observes a length S vector of returns y. Note that S will not generally equal T. Notably, the observation frequency may differ from the frequency of return realization.

  - For convenience, use the notation $t[s]$ to denote the corresponding value of t for a particular s. Because the interval of time between observations is constant (Δt periods), the mapping of observation period to latent return period t is described as:

$$t[s] \equiv P + s * \Delta t$$

  - The observed returns y are a function of vector of unobserved ("latent") returns x.
    - Intuition: The lags reflect a process whereby valuations of investments take up to the end of the lag window to be fully reflected in the returns of an asset.
  - The returns of y follow a moving average process with $P + \Delta t$ terms plus measurement error. The moving average window is parameterized by P unrestricted coefficients and Δt coefficients determined by restrictions.
  - Let ϕ represent the length P vector of unrestricted coefficients and $\tilde{\phi}$ be the length $P + \Delta t$ vector containing both restricted and unrestricted coefficients. Note that both ϕ and $\tilde{\phi}$ use indexing that is the reverse of typical, with $\tilde{\phi}_{P+\Delta t}$ corresponding to the coefficient on the contemporaneous value of x. Then:

$$\begin{aligned} y_s &= \left(\tilde{\phi}_{1:(P+\Delta t)}\right)' x_{(t[s]-P-\Delta t+1):t[s]} + \varepsilon_t^y \\ &= \left(\tilde{\phi}_{(P+1):(P+\Delta t)}\right)' x_{(t[s]-\Delta t+1):t[s]} + \phi' x_{(t[s]-P):(t[s]-\Delta t)} + \varepsilon_t^y \end{aligned}$$

    - If the contemporaneous term $x_{t[s]}$ were unrestricted (say given a coefficient ϕ_{P+1}), the coefficients ϕ and β would only be identified via priors.

- To see the lack of identification, note that doubling β and halving ϕ would lead to the same prediction. Setting the coefficient on $x_{t[s]}$ to 1 creates an implicit scaling restriction, but creates difficulties with respect to the scaling of β. Restricting the coefficient to $\Delta t - \phi' 1$ preserves scaling at the cost of slightly increased in complexity.
- When $\Delta t > 1$, further restrictions help the identification. Consider $\Delta t = 3$, which corresponds to quarterly observed returns and monthly factor returns. If $j \in 1 : \Delta t$ and the sum $\tilde{\phi}_j + \tilde{\phi}_{j+\Delta t} + \cdots + \tilde{\phi}_{j+P-\Delta t}$ does not add to the same value for all j, months that fall earlier in the quarter will have a different long-run impact on NAV than months later in the quarter. This combined with the above restriction implies Δt restrictions.

* Subject to regularity conditions, the use of measurement error in the model is without loss of generality with respect to y following a moving average process. See 3.5 for a discussion.

- The moving average can be written in two forms with matrix notation. To see this, consider the special case where both y and x have the same frequency. Then:

$$\hat{y} = \Phi x = X_L R\phi + x_S$$

$$\hat{y} = \begin{bmatrix} \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} & \cdots & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \phi_1 & \cdots & \phi_P & \tilde{\phi}_{P+1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{T-1} \\ x_T \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_{P-1} & x_P & x_{P+1} \\ x_2 & x_3 & \cdots & x_P & x_{P+1} & x_{P+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{T-P-1} & x_{T-P} & \cdots & x_{T-3} & x_{T-2} & x_{T-1} \\ x_{T-P} & x_{T-P+1} & \cdots & x_{T-2} & x_{T-1} & x_T \end{bmatrix} [R] \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{P-1} \\ \phi_P \end{bmatrix} + \begin{bmatrix} x_{P+1} \\ x_{P+2} \\ \vdots \\ x_{T-1} \\ x_T \end{bmatrix}$$

$$R \equiv \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ -1 & -1 & \cdots & -1 & -1 \end{bmatrix}$$

$$\tilde{\phi}_{P+1} \equiv 1 - 1'\phi$$

$$x_S \equiv x_{t[s]\forall s}$$

  - To generalize to quarterly data and other frequencies, recall $t[s] \equiv P + s * \Delta t$ and define

$$\Phi_{sj} \equiv \begin{cases} \phi_{P-(t[s]-\Delta t-j)} & 1 \leq P - (t[s] - \Delta t - j) \leq P \\ 1 - \left(\iota^{\phi}_{\Delta t-(t[s]-j)}\right)' \phi & t[s] - \Delta t < j \leq t[s] \\ 0 & otherwise \end{cases} \tag{1}$$

$$X_{Lsj} \equiv x_{t[s]-(P+\Delta t-j)} \tag{2}$$

$$\iota^{\phi}_{pl} \equiv \begin{cases} 1 & p + l \mod \Delta t = 0 \\ 0 & otherwise \end{cases} \tag{3}$$

  - The above matrix formulation can be generalized by only including rows of X_L where $t \in \{t[1...S]\}$

and adjusting the restriction matrix. The general version is then:

$$\begin{aligned}\hat{y} = \Phi x =& X_L R\phi + x_S \\ s.t.& \\ x_S \equiv& X_L \iota_{\Delta t}\end{aligned}$$

where $\iota_{\Delta t}$ is a vector of P zeros followed by Δt ones, making x_S the sum of the last Δt columns of X_L.

- The generative process is given by:

$$\begin{aligned}
p\left(y|rest\right) \sim& MN\left(X_L R\phi + x_S, \frac{1}{\tau_y} I\right) (\text{equivelently} MN\left(\Phi x, \frac{1}{\tau_y} I\right)) \\
p\left(x|rest\right) \sim& MN\left(F\beta + r, \frac{1}{\tau_x \tau_y} \Psi^{-1}\right) \\
p\left(\phi|rest\right) \sim& MN\left(\phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right) \\
p\left(\beta|rest\right) \sim& MN\left(\beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta}\left[DA_0 D\right]^{-1}\right) \\
d_k \equiv& (\gamma_k + (1-\gamma_k)\frac{1}{v^2})^{0.5} \\
p\left(\gamma_k\right) \sim& Bern\left(\omega\right) \\
p(\omega) \sim& Beta\left(\kappa_0, \delta_0\right) \\
p\left(\tau_y\right) \sim& Gamma\left(\alpha_{y0}, \zeta_{y0}\right) \\
p\left(\tau_x\right) \sim& Gamma\left(\alpha_{x0}, \zeta_{x0}\right) \\
p\left(\psi_t\right) \sim& Gamma\left(\nu/2, \nu/2\right) \\
p\left(\nu\right) \sim& Gamma\left(\alpha_{\nu 0}, \zeta_{\nu 0}\right) \\
p\left(\tau_\phi\right) \sim& Gamma\left(\alpha_{\phi 0}, \zeta_{\phi 0}\right) \\
p\left(\tau_\beta\right) \sim& Gamma\left(\alpha_{\beta 0}, \zeta_{\beta 0}\right)
\end{aligned}$$

  - Definitions for the above (some of this was already previously described):
    * In the above distributions, MN is the multi-variate normal distribution, $Bern$ is the Bernoulli distribution, $Beta$ is the beta distribution, and $Gamma$ is the gamma distribution with an inverse scale parameterization.
    * All priors are conditionally conjugate except for ν. Wand et al 2011 has a strategy for computing $E\left[q\left(\nu\right)\right]$ in section 4.1, though it requires numerical integration. An alternative is backing out a plausible value for ν from the data and treating ν as known (say by setting ν to match the t-distribution's kurtosis with that of the data.)
    * The rest of the variables are summarized in the following table:

Table 1: Variable Definitions

Variable definitions for the Data Generating Process (DGP). **DGP corresponds to a system with S observations over T periods with each observation dependent on P terms in a moving average and K factors.** Variables are divided into the following types: observed values, local parameters along a particular dimension, global scalar parameters, and hyperparameters.

Variable	Type	Dimensions	Definition/Description
y	Observed	$S \times 1$	Vector of observed returns
$x; X_L$	Local	$T \times 1$; $S \times (P+1)$	x is a vector of gross latent returns; X_L is defined in Equation 2
r	Observed	$T \times 1$	Vector of risk-free returns
ϕ; Φ	Local	$P \times 1$; $S \times T$	ϕ is the moving average window; Φ is defined in Equation 1.
τ_y	Global	-	Precision parameter for independent measurement error
ϕ_0	Hyper	$P \times 1$	Prior estimate for ϕ
M_0	Hyper	$P \times P$	Precision of prior estimate ϕ_0
F	Observed	$T \times K$	Matrix of factor returns, possibly including an intercept
β	Local	$K \times 1$	Regression coefficients of x on F
ψ; Ψ	Local	$T \times 1$; $T \times T$	ψ is a vector of precision weights for x;$\Psi = Diag\left(\psi\right)$
τ_x	Global	-	Precision multiplier parameter for the regression of x on F
τ_β	Global	-	Prior precision multiplier parameter for the prior on ϕ
τ_ϕ	Global	-	Prior precision multiplier parameter for the prior on β
β_0	Hyper	$K \times 1$	Prior mean of β conditional on exclusion ($\gamma = 0$)
β_0^Δ	Hyper	$K \times 1$	Shift in prior mean of β conditional on inclusion ($\gamma = 1$)
A_0	Hyper	$K \times K$	Possibly diagonal prior precision for $\beta_0 + D^{-1}\beta_0^D$
d; D	Local	$K \times 1$; $K \times K$	d is a function of γ and adjusts β for sparsity; $D = Diag\left(d\right)$
γ	Local	$K \times 1$	Vector of variable selection indicators
v	Hyper	-	Variance of the spike distribution as a fraction of the slab variance
ω	Global	-	Probability of variable selection
κ_0; δ_0	Hyper	-	Hyperparameters for prior on ω
ν	Global	-	Non-normality parameter for x; DOF of posterior t distribution
$\alpha_{\phi 0}$; $\zeta_{\phi 0}$	Hyper	-	Hyperparameters for τ_ϕ
$\alpha_{\beta 0}$; $\zeta_{\beta 0}$	Hyper	-	Hyperparameters for τ_β
α_{x0}; ζ_{x0}	Hyper	-	Hyperparameters for τ_x; α_{x0} is shape, ζ_{x0} is inverse scale
α_{y0}; ζ_{y0}	Hyper	-	Hyperparameters for τ_y; α_{y0} is shape, ζ_{y0} is inverse scale
ν_0^-; ν_0^+	Hyper	-	Hyperparameters for prior on ν

- The posterior distribution:

$$
\begin{aligned}
p\left(\Theta|y,F\right) \propto & p\left(y|x,\gamma,\omega,\beta,\phi,\tau_x,\tau_y,\tau_\phi,\tau_\beta,\psi,\nu,F\right) \times p\left(x|\beta,\phi,\tau_x,\tau_y,\psi,F\right) \\
& \times p\left(\phi|\tau_y,\tau_\phi\right) \times p\left(\beta|\gamma,\tau_x,\tau_y,\tau_\beta\right) \times p\left(\gamma|\omega\right) \times p\left(\omega\right) \\
& \times p\left(\psi|\nu\right) \times p\left(\nu\right) \times p\left(\tau_x\right) \times p\left(\tau_y\right) \times p\left(\tau_\phi\right) \times p\left(\tau_\beta\right) \\
= & MN\left(y;\Phi x,\frac{1}{\tau_y}I\right) \times MN\left(x;F\beta + r,\frac{1}{\tau_x}\Psi^{-1}\right) \\
& \times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \times MN\left(\beta;\beta_0 + D^{-1}\beta_0^\Delta,\frac{1}{\tau_x\tau_\beta}\left[DA_0D\right]^{-1}\right) \\
& \times \prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right) \times Beta\left(\omega;\kappa_0,\delta_0\right) \\
& \times \prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right) \times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right) \\
& \times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right) \times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right) \\
& \times Gamma\left(\tau_\phi;\alpha_{\phi 0};\zeta_{\phi 0}\right) \times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)
\end{aligned}
$$

- Approximate the posterior using the Variational Bayes mean-field approach (see Section 2.2).
  - This entails minimizing the KL divergence via

$$\log q_j\left(\Theta_j|D,\Theta_{-j}\right) = E_{\Theta_{-j}} \log\left[p\left(y,\Theta\right)\right]$$

  - The approximate posterior is:

$$\begin{aligned} p\left(\Theta|y\right) \underset{\sim}{\propto} & q\left(\phi\right) \times q\left(\tau_y\right) \times q\left(x\right) \\ & \times q\left(\beta\right) \times q\left(\tau_x\right) \times \prod_{t\in 1:T} q\left(\psi_i\right) \\ & \times q\left(\nu\right) \times \prod_{k\in 1:K} q\left(\gamma_k\right) \times q\left(\omega\right) \\ & \times q\left(\tau_\phi\right) \times q\left(\tau_\beta\right) \end{aligned}$$

1.2 Derivation of MCMC posterior distributions and associated moments

1.2.1 Derivation of $p\left(\phi|rest\right)$

- First, define

$$\begin{aligned} \tilde{y} &\equiv y - x_S \\ \tilde{X}_L &\equiv X_L R \end{aligned}$$

- Then the conditional posterior is given by:

$$\begin{aligned}
\log p\left(\phi|rest\right) =& -\frac{\tau_y}{2}\left(y - X_L R\phi - x_S\right)'\left(y - X_L R\phi - x_S\right) - \frac{\tau_y\tau_\phi}{2}\left(\phi - \phi_0\right)' M_0\left(\phi - \phi_0\right) + c_1^\phi \\
=& -\frac{\tau_y}{2}\left(\tilde{y} - \tilde{X}_L\phi\right)'\left(\tilde{y} - \tilde{X}_L\phi\right) - \frac{\tau_y\tau_\phi}{2}\left(\phi - \phi_0\right)' M_0\left(\phi - \phi_0\right) + c_1^\phi \\
=& -\frac{\tau_y}{2}\left[\phi'\tilde{X}_L'\tilde{X}_L\phi - \tilde{y}'\tilde{X}_L\phi - \phi'\tilde{X}_L'\tilde{y} - \tau_\phi\phi' M_0\phi - \tau_\phi\phi_0' M_0\phi - \tau_\phi\phi' M_0\phi_0\right] + c_2^\phi \\
=& -\frac{\tau_y}{2}\left[\phi'\left(\tilde{X}_L'\tilde{X}_L + \tau_\phi M_0\right)\phi - \left(\tilde{y}'\tilde{X}_L + \tau_\phi\phi_0' M_0\right)\phi - \phi'\left(\tilde{X}_L'\tilde{y} + \tau_\phi M_0'\phi_0\right)\right] + c_2^\phi \\
=& -\frac{1}{2}\left(\phi - \mu_\phi\right)\Lambda_\phi\left(\phi - \mu_\phi\right) + c_3^\phi \\
=& \log MN\left(\phi; \mu_\phi, \Lambda_\phi^{-1}\right) + c_4^\phi
\end{aligned}$$

$s.t.$

$$\begin{aligned}
\Lambda_\phi \equiv& \tau_y\left(\tilde{X}_L'\tilde{X}_L + \tau_\phi M_0\right) \\
\mu_\phi \equiv& \tau_y\Lambda_\phi^{-1}\left(\tilde{X}_L'\tilde{y} + \tau_\phi M_0'\phi_0\right) \\
c_1^\phi \equiv& \frac{S+P}{2}\log\left(\frac{\tau_y}{2\pi}\right) + \frac{P}{2}\log\tau_\phi + \frac{1}{2}\log det\left(M_0\right) \\
&+ \log\Bigg[MN\left(xT^{1/2}; F\beta + T^{1/2}r, \frac{T}{\tau_x\tau_y}\Psi^{-1}\right) \times MN\left(\beta; \beta_0 + D^{-1}\beta_0^\Delta, \frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right) \\
&\times \prod_{k=1}^{K} Bern\left(\gamma_k; \omega\right) \times Beta\left(\omega; \kappa_0, \delta_0\right) \times \prod_{t=1}^{T} Gamma\left(\psi; \nu/2, \nu/2\right) \\
&\times Gamma\left(\nu; \alpha_{\nu 0}, \zeta_{\nu 0}\right) \times Gamma\left(\tau_x; \alpha_{x0}, \zeta_{x0}\right) \times Gamma\left(\tau_y; \alpha_{y0}, \zeta_{y0}\right) \\
&\times Gamma\left(\tau_\beta; \alpha_{\beta 0}, \zeta_{\beta 0}\right) \times Gamma\left(\tau_\phi; \alpha_{\phi 0}, \zeta_{\phi 0}\right)\Bigg] + c^{ev} \\
c_2^\phi \equiv& c_1^\phi - \frac{\tau_y}{2}\left[\tau_\phi\phi_0' M_0\phi + \tilde{y}'\tilde{y}\right] \\
c_3^\phi \equiv& c_2^\phi + \frac{1}{2}\mu_\phi'\Lambda_\phi\mu_\phi \\
c_4^\phi \equiv& c_3^\phi + \frac{P}{2}\log 2\pi - \frac{1}{2}\log det\Lambda_\phi \\
c_{ev} \equiv& -\log p\left(y_t\right)
\end{aligned}$$

  - The last constant makes the full posterior into a valid probability distribution.
  - Note that it may make sense to impose additional restrictions not easily captured by R. For instance, the prior distribution can be truncated to force $\phi \in [0,1]$. With this restriction, the above analysis applies over the interval, with the support of the posterior and prior distribution truncated and re-normalized as appropriate.

- Then the approximate posterior of ϕ is:

$$
\begin{aligned}
\log q(\phi) =& E_{-\phi}\left[-\frac{\tau_y}{2}\left(\tilde{y}-\tilde{X}_L\phi\right)'\left(\tilde{y}-\tilde{X}_L\phi\right)-\frac{\tau_y\tau_\phi}{2}(\phi-\phi_0)'M_0(\phi-\phi_0)\right]+\overline{c}_1^\phi \\
=& -\frac{E[\tau_y]}{2}E_{-\phi}\left[\phi'\tilde{X}_L'\tilde{X}_L\phi-\tilde{y}'\tilde{X}_L\phi-\phi'\tilde{X}_L'\tilde{y}+\tau_\phi\left(\phi'M_0\phi-\phi_0'M_0\phi-\phi'M_0\phi_0\right)\right]+\overline{c}_2^\phi \\
=& -\frac{E[\tau_y]}{2}E_{-\phi}\left[\phi'\left(\tilde{X}_L'\tilde{X}_L+\tau_\phi M\right)\phi-\left(\tilde{y}'\tilde{X}_L+\tau_\phi\phi_0'M_0\right)\phi-\phi'\left(\tilde{X}_L'\tilde{y}+\tau_\phi M_0\phi_0\right)\right]+\overline{c}_2^\phi \\
=& -\frac{E[\tau_y]}{2}\Bigg(\phi'\left(E\left[\tilde{X}_L'\tilde{X}_L\right]+E[\tau_\phi]M_0\right)\phi-\left(E\left[\tilde{y}'\tilde{X}_L\right]+E[\tau_\phi]\phi_0'M_0\right)\phi \\
&-\phi'\left(E\left[\tilde{X}_L'\tilde{y}\right]+E[\tau_\phi]E[M_0]\phi_0\right)\Bigg)+\overline{c}_2^\phi \\
=& -\frac{1}{2}\left((\phi-\overline{\mu}_\phi)'\overline{\Lambda}_\phi(\phi-\overline{\mu}_\phi)\right)+\overline{c}_3^\phi \\
=& \log MN\left(\phi;\overline{\mu}_\phi,\overline{\Lambda}_\phi^{-1}\right)+\overline{c}_4^\phi \\
& s.t. \\
\overline{\Lambda}_\phi \equiv& E[\tau_y]\left(E\left[\tilde{X}_L'\tilde{X}_L\right]+E[\tau_\phi]E[M_0]\right) \\
\overline{\mu}_\phi \equiv& E[\tau_y]\Lambda_\phi^{-1}\left(E\left[\tilde{X}_L'\tilde{y}\right]+E[\tau_\phi]E[M_0]\phi_0\right) \\
\overline{c}_1^\phi \equiv& \frac{S+P}{2}E\left[\log\left(\frac{\tau_y}{2\pi}\right)\right]+0.5\log det(M_0)+\frac{P}{2}\log\tau_\phi \\
&+E_{-\phi}\log\Bigg[MN\left(x;F\beta,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^\Delta,\frac{1}{\tau_x\tau_y\tau_\beta}[DA_0D]^{-1}\right) \\
&\times\prod_{k=1}^{K}Bern(\gamma_k;\omega)\times Beta(\omega;\kappa_0,\delta_0)\times\prod_{t=1}^{T}Gamma(\psi;\nu/2,\nu/2) \\
&\times Gamma(\nu;\alpha_{\nu0},\zeta_{\nu0})\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
&\times Gamma(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0})\times Gamma(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0})\Bigg]+c^{ev} \\
\overline{c}_2^\phi \equiv& \overline{c}_1^\phi-\frac{E[\tau_y]}{2}\left(E[\tilde{y}'\tilde{y}]+E[\tau_\phi]\phi_0'E[M_0]\phi_0\right) \\
\overline{c}_3^\phi \equiv& \overline{c}_2^\phi+\frac{1}{2}\left(\overline{\mu}_\phi'\overline{\Lambda}_\phi\overline{\mu}_\phi\right) \\
\overline{c}_4^\phi \equiv& \overline{c}_3^\phi+\frac{P}{2}\log 2\pi-\frac{1}{2}\log\left(det\left(\overline{\Lambda}_\phi\right)\right)
\end{aligned}
$$

- To derive the expectation of $E[X_L'X_L]$, need Matrix Cookbook equation 326 which states if x is an $N\times 1$ random vector and A and B matrices of width N, then $E\left[(Ax)'Bx\right]=Tr(A\Sigma B')+\mu'A'B\mu$
  - Define X_L^p as

$$X_L=\begin{bmatrix}X_L^1 & X_L^2 & \dots & X_L^{P+\Delta t}\end{bmatrix}$$

  - Assume (this will be shown later) that $x\sim MV\left(\mu_x,\Lambda_x^{-1}\right)$. The plan will be to create a weighting matrix $E[A^p x]$ such that $E\left[\left(X_L^i\right)'X_L^j\right]=E\left[x'\left(A^i\right)'A^j x\right]$, which will allow us to use the Matrix Cookbook formula.

- The weighting matrix for column p takes the form of a sparse $S \times T$ selector matrix $\iota\left(X_L^p\right)$. The matrix is zero everywhere except for $\iota\left(X_L^p\right)_{s,t[s]-(P+\Delta t)+p}$ for all $s \in 1:S$.
  * Intuitively, $\iota\left(X_L^p\right)$ transforms vector x into X_L^p.
  * Each row of $\iota\left(X_L^p\right)$ contains a single non-zero. Similarly column j of $\iota\left(X_L^p\right)$ contains a single 1 if and only if $x_j \in X_L^p$.
  * Another way of writing this:

$$\iota\left(X_L^p\right)_{sj} \equiv \begin{cases} 1 & j = t\left[s\right] - \left(P + \Delta t\right) + p \\ 0 & otherwise \end{cases}$$

- From these definitions,

$$\begin{aligned} E\left[X_L' X_L\right]_{ij} \equiv & E\left[\left(\iota\left(X_L^i\right) x\right)' \left(\iota\left(X_L^j\right) x\right)\right] \\ = & Tr\left(\iota\left(X_L^i\right) \Lambda_x^{-1} \iota\left(X_L^j\right)'\right) + \mu_x' \iota\left(X_L^i\right)' \iota\left(X_L^j\right) \mu_x \end{aligned}$$

  which fully characterizes $E\left[X_L' X_L\right]$.
- Note that $E\left[\tilde{X}_L' \tilde{X}_L\right] = R' E\left[X_L' X_L\right] R$

- The same approach applies to $E\left[X_L' \tilde{y}\right]$:

$$\begin{aligned} E\left[X_L' \tilde{y}\right] = & E\left[X_L\right]' y - E\left[X_L' x_S\right] \\ & s.t. \\ E\left[X_L' x_S\right]_i = & \sum_{l=1}^{l=\Delta t} E\left[\left(\iota\left(X_L^i\right) x\right)' \left(\iota\left(X_L^{P+l}\right) x\right)\right] \\ = & \sum_{l=1}^{l=\Delta t} \left[Tr\left(\iota\left(X_L^i\right) \Lambda_x^{-1} \iota\left(X_L^{P+l}\right)'\right) + \mu_x' \iota\left(X_L^i\right)' \iota\left(X_L^{P+l}\right) \mu_x\right] \end{aligned}$$

- The above moments are no longer accurate if the distribution is truncated. The actual moments for a truncated multivariate normal are available in the literature if need arises.

1.2.2 Derivation of $p(x|rest)$

- The conditional posterior is characterized as:

$$
\begin{aligned}
\log p(x|rest) =& -\frac{\tau_y}{2}\left[(y-\Phi x)'(y-\Phi x)+\tau_x((x-r)-F\beta)'\Psi((x-r)-F\beta)\right]+c_1^x \\
=& -\frac{\tau_y}{2}\left[x'\Phi'\Phi x - x'\Phi' y - y'\Phi x + \tau_x x'\Psi x + \tau_x x'\Psi(r+F\beta)+(r+F\beta)'\Psi x\tau_x\right]+c_2^x \\
=& -\frac{\tau_y}{2}\left[x'(\Phi'\Phi+\tau_x\Psi)x - x'(\Phi' y+\tau_x\Psi(r+F\beta)) - \left(y'\Phi+\tau_x(r+F\beta)'\Psi\right)x\right]+c_2^x \\
=& -\frac{1}{2}(x-\mu_x)'\Lambda_x(x-\mu_x)+c_3^x \\
=& \log MN\left(x;\mu_x,\Lambda_x^{-1}\right)+c_4^x \\
& s.t. \\
\Lambda_x \equiv& \tau_y(\Phi'\Phi+\tau_x\Psi) \\
\mu_x \equiv& \tau_y\Lambda_x^{-1}(\Phi' y+\tau_x\Psi(r+F\beta)) \\
c_1^x \equiv& \frac{S+T}{2}\log\left(\frac{\tau_y}{2\pi}\right)+\frac{T}{2}\log\tau_x+\frac{1}{2}\log Det(\Psi) \\
& +\log\left[MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}[DA_0D]^{-1}\right)\right. \\
& \times\prod_{k=1}^{K}Bern(\gamma_k;\omega)\times Beta(\omega;\kappa_0,\delta_0)\times\prod_{t=1}^{T}Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
& \times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
& \times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
& \left.\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0})\times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\right]+c^{ev} \\
c_2^x =& c_1^x-\frac{\tau_y\tau_x}{2}(r+F\beta)'\Psi(r+F\beta)-\frac{\tau_y}{2}y'y \\
c_3^x =& c_2^x+\frac{\mu_x'\Lambda_x\mu_x}{2} \\
c_4^x =& c_3^x+\frac{T}{2}\log 2\pi-\log det\Lambda_x
\end{aligned}
$$

- The approximate posterior is:

$$
\begin{aligned}
\log q\left(x\right)= & E_{-x}\left[\frac{-\tau_{y}}{2}\left(\left(y-\Phi x\right)'\left(y-\Phi x\right)+\tau_{x}\left(\left(x-r\right)-F\beta\right)'\Psi\left(\left(x-r\right)-F\beta\right)\right)\right]+\overline{c}_{1}^{x}\\
= & \frac{-E\left[\tau_{y}\right]}{2}\Bigg(x'\left[E\left[\Phi'\Phi\right]+E\left[\tau_{x}\right]E\left[\Psi\right]\right]x-\left[y'E\left[\Phi\right]+E\left[\tau_{x}\right]\left(\mu_{\beta}'F'+r'\right)E\left[\Psi\right]\right]x\\
 & -x'\left[E\left[\Phi\right]'y+E\left[\tau_{x}\right]E\left[\Psi\right]\left(F\mu_{\beta}+r\right)\right]\Bigg)+\overline{c}_{2}^{x}\\
= & -\frac{1}{2}\left(x-\overline{\mu}_{x}\right)'\overline{\Lambda}_{x}\left(x-\overline{\mu}_{x}\right)+\overline{c}_{3}^{x}\\
= & \log\left(N\left(x;\overline{\mu}_{x},\overline{\Lambda}_{x}^{-1}\right)\right)+\overline{c}_{4}^{x}\\
 & s.t.\\
\overline{\Lambda}_{x}\equiv & E\left[\tau_{y}\right]\left(E\left[\Phi'\Phi\right]+E\left[\tau_{x}\right]E\left[\Psi\right]\right)\\
\overline{\mu}_{x}\equiv & E\left[\tau_{y}\right]\Lambda_{x}^{-1}\left(E\left[\Phi\right]'y+E\left[\tau_{x}\right]E\left[\Psi\right]\left(F\overline{\mu}_{\beta}+r\right)\right)\\
\overline{c}_{1}^{x}\equiv & E_{-x}\Bigg[\frac{S+T}{2}\log\frac{\tau_{y}}{2\pi}+\frac{T}{2}\log\tau_{x}+\frac{1}{2}\log det\Psi\\
 & +\log\Bigg(MN\left(\phi;\phi_{0},\frac{1}{\tau_{y}\tau_{\phi}}M_{0}^{-1}\right)\times MN\left(\beta;\beta_{0}+D^{-1}\beta_{0}^{\Delta},\frac{1}{\tau_{x}\tau_{y}\tau_{\beta}}\left[DA_{0}D\right]^{-1}\right)\\
 & \times\prod_{k=1}^{K}Bern\left(\gamma_{k};\omega\right)\times Beta\left(\omega;\kappa_{0},\delta_{0}\right)\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right)\\
 & \times Gamma\left(\tau_{x};\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_{y};\alpha_{y0},\zeta_{y0}\right)\\
 & \times Gamma\left(\tau_{\beta};\alpha_{\beta0},\zeta_{\beta0}\right)\times Gamma\left(\tau_{\phi};\alpha_{\phi0},\zeta_{\phi0}\right)\Bigg)\Bigg]+c^{ev}\\
\overline{c}_{2}^{x}\equiv & \overline{c}_{1}^{x}-E_{-x}\left[\frac{\tau_{y}}{2}\left(\tau_{x}\left(\beta'F'+r'\right)\Psi\left(F\beta+r\right)+y'y\right)\right]\\
\overline{c}_{3}^{x}\equiv & \overline{c}_{2}^{x}+\frac{\overline{\mu}_{x}'\overline{\Lambda}_{x}\overline{\mu}_{x}}{2}\\
\overline{c}_{4}^{x}\equiv & \overline{c}_{3}^{x}+\frac{T}{2}\log\left(2\pi\right)-\frac{1}{2}\log\left(det\overline{\Lambda}_{x}\right)
\end{aligned}
$$

- As an aside, in the code base we can test the expectation by verifying that, for any arbitrary value of x and simulated draws μ_{xj}, Λ_{xj}

$$
\begin{aligned}
-\frac{1}{2}\left(x-\mu_{x}\right)'\Lambda_{x}\left(x-\mu_{x}\right)+\frac{1}{2}\mu_{x}'\Lambda_{x}\mu_{x}= & \frac{1}{J}\sum_{j}\left(x-\hat{\mu}_{xj}\right)'\hat{\Lambda}_{xj}\left(x-\hat{\mu}_{xj}\right)+\frac{1}{2}\hat{\mu}_{xj}'\Lambda_{xj}\hat{\mu}_{x}\\
 & s.t.\\
\hat{\Lambda}_{xj}= & \tau_{yj}\left(\Phi_{j}'\Phi_{j}+\tau_{xj}\Psi\right)\\
\hat{\mu}_{xj}= & \tau_{yj}\Lambda_{xj}\left(\Phi_{j}'y+\tau_{xj}\Psi F\beta_{j}\right)
\end{aligned}
$$

where all indexed variables are drawn from their respective conditional posterior distributions.

  - The last term needs to be added back for testing purposes (it only affects the constant of proportionality)

- The above solution depends on knowing the $T \times T$ matrix $E\left[\Phi'\Phi\right]$. To begin, denote each row of Φ as Φ'_s such that

$$\Phi \equiv \begin{bmatrix} \Phi'_1 \\ \Phi'_2 \\ \vdots \\ \Phi'_S \end{bmatrix}$$

  - This implies the expectation can be written as:

$$E\Phi'\Phi = \sum_{s \in 1:S} E\Phi_s \Phi'_s$$

  - Each outer product then consists of the second moment matrix for $\tilde{\phi}$ padded by zeros:

$$\begin{aligned} E\left[\Phi_s \Phi'_s\right] &= \begin{bmatrix} \mathbf{0}^{UL} & \cdots & \vdots \\ \vdots & \tilde{M} & \vdots \\ \vdots & \cdots & \mathbf{0}^{LR} \end{bmatrix} \\ s.t. & \\ \mathbf{0}^{UL} &\equiv 0_{(t[s]-P-\Delta t)\times(t[s]-P-\Delta t)} \\ \tilde{M} &\equiv E\left[\tilde{\phi}\tilde{\phi}'\right] \\ \mathbf{0}^{LR} &\equiv 0_{(T-t[s])\times(T-t[s])} \end{aligned}$$

    * The extra Δt rows and columns of M are necessary due to the restriction. To see this, note $\tilde{M}$ can be written compactly as $E\left[\tilde{\phi}\tilde{\phi}'\right]$.

  - With ι_l^{ϕ} defined in Equation 3, construct $\tilde{M}$ as follows:

$$\tilde{M} = E\left[\tilde{\phi}_i \tilde{\phi}_j\right] = \begin{cases} E\left[\phi_i \phi'_j\right] & i \leq P \cap j \leq P \\ E\left[\phi_i \left(1 - \phi' \iota_{j-P}^{\phi}\right)\right] & i \leq P \cap j > P \\ E\left[\left(1 - \phi' \iota_{i-P}^{\phi}\right)\phi_j\right] & i > P \cap j \leq P \\ E\left[\left(1 - \phi' \iota_{i-P}^{\phi}\right)\left(1 - \phi' \iota_{j-P}^{\phi}\right)\right] & i > P \cap j > P \end{cases}$$

  - Define ι^{ϕ} as a $P \times \Delta t$ indicator matrix such that $\iota^{\phi} \equiv \begin{bmatrix} \iota_1^{\phi} & \ldots & \iota_{\Delta t}^{\phi} \end{bmatrix}$. Then addressing each part in turn:

$$\begin{aligned} E\left[\phi\phi'\right] &= \Lambda_{\phi}^{-1} + \mu_{\phi}\mu'_{\phi} \\ E\left[\phi\left(1_{\Delta t} - \left(\iota^{\phi}\right)'\phi\right)'\right] &= \mu_{\phi} 1'_{\Delta t} - E\left[\phi\phi'\right]\iota^{\phi} \\ E\left[\left(1_{\Delta t} - \left(\iota^{\phi}\right)'\phi\right)\left(1_{\Delta t} - \left(\iota^{\phi}\right)'\phi\right)'\right] &= 1_{\Delta t}1'_{\Delta t} - 1_{\Delta t}\mu'_{\phi}\iota^{\phi} - \left(\iota^{\phi}\right)'\mu_{\phi}1'_{\Delta t} + \left(\iota^{\phi}\right)' E\left[\phi\phi'\right]\iota^{\phi} \end{aligned}$$

1.2.3 Derivation of $p\left(\tau_y|rest\right)$

- Let $\tilde{\beta} \equiv \beta - \beta_0$. Then the conditional posterior is:

$$
\begin{aligned}
\log p\left(\tau_y|rest\right) =& \frac{-\tau_y}{2}\left(\tilde{y} - \tilde{X}_L\phi\right)'\left(\tilde{y} - \tilde{X}_L\phi\right) - \frac{\tau_y\tau_\phi}{2}\left(\phi - \phi_0\right)' M_0\left(\phi - \phi_0\right) \\
& - \frac{\tau_x\tau_y}{2}\left((x-r) - F\beta\right)'\Psi\left((x-r) - F\beta\right) \\
& - \frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) \\
& - \tau_y\zeta_{y0} + \left(\frac{S+T+P+K}{2} + \alpha_{y0} - 1\right)\log\tau_y + c_1^{\tau_y} \\
=& \log Gamma\left(\tau_y;\alpha_y,\zeta_y\right) + c_2^{\tau_y} \\
s.t.& \\
\alpha_y \equiv& \frac{S+T+P+K}{2} + \alpha_{y0} \\
\zeta_y \equiv& \zeta_{y0} + \frac{1}{2}\Bigg[\left(\tilde{y} - \tilde{X}_L\phi\right)'\left(\tilde{y} - \tilde{X}_L\phi\right) + \tau_\phi\left(\phi - \phi_0\right)' M_0\left(\phi - \phi_0\right) \\
& + \frac{\tau_x}{2}\left((x-r) - F\beta\right)'\Psi\left((x-r) - F\beta\right) \\
& + \frac{\tau_x\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta},\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)\Bigg] \\
c_1^{\tau_y} \equiv& + \alpha_{y0}\log\zeta_{y0} - \log\Gamma\left(\alpha_0\right) - \frac{S+T+K+P}{2}\log 2\pi + \frac{T+K}{2}\log\tau_x \\
& + \frac{1}{2}\log detM_0 + \frac{1}{2}\log det\Psi + \frac{1}{2}\log det\left(DA_0D\right) + \frac{P}{2}\log\tau_\phi + \frac{K}{2}\log\tau_\beta \\
& + \log\Bigg(\prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right) \times Beta\left(\omega;\kappa_0,\delta_0\right) \times \prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right) \\
& \times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right) \times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right) \\
& \times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right) \times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\Bigg)\Bigg] + c^{ev} \\
c_2^{\tau_y} \equiv& c_1^{\tau_y} + \log\Gamma\left(\alpha_y\right) - \alpha_y\log\zeta_y
\end{aligned}
$$

- Then the approximate unconditional posterior for τ_y is:

$$\begin{aligned}
\log q\left(\tau_y\right) =& E_{-\tau_y}\left[\frac{-\tau_y}{2}\left(\tilde{y}-\tilde{X}_L\phi\right)'\left(\tilde{y}-\tilde{X}_L\phi\right)-\frac{\tau_y\tau_\phi}{2}\left(\phi-\phi_0\right)'M_0\left(\phi-\phi_0\right)\right.\\
&-\frac{\tau_x\tau_y}{2}\left(\left(x-r\right)-F\beta\right)'\Psi\left(\left(x-r\right)-F\beta\right)\\
&-\frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta}-D^{-1}\beta_0^{\Delta}\right)'DA_0D\left(\tilde{\beta}-D^{-1}\beta_0^{\Delta}\right)\\
&\left.-\tau_y\zeta_{y0}+\left(\frac{S+T+P+K}{2}+\alpha_{y0}-1\right)\log\tau_y\right]+\bar{c}_1^{\tau_y}\\
=& \log\left(Gamma\left(\tau_y;\overline{\alpha}_y,\overline{\zeta}_y\right)\right)+\bar{c}_2^{\tau_y}\\
s.t.&\\
\overline{\alpha}_y \equiv& \frac{S+T+P+K}{2}+\alpha_{y0}\\
\overline{\zeta}_y \equiv& \frac{1}{2}\left(E\left[\tilde{y}'\tilde{y}\right]+E\left[\phi'\tilde{X}_L'\tilde{X}_L\phi\right]-E\left[\tilde{y}'\tilde{X}_L\right]\mu_\phi-\mu_\phi'E\left[\tilde{X}_L'\tilde{y}\right]\right.\\
&+E\left[g_y^{-1}\right]\left(E\left[\phi'M_0\phi\right]+\phi_0M_0\phi_0-\phi_0'M_0\mu_\phi-\mu_\phi'M_0\phi_0\right)\\
&+E\left[\tau_x\right]\left(E\left[x'\Psi x\right]+E\left[\left(\beta'F'+r'\right)\Psi\left(F\beta+r\right)\right]-\mu_x'E\left[\Psi\right]\left(F\mu_\beta+r\right)-\left(\mu_\beta'F'+r'\right)E\left[\Psi\right]\mu_x\right)\\
&\left.+E\left[\tau_x\right]E\left[\tau_\beta\right]\left(E\left[\tilde{\beta}'DA_0D\tilde{\beta}\right]+\left(\beta_0^{\Delta}\right)'A_0\beta_0^{\Delta}-\left(\beta_0^{\Delta}\right)'A_0E\left[D\right]E\left[\tilde{\beta}\right]-E\left[\tilde{\beta}\right]'E\left[D\right]A_0\beta_0^{\Delta}\right)\right)+\zeta_{y0}\\
\bar{c}_1^{\tau_y} \equiv& E_{-\tau_y}\left[-\frac{S+T+P+K}{2}\log 2\pi+\frac{1}{2}\log det\left(\Psi\right)+\frac{1}{2}\log det\left(M_0\right)+\frac{1}{2}\log det\left(DA_0D\right)\right.\\
&+\frac{T+K}{2}\log\tau_x+\alpha_{y0}\log\zeta_{y0}-\log\Gamma\left(\alpha_{y0}\right)+\frac{P}{2}\log\tau_\phi+\frac{K}{2}\log\tau_\beta\\
&+\log\left(\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\right.\\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Unif\left(\nu;\nu_0^{-},\nu_0^{+}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\\
&\left.\left.\times Gamma\left(\tau_\beta;\tau_{\beta0},\zeta_{\beta0}\right)\times Gamma\left(\tau_\phi;\tau_{\phi0},\zeta_{\phi0}\right)\right)\right]+c^{ev}\\
\bar{c}_2^{\tau_y} \equiv& \bar{c}_1^{\tau_y}-\overline{\alpha}_y\log\overline{\zeta}_y+\log\Gamma\left(\overline{\alpha}_y\right)
\end{aligned}$$

- The approximate posterior depends on the moments of multiple quadratic forms:
  - Start with $E\left[\phi'M_0\phi\right]$. Use Matrix Cookbook formula 318:

$$E\left[\phi'A_0\phi\right]=Tr\left(M_0\overline{\Lambda}_\phi^{-1}\right)+\overline{\mu}_\phi'M_0\overline{\mu}_\phi$$

  - Next compute $E\left[x'\Psi x\right]$. Since under the approximation Ψ is independent of x,the variables are independent, this is a straight-forward re-application of the same formula:

$$E\left[x'\Psi x\right]=Tr\left(E\left[\Psi\right]\overline{\Lambda}_x^{-1}\right)+\overline{\mu}_x'E\left[\Psi\right]\overline{\mu}_x$$

– A similar pattern applies to $E\left[\beta' F' \Psi F \beta\right]$:

$$E\left[\beta' F' \Psi F \beta\right] = Tr\left(F' E\left[\Psi\right] F \overline{\Lambda}_{\beta}^{-1}\right) + \overline{\mu}_{\beta}' F' E\left[\Psi\right] F \overline{\mu}_{\beta}$$

  * Hence:

$$\begin{aligned} E\left[\left(\beta' F' + r'\right) \Psi \left(F\beta + r\right)\right] =& Tr\left(F' E\left[\Psi\right] F \overline{\Lambda}_{\beta}^{-1}\right) + \overline{\mu}_{\beta}' F' E\left[\Psi\right] F \overline{\mu}_{\beta} \\ &+ r' E\left[\Psi\right] F \overline{\mu}_{\beta} + \overline{\mu}_{\beta}' F' E\left[\Psi\right] r + r' E\left[\Psi\right] r \\ =& Tr\left(F' E\left[\Psi\right] F \overline{\Lambda}_{\beta}^{-1}\right) + \left(\overline{\mu}_{\beta}' F' + r'\right) E\left[\Psi\right] \left(r + F\overline{\mu}_{\beta}\right) \end{aligned}$$

– To compute $E\left[\phi' \tilde{X}_L' \tilde{X}_L \phi\right]$, reference 1.2.2 for the derivation of $E\left[\tilde{X}_L' \tilde{X}_L\right]$. Then the prior pattern applies:

$$E\left[\phi' \tilde{X}_L' \tilde{X}_L \phi\right] = Tr\left(E\left[\tilde{X}_L' \tilde{X}_L\right] \overline{\Lambda}_{\phi}^{-1}\right) + \overline{\mu}_{\phi}' E\left[\tilde{X}_L' \tilde{X}_L\right] \overline{\mu}_{\phi}$$

– Compute $E\left[D\right]$ as a straight-forward discrete expectation:

$$E\left[D\right] = \overline{p}_{\gamma} + \left(1 - \overline{p}_{\gamma}\right) v^{-1}$$

– Calculate $E\left[\beta' D A_0 D \beta\right]$ in two steps:

1. First compute $E\left[D A_0 D\right]$. This step has a general case and a special case for when the matrix is diagonal.
   * Start with the diagonal case, which is relatively straight forward. This version assumes that A_0 is a diagonal matrix. The expectation of each element is given by:

$$E\left[d_k a_{0k} d_k\right] = a_{0k} \overline{p}_{\gamma k} + \frac{\left(1 - \overline{p}_{\gamma k}\right) a_{0k}}{v^2}$$

   * Then the general case
     · The expectation of each element is given by

$$\begin{aligned} E\left[D A_0 D\right] =& E\left[d d'\right] \odot A_0 \\ =& \begin{bmatrix} E\left[d_1^2\right] & E\left[d_1\right] E\left[d_2\right] & \cdots & E\left[d_1\right] E\left[d_K\right] \\ E\left[d_2\right] E\left[d_1\right] & E\left[d_2^2\right] & \cdots & E\left[d_2\right] E\left[d_K\right] \\ \vdots & \vdots & \ddots & \vdots \\ E\left[d_K\right] E\left[d_1\right] & E\left[d_K\right] E\left[d_2\right] & \cdots & E\left[d_K^2\right] \end{bmatrix} \odot A_0 \\ E\left[d_k^2\right] =& \overline{p}_{\gamma k} + \frac{\left(1 - \overline{p}_{\gamma k}\right)}{v^2} \end{aligned}$$

     where the last formula corresponds to the squared expectation of a Bernoulli.
2. Either way, apply the typical formulas:

$$\begin{aligned} E\left[\beta' D A_0 D \beta\right] =& Tr\left(E\left[D A_0 D\right] \overline{\Lambda}_{\beta}^{-1}\right) + \overline{\mu}_{\beta}' E\left[D A_0 D\right] \overline{\mu}_{\beta} \\ E\left[\tilde{\beta}' D A_0 D \tilde{\beta}\right] =& Tr\left(E\left[D A_0 D\right] \overline{\Lambda}_{\beta}^{-1}\right) + \left(\overline{\mu}_{\beta} - \beta_0\right)' E\left[D A_0 D\right] \left(\overline{\mu}_{\beta} - \beta_0\right) \end{aligned}$$

– Finally, recall that $\tilde{y} = y - x_S$. Then $E\left[\tilde{y}'\tilde{y}\right]$ (see section 1.2.1) for a justification and definition of $\iota\left(X_L^{P+1}\right)$:

$$
\begin{aligned}
E\left[\tilde{y}'\tilde{y}\right] =& E\left[\left(y - x_S\right)'\left(y - x_S\right)\right] \\
=& y'y - 2y'\mu_{xS} + E\left[x_S'x_S\right] \\
& s.t. \\
E\left[x_S'x_S\right] \equiv& \sum_{i=1}^{i=\Delta t}\sum_{j=1}^{j=\Delta t} E\left[x'\iota\left(X_L^{P+i}\right)'\iota\left(X_L^{P+j}\right)x\right] \\
& \sum_{i=1}^{i=\Delta t}\sum_{j=1}^{j=\Delta t}\left(tr\left[\iota\left(X_L^{P+i}\right)'\Lambda_x^{-1}\iota\left(X_L^{P+j}\right)\right] + \mu_x'\iota\left(X_L^{P+i}\right)'\iota\left(X_L^{P+j}\right)\mu_x\right)
\end{aligned}
$$

1.2.4 Derivation of $p\left(\tau_x|rest\right)$

- The conditional posterior:

$$
\begin{aligned}
\log p\left(\tau_x|rest\right) =& -\frac{\tau_x\tau_y}{2}\left(\left(x-r\right) - F\beta\right)'\Psi\left(\left(x-r\right) - r\right) - \frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) \\
& - \tau_x\zeta_{0x} + \left(\frac{T+K}{2} + \alpha_{0x} - 1\right)\tau_x + c_1^{\tau_x} \\
=& \log Gamma\left(\tau_x;\alpha_x,\zeta_x\right) + c_2^{\tau_x} \\
& s.t. \\
\alpha_x \equiv& \frac{T+K}{2} + \alpha_{0x} \\
\zeta_x \equiv& \frac{\tau_y}{2}\left(\left(x-r\right) - F\beta\right)'\Psi\left(\left(x-r\right) - F\beta\right) + \frac{\tau_y\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) + \zeta_{0x} \\
c_1 \equiv& \frac{T+K}{2}\log\frac{\tau_y}{2\pi} + \frac{1}{2}\log detDA_0D + \frac{1}{2}\log det\Psi + \frac{K}{2}\log\tau_\beta \\
& + \alpha_{y0}\log\zeta_{y0} - \log\Gamma\left(\alpha_{y0}\right) \\
& + \log\left(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\right. \\
& \times\prod_{k=1}^{K} Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right) \\
& \times\prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right) \\
& \left.\times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\right) + c^{ev} \\
c_2 \equiv& c_1 + \log\Gamma\left(\alpha_x\right) - \alpha_x\log\zeta_x
\end{aligned}
$$

- The approximate posterior:

$$
\begin{aligned}
\log q\left(\tau_x\right) =& E_{-\tau_x}\Bigg[-\frac{\tau_x\tau_y}{2}\left((x-r)-F\beta\right)'\Psi\left((x-r)-F\beta\right)-\frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta}-D^{-1}\beta_0^{\Delta}\right)'DA_0D\left(\tilde{\beta}-D^{-1}\beta_0^{\Delta}\right)\\
&-\tau_x\zeta_{x0}+\left(\frac{T+K}{2}+\alpha_{x0}-1\right)\log\tau_x\Bigg]+\overline{c}_1^{\tau_x}\\
=&\log Gamma\left(\tau_x;\overline{\alpha}_x,\overline{\zeta}_x\right)+\overline{c}_2^{\tau_x}\\
s.t.&\\
\overline{\alpha}_x \equiv& \frac{T+K}{2}+\alpha_{x0}\\
\overline{\zeta}_x \equiv& \zeta_{x0}+\frac{E\left[\tau_y\right]}{2}\left(E\left[x'\Psi x\right]+E\left[\left(\beta'F'+r'\right)\Psi\left(F\beta+r\right)\right]-\mu_x'E\left[\Psi\right]\left(F\mu_\beta+r\right)-\left(\mu_\beta'F'+r'\right)E\left[\Psi\right]\mu_x\right)\\
&+\frac{E\left[\tau_y\right]E\left[\tau_\beta\right]}{2}\left(E\left[\tilde{\beta}'DA_0D\tilde{\beta}\right]+\left(\beta_0^{\Delta}\right)'A_0\beta_0^{\Delta}-\left(\beta_0^{\Delta}\right)'A_0E\left[D\right]E\left[\tilde{\beta}\right]-E\left[\tilde{\beta}\right]'E\left[D\right]A_0\beta_0^{\Delta}\right)\\
\overline{c}_1^{\tau_x} =& E_{-\tau_x}\Bigg[-\frac{T+K}{2}\log 2\pi+\frac{1}{2}\log det\left(\Psi\right)+\frac{1}{2}\log det\left(DA_0D\right)\\
&+\frac{T+K}{2}\log\tau_y+\alpha_{x0}\log\zeta_{x0}-\log\Gamma\left(\alpha_{x0}\right)+\frac{K}{2}\log\tau_\beta\\
&+\log\Bigg(MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\\
&\times\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\\
&\times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\Bigg)\Bigg]+c^{ev}\\
\overline{c}_2^{\tau_x} =& \overline{c}_1^{\tau_x}-\overline{\alpha}_x\log\overline{\zeta}_x+\log\Gamma\left(\overline{\alpha}_x\right)
\end{aligned}
$$

- Refer to Section 1.2.3 for derivations of the above expectations

1.2.5 Derivation of $p(\tau_\phi|rest)$

$$\begin{aligned}
\log p(\tau_\phi|rest) =& \frac{P}{2}\log(\tau_\phi) - \frac{\tau_y\tau_\phi}{2}(\phi-\phi_0)' M_0(\phi-\phi_0) + (\alpha_{\phi 0}-1)\log\tau_\phi - \zeta_{\phi 0}\tau_\phi + c_1^{\tau\phi} \\
=& \log Gamma(\tau_\phi;\alpha_\phi,\zeta_\phi) + c_2^{\tau\phi} \\
& s.t. \\
\alpha_\phi =& \alpha_{\phi 0} + \frac{P}{2} \\
\zeta_\phi =& \zeta_{\phi 0} + \frac{\tau_y}{2}(\phi-\phi_0)' M_0(\phi-\phi_0) \\
c_1 =& \alpha_{\phi 0}\log\zeta_{\phi 0} - \log\Gamma(\alpha_{\phi 0}) + \frac{1}{2}\log Det(M_0) + \frac{P}{2}\log\left(\frac{\tau_y}{2\pi}\right) \\
& + \log\Bigg[MN\left(y;\Phi x, \frac{1}{\tau_y}I\right) \times MN\left(x; F\beta + r, \frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \times MN\left(\beta;\beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x\tau_y\tau_\beta}[DA_0D]^{-1}\right) \\
& \times \prod_{k=1}^{K} Bern(\gamma_k;\omega) \times Beta(\omega;\kappa_0,\delta_0) \times \prod_{t=1}^{T} Gamma(\psi;\nu/2,\nu/2) \times Gamma(\tau_\beta;\tau_{\beta 0},\zeta_{\beta 0}) \\
& \times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \times Gamma(\tau_x;\alpha_{x0},\zeta_{x0}) \times Gamma(\tau_y;\alpha_{y0},\zeta_{y0})\Bigg] + c^{ev} \\
c_2 =& c_1 - \alpha_\phi\log\zeta_\phi + \log\Gamma(\alpha_\phi)
\end{aligned}$$

Approximate posterior:

$$\begin{aligned}
\log q(\tau_\phi) =& E_{-\tau\phi}\left[-\frac{P}{2}\log(\tau_\phi) - \frac{\tau_y\tau_\phi}{2}(\phi-\phi_0)' M_0(\phi-\phi_0) + (\alpha_\phi - 1)\log\tau_\phi - \zeta_{\phi 0}\tau_\phi\right] + \overline{c}_1^{\tau\phi} \\
=& \log InvGamma\left(\tau_\phi;\overline{\alpha}_\phi,\overline{\zeta}_\phi\right) + \overline{c}_2^{\tau\phi} \\
& s.t. \\
\overline{\alpha}_\phi =& \alpha_{\phi 0} + \frac{P}{2} \\
\overline{\zeta}_\phi =& \zeta_{\phi 0} + E[\phi' M_0\phi] + \phi_0 M_0\phi_0 - \phi_0' M_0\mu_\phi - \mu_\phi' M_0\phi_0 \\
\overline{c}_1^{\tau\phi} =& \frac{P}{2}E\left[\log\left(\frac{\tau_y}{2\pi}\right)\right] + \frac{1}{2}\log det(M_0) + \alpha_{\phi 0}\log\zeta_{\phi 0} - \log\Gamma(\alpha_{\phi 0}) \\
& + E_{-gy}\log\Bigg[MN\left(y;\Phi x, \frac{1}{\tau_y}I\right) \times MN\left(x; F\beta, \frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \times MN\left(\beta;\beta_0 + D^{-1}\beta_0^{\Delta}, \frac{1}{\tau_x\tau_y\tau_\beta}[DA_0D]^{-1}\right) \\
& \times \prod_{k=1}^{K} Bern(\gamma_k;\omega) \times Beta(\omega;\kappa_0,\delta_0) \times \prod_{t=1}^{T} Gamma(\psi;\nu/2,\nu/2) \times Gamma(\tau_\beta;\tau_{\beta 0},\zeta_{\beta 0}) \\
& \times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \times Gamma(\tau_x;\alpha_{x0},\zeta_{x0}) \times Gamma(\tau_y;\alpha_{y0},\zeta_{y0})\Bigg] + c^{ev} \\
\overline{c}_2^{\tau\phi} =& c_1^{\tau\phi} - \overline{\alpha}_\phi\log\overline{\zeta}_\phi + \log\Gamma(\overline{\alpha}_\phi)
\end{aligned}$$

1.2.6 Derivation of $p(\tau_\beta|rest)$

Conditional posterior:

$$
\begin{aligned}
\log p\left(\tau_\beta|rest\right) =& -\frac{K}{2}\log\left(\tau_\beta\right) - \frac{\tau_y\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) + \left(\alpha_\beta - 1\right)\log\tau_\beta - \zeta_{x0}\tau_\beta + c_1^{\tau\beta} \\
=& \log Gamma\left(\tau_\beta; \alpha_\beta, \zeta_\beta\right) + c_2^{\tau\beta} \\
s.t.& \\
\alpha_\beta \equiv& \alpha_{\beta 0} + \frac{K}{2} \\
\zeta_\beta \equiv& \zeta_{\beta 0} + \frac{\tau_x\tau_y}{2}\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^{\Delta}\right) \\
c_1^{\tau\beta} \equiv& \alpha_{\beta 0}\log\zeta_{\beta 0} - \log\Gamma\left(\alpha_{\beta 0}\right) + \frac{K}{2}\log\frac{\tau_x\tau_y}{2\pi} + \frac{1}{2}\log detDA_0D \\
&+ \log\left(MN\left(\phi; \phi_0, \frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \times MN\left(y; \Phi x, \frac{1}{\tau_y}I\right)\right. \\
&\times \prod_{k=1}^{K} Bern\left(\gamma_k; \omega\right) \times Beta\left(\omega; \kappa_0, \delta_0\right) \\
&\times \prod_{t=1}^{T} Gamma\left(\psi; \nu/2, \nu/2\right) \times Gamma\left(\nu; \alpha_{\nu 0}, \zeta_{\nu 0}\right) \times Gamma\left(\tau_\phi; \tau_{\phi 0}, \zeta_{\phi 0}\right) \\
&\left.\times Gamma\left(\tau_y; \alpha_{y0}, \zeta_{y0}\right) \times Gamma\left(\tau_x; \alpha_{x0}, \zeta_{x0}\right)\right) + c^{ev} \\
c_2^{\tau\beta} \equiv& c_1^{\tau\beta} - \alpha_\beta\log\zeta_\beta + \log\Gamma\left(\alpha_\beta\right)
\end{aligned}
$$

Approximate unconditional posterior

$$\begin{aligned}
\log q\left(\tau_\beta\right) =& E_{-\tau\beta}\Bigg[-\frac{K}{2}\log\left(\tau_\beta\right)-\frac{\tau_y\tau_x\tau_\beta}{2}\left(\tilde{\beta}-D^{-1}\beta_0^\Delta\right)' DA_0D\left(\tilde{\beta}-D^{-1}\beta_0^\Delta\right)\\
&+\left(\alpha_{x0}^g-1\right)\log\tau_\beta\Bigg]-\zeta_\beta\tau_\beta+\overline{c}_1^{\tau\beta}\\
=& \log Gamma\left(\tau_\beta;\overline{\alpha}_\beta,\overline{\zeta}_\beta\right)+\overline{c}_2^{\tau\beta}\\
s.t.&\\
\overline{\alpha}_\beta =& \alpha_{\beta 0}+\frac{K}{2}\\
\overline{\zeta}_\beta =& \zeta_{\beta 0}+\frac{E\left[\tau_x\right]E\left[\tau_y\right]}{2}\left(E\left[\tilde{\beta}' DA_0D\tilde{\beta}\right]+\left(\beta_0^\Delta\right)' A_0\beta_0^\Delta-\left(\beta_0^\Delta\right)' A_0E\left[D\right]E\left[\tilde{\beta}\right]-E\left[\tilde{\beta}\right]' E\left[D\right]A_0\beta_0^\Delta\right)\\
\overline{c}_1^{\tau\beta} \equiv& E_{-\tau\beta}\Bigg[\alpha_{\beta 0}\log\zeta_{\beta 0}-\log\Gamma\left(\alpha_{x0}^g\right)+\frac{K}{2}\log\frac{\tau_x\tau_y}{2\pi}+\frac{1}{2}\log detDA_0D\\
&+\log\Bigg(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\\
&\times\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right)\times Gamma\left(\tau_\phi;\tau_{\phi 0},\zeta_{\phi 0}\right)\\
&\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\Bigg)\Bigg]+c^{ev}\\
\overline{c}_2^{\tau\beta} \equiv& -\overline{\alpha}_\beta\log\overline{\zeta}_\beta+\log\Gamma\left(\overline{\alpha}_\beta\right)
\end{aligned}$$

1.2.7 Derivation of $p\left(\beta|rest\right)$

Note that setting the multi-variate distribution to $MN\left(\beta;\beta_0+D^{-1}\beta_0^\Delta,\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)$ greatly improves tractability, particularly for the approximate unconditional posterior approximation.

First, define $\tilde{\beta}_0^{\Delta} \equiv \beta_0 - D^{-1}\beta_0$. The the conditional posterior $p(\beta|rest)$ is:

$$
\begin{aligned}
\log p(\beta|rest) =& -\frac{\tau_x\tau_y}{2}\left[\left((x-r)-F\beta\right)'\Psi\left((x-r)-F\beta\right)+\tau_\beta\left(\beta-\beta_0-D^{-1}\beta_0^{\Delta}\right)'DA_0D\left(\beta-\beta_0-D^{-1}\beta_0^{\Delta}\right)\right]+c_1^{\beta}\\
=& -\frac{\tau_x\tau_y}{2}\Bigg[\beta'F'\Psi F\beta+\beta'F'\Psi(x-r)+(x-r)'\Psi F\beta\\
&+\tau_\beta\beta'DA_0D\beta-\tau_\beta 2\beta'DA_0D\left(\beta_0+D^{-1}\beta_0^{\Delta}\right)\Bigg]+c_2^{\beta}\\
=& -\frac{1}{2}(\beta-\mu_\beta)'\Lambda_\beta(\beta-\mu_\beta)+c_3^{\beta}\\
=& \log MN(\beta;\mu_\beta,\Lambda_\beta)+c_4^{\beta}\\
s.t.&\\
\Lambda_\beta \equiv& \frac{\tau_x\tau_y}{2}\left[F'\Psi F+\tau_\beta DA_0D\right]\\
\mu_\beta \equiv& \frac{\tau_x\tau_y}{2}\Lambda_\beta^{-1}\left[F'\Psi(x-r)+\tau_\beta DA_0\left(D\beta_0+\beta_0^{\Delta}\right)\right]\\
c_1^{\beta} \equiv& \frac{T+K}{2}\log\frac{\tau_x\tau_y}{2\pi}+\frac{1}{2}\log detDA_0D+\frac{1}{2}\log det\Psi+\frac{K}{2}\log\tau_\beta\\
&-\frac{T}{2}\log T+\log\Bigg(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\\
&\times\prod_{k=1}^{K}Bern(\gamma_k;\omega)\times Beta(\omega;\kappa_0,\delta_0)\\
&\times\prod_{t=1}^{T}Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu0},\zeta_{\nu0})\\
&\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0})\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\\
&\times Gamma(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0})\times Gamma(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0})\Bigg)+c^{ev}\\
c_2^{\beta} =& c_1^{\beta}-\frac{\tau_x\tau_y}{2}\left((x-r)'\Psi(x-r)+\tau_\beta\left(D\beta_0+\beta_0^{\Delta}\right)'A_0\left(D\beta_0+\beta_0^{\Delta}\right)\right)\\
c_3^{\beta} =& c_2^{\beta}+\frac{\mu_\beta'\Lambda_\beta\mu_\beta}{2}\\
c_4^{\beta} =& c_3^{\beta}-\frac{1}{2}\log\Lambda_\beta+\frac{K}{2}\log 2\pi
\end{aligned}
$$

Derivation of the approximate unconditional posterior $q(\beta)$

$$
\begin{aligned}
\log q(\beta) =& -\frac{E[\tau_x]E[\tau_y]}{2}E_{-\beta}\Bigg[((x-r)-F\beta)'\Psi((x-r)-F\beta) \\
&+\tau_\beta\left(\beta-\beta_0-D^{-1}\beta_0^\Delta\right)'DA_0D\left(\beta-\beta_0-D^{-1}\beta_0^\Delta\right)\Bigg]+\overline{c}_1^\beta \\
=& -\frac{E[\tau_x]E[\tau_y]}{2}E_{-\beta}\Bigg[\beta'F'\Psi F\beta-(x'-r')\Psi F\beta-\beta'F'\Psi(x-r) \\
&+\tau_\beta\beta'DA_0D\beta-\tau_\beta 2\beta'DA_0D\left(\beta_0+D^{-1}\beta_0^\Delta\right)\Bigg]+\overline{c}_2^\beta \\
=& -\frac{1}{2}\left(\beta-\overline{\mu}_\beta\right)'\overline{\Lambda}_\beta\left(\beta-\overline{\mu}_\beta\right)+\overline{c}_3^\beta \\
=& \log\left[N\left(\beta;\overline{\mu}_\beta,\overline{\Lambda}_\beta\right)\right]+\overline{c}_4^\beta \\
& s.t. \\
\Lambda_\beta \equiv& E[\tau_x]E[\tau_y](F'E[\Psi]F+E[\tau_\beta]E[DA_0D]) \\
\mu_\beta \equiv& E[\tau_x]E[\tau_y]\overline{\Lambda}_\beta^{-1}\left(F'E[\Psi](\overline{\mu}_x-r)+E[\tau_\beta]\left(E[DA_0D]\beta_0+E[D]A_0\beta_0^\Delta\right)\right) \\
\overline{c}_1^\beta \equiv& E_{-\beta}\Bigg[\frac{T+K}{2}\log\left(\frac{\tau_x\tau_y}{2\pi}\right)+\frac{1}{2}\log det(\Psi)+\frac{1}{2}\log det(DA_0D)+\frac{K}{2}\log\tau_\beta \\
&+\log\Bigg(MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \\
&\times\prod_{k=1}^{K}Bern(\gamma_k;\omega)\times Beta(\omega;\kappa_0,\delta_0) \\
&\times\prod_{t=1}^{T}Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
&\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
&\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0})\times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\Bigg)\Bigg]+c^{ev} \\
\overline{c}_2^\beta \equiv& \overline{c}_1^\beta-E_{-\beta}\left[\frac{\tau_y\tau_x}{2}\left(\tau_\beta\left(D\beta_0+\beta_0^\Delta\right)'DA_0D\left(D\beta_0+\beta_0^\Delta\right)+(x-r)'\Psi(x-r)\right)\right] \\
\overline{c}_3^\beta \equiv& \overline{c}_2^\beta+\frac{1}{2}\overline{\mu}_\beta'\overline{\Lambda}_\beta\overline{\mu}_\beta \\
\overline{c}_4^\beta \equiv& \overline{c}_3^\beta+\frac{K}{2}\log(2\pi)-\frac{1}{2}\log det\left(\overline{\Lambda}_\beta\right)
\end{aligned}
$$

- Testing note- remember that the expectations are within $\overline{\mu}_\beta$ and $\overline{\Lambda}_\beta$, hence plugging in draws for the log normal distribution will not provide a consistent estimate.

1.2.8 Derivation of $p(\gamma)$ (Diagonal Case)

- The below derives the conditional posterior for γ_k in the scenario where each γ_k is conditionally independent of the other values of γ (denoted as γ_{-k}). In other words, $p(\gamma_k|\gamma_{-k},rest)=p(\gamma_k|rest)$. Note that this does not imply unconditionally that $\gamma_k\perp\gamma_{-k}$ as other variables (e.g. β) influence both γ_k and γ_{-k}.

- Practically this implies A_0 is diagonal, such that $a_0 \equiv diag\left(A_0\right)$
- Also recall $d_k^2 \equiv \gamma_k + \frac{1-\gamma_k}{v^2}$.

- As the only discrete distribution, the derivation for $p\left(\gamma\right)$ proceeds somewhat differently than others.
  - The distribution for p_k with a conditionally independent prior for β is given by $\frac{\tilde{p}(\gamma_k=1)}{\tilde{p}_k(\gamma_k=0)+\tilde{p}_k(\gamma_k=1)}$

$$
\begin{aligned}
\log p\left(\gamma_k\right) =& \log d_k - \frac{\tau_x\tau_y\tau_\beta d_k^2 a_{0k}}{2}\left(\tilde{\beta}_k - \frac{\beta_{0k}^{\Delta}}{d_k}\right)^2 + \gamma_k \log\omega + \left(1-\gamma_k\right)\log\left(1-\omega\right) + c_1^{\gamma k} \\
=& \log\left(d_k\right) - \frac{\tau_x\tau_y\tau_\beta d_k^2 a_{0k}}{2}\left(\tilde{\beta}_k - \frac{2\beta_{0k}^{\Delta}\tilde{\beta}_k}{d_k}\right) + \gamma_k \log\omega + \left(1-\gamma_k\right)\log\left(1-\omega\right) + c_2^{\gamma_k} \\
=& \gamma_k \log p_{\gamma k} + \left(1-\gamma_k\right)\log\left(1-p_{\gamma k}\right) + c_3^{\gamma_k} \\
& s.t.
\end{aligned}
$$

$$
\begin{aligned}
p_{\gamma k} =& \frac{\tilde{p}_{\gamma k}|_1}{\tilde{p}_{\gamma k}|_0 + \tilde{p}_{\gamma k}|_1} \\
\tilde{p}_{\gamma k}|_1 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta a_{0k}}{2}\left(\tilde{\beta}_k^2 - 2\beta_{0k}^{\Delta}\tilde{\beta}_k\right)\right)\omega \\
\tilde{p}_{\gamma k}|_0 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta a_{0k}}{2}\left(\frac{\tilde{\beta}_k^2}{v^2} - \frac{2\beta_{0k}^{\Delta}\tilde{\beta}_k}{v}\right)\right)\frac{1-\omega}{v} \\
c_1^{\gamma k} \equiv& \frac{1}{2}\log\frac{a_{0k}\tau_x\tau_y\tau_\beta}{2\pi} \\
& + \log\Bigg(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right) \\
& \times \prod_{j=1,j\neq k}^{K}\left(N\left(\beta_j;\frac{\beta_0^{\Delta}}{d_j}+\beta_0,\frac{1}{\tau_x\tau_y\tau_\beta a_{0j}d_j^2}\right)\times Bern\left(\gamma_j;\omega\right)\right)\times Beta\left(\omega;\kappa_0,\delta_0\right) \\
& \times \prod_{t=1}^{T} Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu 0},\zeta_{\nu 0}\right) \\
& \times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right) \\
& \times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0}\right)\Bigg) + c^{ev} \\
c_2^{\gamma k} \equiv& c_1^{\gamma k} - \frac{\tau_x\tau_y\tau_\beta a_{0k}\left(\beta_{0k}^{\Delta}\right)^2}{2} \\
c_3^{\gamma k} \equiv& c_2^{\gamma k} + \log\left(\tilde{p}_{\gamma k}|_1 + \tilde{p}_{\gamma k}|_0\right)
\end{aligned}
$$

  - Note that the normalization is accounted for in $c_3^{\gamma k}$. The normalization is fully revealed as the true probabilities must add to one.

- Similarly the approximate distribution for any γ_k is given by $\frac{\tilde{q}_k(\gamma_k=1)}{\tilde{q}_k(\gamma_k=0)+\tilde{q}_k(\gamma_k=1)}$

– Begin with the relevant (approximate) priors:

$$\begin{aligned}
\log q(\gamma_k) =& E_{-\gamma_k}\Bigg[\log(d_k) - \frac{\tau_x\tau_y\tau_\beta d_k^2 a_{0k}}{2}\left(\tilde{\beta}_k - \frac{\beta_{0k}}{d_k}\right)^2 \\
&+ \gamma_k \log(\omega) + (1-\gamma_k)\log(1-\omega)\Bigg] + \overline{c}_1^{\gamma_k} \\
=& -\frac{E[\tau_x]E[\tau_y]E[\tau_\beta]a_{0k}}{2}\left(d_k^2 E\left[\tilde{\beta}_k^2\right] - 2d_k E\left[\tilde{\beta}_k\right]\beta_{0k}^{\Delta}\right) \\
&+ \log(d_k) + \gamma_k E[\log(\omega)] + (1-\gamma_k)E[\log(1-\omega)] + \overline{c}_2^{\gamma_k} \\
=& \gamma_k \log \overline{p}_{\gamma k} + (1-\gamma_k)\log\left(1-\overline{p}_{\gamma k}\right) + \overline{c}_3^{\gamma_k} \\
& s.t. \\
\overline{p}_{\gamma k} \equiv& \frac{\tilde{q}(\gamma_k)|_1}{\tilde{q}(\gamma_k)|_1 + \tilde{q}(\gamma_k)|_0} \\
\tilde{q}(\gamma_k)|_1 =& \exp\left(-\frac{E[\tau_x]E[\tau_y]E[\tau_\beta]a_{0k}}{2}\left(E\left[\tilde{\beta}_k^2\right] - 2E\left[\tilde{\beta}_k\right]\beta_{0k}^{\Delta}\right) + E[\log(\omega)]\right) \\
\tilde{q}(\gamma_k)|_0 =& \exp\left(-\log(v) - \frac{E[\tau_x]E[\tau_y]E[\tau_\beta]a_{0k}}{2}\left(\frac{E\left[\tilde{\beta}_k^2\right]}{v^2} - \frac{2E\left[\tilde{\beta}_k\right]\beta_{0k}^{\Delta}}{v}\right) + E[\log(1-\omega)]\right) \\
\overline{c}_1^{\gamma_k} \equiv& E_{-\gamma_k}\Bigg[\frac{1}{2}\log\left(\frac{\tau_x\tau_y\tau_\beta a_{0k}}{2\pi}\right) \\
&+ \log\Bigg(MN\left(y;\Phi x, \frac{1}{\tau_y}I\right) \times MN\left(\phi;\phi_0, \frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \times MN\left(x; F\beta + r, \frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \\
&\prod_{j=1, j\neq k}^{K}\left(N\left(\beta_j; \frac{\beta_{0j}^{\Delta}}{d_j} + \beta_{0j}, \frac{1}{\tau_x\tau_y\tau_\beta d_j^2 a_{0j}}\right) \times Bern(\gamma_j;\omega)\right) \times Beta(\omega;\kappa_0,\delta_0) \\
&\times \prod_{t=1}^{T} Gamma(\psi;\nu/2,\nu/2) \times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
&\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0}) \times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
&\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}) \times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\Bigg)\Bigg] + c^{ev} \\
\overline{c}_2^{\gamma_k} \equiv& \overline{c}_1^{\gamma_k} - \frac{1}{2}E[\tau_x]E[\tau_y]E[\tau_\beta]a_{0k}\left(\beta_{0k}^{\Delta}\right)^2 \\
\overline{c}_3^{\gamma_k} \equiv& \overline{c}_2^{\gamma_k} + \log\left(\tilde{q}(\gamma_k)|_1 + \tilde{q}(\gamma_k)|_0\right)
\end{aligned}$$

- Compute the moments:
  - Derivation of $E\log\omega$ and $E\log(1-\omega)$
    - In subsequent sections, we show $q(\omega) = Beta\left(\omega;\overline{\kappa},\overline{\delta}\right)$
    - Plugging in the results from, Section 3.2:

$$\begin{aligned}
E\log(\omega) =& \digamma(\overline{\kappa}) - \digamma\left(\overline{\kappa}+\overline{\delta}\right) \\
E\log(1-\omega) =& \digamma\left(\overline{\delta}\right) - \digamma\left(\overline{\kappa}+\overline{\delta}\right)
\end{aligned}$$

where $\digamma(\cdot)$ is the digamma function.

- Derivation of $E\left[\beta_k^2\right]$ and $E\left[\beta_k\right]$:
    - These are just the marginals:

$$
\begin{aligned}
E\left[\beta_k\right] =& \overline{\mu}_{\beta k} \\
E\left[\beta_k^2\right] =& \frac{1}{\overline{\Lambda}_{\beta k}} + \overline{\mu}_{\beta k}^2
\end{aligned}
$$

    - Hence:

$$
\begin{aligned}
E\left[\tilde{\beta}_k\right] =& \overline{\mu}_{\beta k} - \beta_{0k} \\
E\left[\beta_k^2\right] =& \frac{1}{\overline{\Lambda}_{\beta k}} + \left(\overline{\mu}_{\beta k} - \beta_{0k}\right)^2
\end{aligned}
$$

1.2.9 Derivation of $p(\gamma)$ (General Case)

- As the only discrete distribution, the derivation for $p(\gamma)$ proceeds somewhat differently than others.
    - The distribution for q_k conditional on a conditionally independent prior is given by $\frac{\tilde{p}(\gamma_k=1)}{\tilde{p}_k(\gamma_k=0)+\tilde{p}_k(\gamma_k=1)}$
- In contrast with the conditionally independent approach, the below generalization allows for off-diagonal

terms for A_0.

$$
\begin{aligned}
p(\gamma_k) =& \log d_k - \frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta} - D^{-1}\beta_0^\Delta\right)' DA_0D\left(\tilde{\beta} - D^{-1}\beta_0^\Delta\right) \\
& + \gamma_k \log\omega + (1-\gamma_k)\log(1-\omega) + c_1^{\gamma k} \\
=& \log d_k - \frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta}' DA_0D\tilde{\beta} - 2\tilde{\beta}' DA_0\beta_0^\Delta\right) \\
& + \gamma_k \log\omega + (1-\gamma_k)\log(1-\omega) + c_2^{\gamma k} \\
=& \gamma_k \log p_{\gamma k} + (1-\gamma_k)\log(1-p_{\gamma k}) + c_3^{\gamma k} \\
s.t.& \\
p_{\gamma k} \equiv& \frac{\tilde{p}_{\gamma k}|_1}{\tilde{p}_{\gamma k}|_0 + \tilde{p}_{\gamma k}|_1} \\
\tilde{p}_{\gamma k}|_1 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta}{2}\left[\left(\tilde{\beta}' DA_0D\tilde{\beta} - 2\tilde{\beta}' DA_0\beta_0^\Delta\right)\right]_{d_k=1}\right)\omega \\
\tilde{p}_{\gamma k}|_0 \equiv& \exp\left(-\frac{\tau_x\tau_y\tau_\beta}{2}\left[\left(\tilde{\beta}' DA_0D\tilde{\beta} - 2\tilde{\beta}' DA_0\beta_0^\Delta\right)\right]_{d_k=v^{-1}}\right)\frac{1-\omega}{v} \\
c_1^{\gamma k} \equiv& \frac{K}{2}\log\frac{\tau_x\tau_y\tau_\beta}{2\pi} + \frac{1}{2}\log detA_0 + \sum_{j=1, j\neq k}^{K}\log d_j \\
& + \log\left(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\right. \\
& \times \prod_{j=1,j\neq k}^{K} Bern(\gamma_j;\omega)\times Beta(\omega;\kappa_0,\delta_0) \\
& \times \prod_{t=1}^{T} Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
& \times Gamma(\tau_y;\alpha_{y0},\zeta_{y0})\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0}) \\
& \left.\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0})\times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\right) + c^{ev} \\
c_2^{\gamma k} \equiv& c_1^{\gamma k} - \frac{\tau_x\tau_y}{2g_x}\left(\beta_0^\Delta\right)' A_0\beta_0^\Delta \\
c_3^{\gamma k} \equiv& c_2^{\gamma k} + \log\left(\tilde{p}_{\gamma k}|_1 + \tilde{p}_{\gamma k}|_0\right)
\end{aligned}
$$

- Note that the normalization is accounted for in $c_3^{\gamma k}$. The normalization is fully revealed as the true probabilities must add to one.
- The calculations need to be computed carefully to avoid overflow/underflow conditions.
  * A straight forward approach is to normalize the numerator and denominator

$$
\begin{aligned}
p_{\gamma k} =& \frac{\exp\left(\log\tilde{p}_{\gamma k}|_1\right)}{\exp\left(\log\tilde{p}_{\gamma k}|_0\right) + \exp\left(\log\tilde{p}_{\gamma k}|_1\right)} \\
=& \frac{\exp\left(\log\tilde{p}_{\gamma k}|_1 - \log h\right)}{\exp\left(\log\tilde{p}_{\gamma k}|_0 - \log h\right) + \exp\left(\log\tilde{p}_{\gamma k}|_1 - \log h\right)} \\
s.t.& \\
h \equiv& \max\left(\tilde{p}_{\gamma k}|_1, \tilde{p}_{\gamma k}|_0\right)
\end{aligned}
$$

* Note that storing the log of this value is necessary for recovering the log pdf in an overflow/underflow situation.

- The general approach creates performance issues for the approximation. Each value of γ_k affects each other value- hence the moments involving D must be recalculated K times. For comparison, x, β, and ϕ can all be drawn as a vector simultaneously.
  - As the only discrete distribution, the derivation for $q(\gamma)$ proceeds somewhat differently than others.
    * The distribution for any γ_k is given by $\frac{q_k(\gamma_k=1)}{\tilde{q}_k(\gamma_k=0)+\tilde{q}_k(\gamma_k=1)}$
  - Begin with the relevant (approximate) priors:

$$
\begin{aligned}
\log q(\gamma_k) =& E_{-\gamma_k}\Bigg[\log d_k - \frac{\tau_x\tau_y\tau_\beta}{2}\left(\tilde{\beta}-D^{-1}\beta_0^\Delta\right)' DA_0D\left(\tilde{\beta}-D^{-1}\beta_0^\Delta\right) \\
&+\gamma_k\log(\omega)+(1-\gamma_k)\log(1-\omega)\Bigg]+\bar{c}_1^{\gamma k} \\
=& -\frac{E[\tau_x]E[\tau_y]E[\tau_\beta]}{2}\left(E_{-\gamma k}\left[\tilde{\beta}'DA_0D\tilde{\beta}-2\tilde{\beta}'DA_0\beta_0^\Delta\right]\right) \\
&+\gamma_k E\log(\omega)+(1-\gamma_k)E\log(1-\omega)+\log d_k\Bigg]+\bar{c}_2^{\gamma_k} \\
=&\gamma_k\log\overline{p}_{\gamma k}+(1-\gamma_{\gamma k})\log\left(1-\overline{p}_{\gamma k}\right)+\bar{c}_3^{\gamma_k} \\
&s.t. \\
\overline{p}_{\gamma k} \equiv& \frac{\tilde{q}(\gamma_k)|_1}{\tilde{q}(\gamma_k)|_1+\tilde{q}(\gamma_k)|_0} \\
\tilde{q}(\gamma_k)|_1 =& \exp\left(-\frac{E[\tau_x]E[\tau_y]E[\tau_\beta]}{2}\left(E_{-\gamma k}\left[\tilde{\beta}'DA_0D\tilde{\beta}-2\tilde{\beta}'DA_0\beta_0^\Delta|d_k=1\right]\right)+E\log(\omega)\right) \\
\tilde{q}(\gamma_k)|_0 =& \exp\left(-\log(v)-\frac{E[\tau_x]E[\tau_y]E[\tau_\beta]}{2}\left(E_{-\gamma k}\left[\tilde{\beta}'DA_0D\tilde{\beta}-2\tilde{\beta}'DA_0\beta_0^\Delta|d_k=v^{-1}\right]\right)+E\log(1-\omega)\right) \\
\bar{c}_1^{\gamma k} \equiv& E_{-\gamma_k}\Bigg[\frac{K}{2}\log\left(\frac{\tau_x\tau_y\tau_\beta}{2\pi}\right)+\frac{1}{2}\log det(A_0)+\sum_{j=1,j\neq k}^{K}\log d_j \\
&+\log\Bigg(MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \\
&\prod_{j=1,j\neq k}^{K}(Bern(\gamma_j;\omega))\times Beta(\omega;\kappa_0,\delta_0) \\
&\times\prod_{t=1}^{T}Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
&\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
&\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0})\times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\Bigg)\Bigg]+c^{ev} \\
\bar{c}_2^{\gamma_k} \equiv& \bar{c}_1^{\gamma k}-\frac{1}{2}E_{-\gamma k}\left[\tau_x\tau_y\tau_\beta\left(\beta_0^\Delta\right)' A_0\beta_0^\Delta\right] \\
\bar{c}_3^{\gamma_k} \equiv& \bar{c}_2^{\gamma k}+\log\left(\tilde{q}(\gamma_k)|_1+\tilde{q}(\gamma_k)|_0\right)
\end{aligned}
$$

- The unconditional approximation requires an analogous approach to the conditional posterior with respect to the numerical calculation of the (log) probability.
- Compute the moments:
  - Derivation of $E\left[\tilde{\beta}' D A_0 D \tilde{\beta} - \tilde{\beta}' D A_0 \beta_0 - \beta_0' A_0 D \tilde{\beta} | d_k\right]$
    * Because by the mean-field approximation all variables are independent, the expectations separate and the result is the same as the unconditional but with d_k^2 and d_k substituted in for $E\left[d_k\right]$ and $E\left[d_k^2\right]$ (d_k is now a constant).
  - Derivation of $E \log \omega$ and $E \log\left(1-\omega\right)$
    * In subsequent sections, we show $q\left(\omega\right) = Beta\left(\omega; \overline{\kappa}, \overline{\delta}\right)$
    * Plugging in the results from, Section 3.2:

$$
\begin{aligned}
E \log\left(\omega\right) =& \digamma\left(\overline{\kappa}\right) - \digamma\left(\overline{\kappa} + \overline{\delta}\right) \\
E \log\left(1-\omega\right) =& \digamma\left(\overline{\delta}\right) - \digamma\left(\overline{\kappa} + \overline{\delta}\right)
\end{aligned}
$$

      where $\digamma\left(\cdot\right)$ is the digamma function.

1.2.10 Derivation $p\left(\omega\right)$

- The derivation for $p\left(\omega\right)$ is unique in that it is the only beta distributed variable. The conjugation is otherwise straight forward.

- The conditional posterior:

$$
\begin{aligned}
\log p(\omega) =& (\kappa_0 - 1)\log\omega + (\delta_0 - 1)\log(1-\omega) \\
&+ \sum_{k=1}^{K}\left[\gamma_k \log\omega + (1-\gamma_k)\log(1-\omega)\right] + c_1^{\omega} \\
=& (\kappa - 1)\log\omega + (\delta - 1)\log(1-\omega) + c_2^{\omega} \\
=& \log Beta(\kappa,\delta) + c_3^{\omega} \\
s.t.& \\
\kappa \equiv& \kappa_0 + \sum_{k=1}^{K}\gamma_k \\
\delta \equiv& \delta_0 + K - \sum_{k=1}^{K}\gamma_k \\
c_1^{\omega} \equiv& -\log B(\kappa_0,\delta_0) + \log\left(MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\right. \\
&\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right) \\
&\times \prod_{t=1}^{T} Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu0},\zeta_{\nu0}) \\
&\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0})\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0}) \\
&\left.\times Gamma(\tau_\beta;\tau_{\beta0},\zeta_{\beta0})\times Gamma(\tau_\phi;\tau_{\phi0},\zeta_{\phi0})\right) + c^{ev} \\
c_2^{\omega} \equiv& c_1^{\omega} + \log B(\kappa,\delta)
\end{aligned}
$$

- The approximate unconditional posterior:

$$\begin{aligned}
\log q(\omega) =& E_{-\omega}\Bigg[(\kappa_0-1)\log(\omega)+(\delta_0-1)\log(1-\omega) \\
&+\sum_k\left(\gamma_k\log(\omega)+(1-\gamma_k)\log(1-\omega)\right)+\overline{c}_1^{\omega}\Bigg] \\
=&(\overline{\kappa}-1)\log(\omega)+\left(\overline{\delta}-1\right)\log(1-\omega)+\overline{c}_1^{\omega} \\
=&\log\left(Beta\left(\overline{\kappa},\overline{\delta}\right)\right)+\overline{c}_2^{\omega} \\
s.t.& \\
\overline{\kappa}\equiv&\kappa_0+\sum_k\overline{p}_{\gamma k} \\
\overline{\delta}\equiv&\delta_0+K-\sum_k\overline{p}_{\gamma k} \\
\overline{c}_1^{\omega}\equiv& E_{-\omega}\Bigg[-\log B(\kappa_0,\delta_0)+\log\Bigg(MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right) \\
&\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right) \\
&\times\prod_{t=1}^{T}Gamma(\psi;\nu/2,\nu/2)\times Gamma(\nu;\alpha_{\nu 0},\zeta_{\nu 0}) \\
&\times Gamma(\tau_x;\alpha_{x0},\zeta_{x0})\times Gamma(\tau_y;\alpha_{y0},\zeta_{y0}) \\
&\times Gamma(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0})\times Gamma(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0})\Bigg)\Bigg]+c^{ev} \\
\overline{c}_2^{\omega}\equiv&\overline{c}_1^{\omega}+\log B(\kappa,\delta)
\end{aligned}$$

1.2.11 Derivation of $p(\psi_t)$

- The derivation for $p(\psi_t)$ is straight forward as the diagonal matrix of Ψ allows for component-wise treatment.

- The conditional posterior:

$$
\begin{aligned}
\log p\left(\psi_t\right) =& -\frac{\tau_x \tau_y \psi_t}{2}\left(\left(x_t-r_t\right)-f_t^{\prime} \beta\right)^2+\frac{1}{2} \log \psi_t+\left(\frac{\nu}{2}-1\right) \log \psi_t-\frac{\nu \psi_t}{2}+c_1^{\psi t} \\
=& \log Gamma\left(\psi_t, \alpha_{\psi t}, \zeta_{\psi t}\right)+c_2^{\psi t} \\
& s.t. \\
\alpha_{\psi t} \equiv& \frac{\nu+1}{2} \\
\zeta_{\psi t} \equiv& \frac{\nu}{2}+\frac{\tau_x \tau_y}{2}\left(\left(x_t-r_t\right)-f_t^{\prime} \beta\right)^2 \\
c_1^{\psi t} \equiv& \frac{\nu}{2} \log \frac{\nu}{2}-\log \Gamma\left(\frac{\nu}{2}\right)+\frac{1}{2} \log \left(\frac{\tau_x \tau_y}{2 \pi}\right) \\
& +\log \left(\prod_{j=1, j \neq t}^T\left[N\left(x_t ; f_t^{\prime} \beta+r, \frac{1}{\tau_x \tau_y \psi_t}\right) \times Gamma\left(\psi_j ; \frac{\nu}{2}, \frac{\nu}{2}\right)\right]\right. \\
& \times MN\left(\phi ; \phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right) \times MN\left(y ; \Phi x, \frac{1}{\tau_y} I\right) \\
& \times MN\left(\beta ; \beta_0+D^{-1} \beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta}\left[D A_0 D\right]^{-1}\right) \\
& \times \prod_{k=1}^K Bern\left(\gamma_k ; \omega\right) \times Beta\left(\omega ; \kappa_0, \delta_0\right) \times Gamma\left(\nu ; \alpha_{\nu 0}, \zeta_{\nu 0}\right) \\
& \times Gamma\left(\tau_y ; \alpha_{y 0}, \zeta_{y 0}\right) \times Gamma\left(\tau_x ; \alpha_{x 0}, \zeta_{x 0}\right) \\
& \left.\times Gamma\left(\tau_\beta ; \alpha_{\beta 0}, \zeta_{\beta 0}\right) \times Gamma\left(\tau_\phi ; \alpha_{\phi 0}, \zeta_{\phi 0}\right)\right)+c^{ev} \\
c_2^{\psi t} \equiv& c_1^{\psi t}-\alpha_{\psi t} \log \zeta_{\psi t}+\log \Gamma\left(\alpha_{\psi t}\right)
\end{aligned}
$$

- The approximate unconditional posterior:

$$
\begin{aligned}
\log q\left(\psi_t\right) =& E_{-\psi t}\left[-\frac{\tau_x \tau_y \psi_t}{2}\left(\left(x_t-r_t\right)-f_t^{\prime} \beta\right)^2+\frac{1}{2} \log \psi_t+\left(\frac{\nu}{2}-1\right) \log \psi_t-\frac{\nu \psi_t}{2}\right]+\bar{c}_1^{\psi t} \\
=& \log \left(\operatorname{Gamma}\left(\psi_t ; \alpha_{\psi t}, \zeta_{\psi t}\right)\right)+\bar{c}_2^{\psi t} \\
& s.t. \\
\bar{\alpha}_{\psi t} \equiv& \frac{E[\nu]}{2}+\frac{1}{2} \\
\bar{\zeta}_{\psi t} \equiv& \frac{E\left[\tau_x\right] E\left[\tau_y\right]}{2}\left(E\left[x_t^2\right]-2 \mu_{x t}\left(f_t^{\prime} \mu_\beta+r_t\right)+E\left[\left(\beta^{\prime} f_t+r_t\right)\left(f_t^{\prime} \beta+r_t\right)\right]\right)+\frac{E[\nu]}{2} \\
\bar{c}_1^{\psi t} \equiv& E_{-\psi t}\left[\frac{1}{2} \log \frac{\tau_x \tau_y}{2 \pi}+\frac{\nu}{2} \log \frac{\nu}{2}-\log \Gamma\left(\frac{\nu}{2}\right)\right. \\
& +\sum_{j=1, j \neq t}^T \log N\left(x_t ; f_t^{\prime} \beta+r, \frac{1}{\tau_x \tau_y \psi_t}\right)+\sum_{j=1, j \neq t}^T \log \operatorname{Gamma}\left(\psi_j ; \nu / 2, \nu / 2\right) \\
& +\log \left(M N\left(y ; \Phi x, \frac{1}{\tau_y} I\right) \times M N\left(\phi ; \phi_0, \frac{1}{\tau_y \tau_\phi} M_0^{-1}\right) \times M N\left(\beta ; \beta_0+D^{-1} \beta_0^{\Delta}, \frac{1}{\tau_x \tau_y \tau_\beta}\left[D A_0 D\right]^{-1}\right)\right. \\
& \times \prod_{k=1}^K \operatorname{Bern}\left(\gamma_k ; \omega\right) \times \operatorname{Beta}\left(\omega ; \kappa_0, \delta_0\right) \\
& \times \operatorname{Unif}\left(\nu ; \nu_0^{-}, \nu_0^{+}\right) \times \operatorname{Gamma}\left(\tau_x ; \alpha_{y 0}, \zeta_{y 0}\right) \times \operatorname{Gamma}\left(\tau_y ; \alpha_{y 0}, \zeta_{y 0}\right) \\
& \left.\left.\times \operatorname{Gamma}\left(\tau_\beta ; \alpha_{\beta 0}, \zeta_{\beta 0}\right) \times \operatorname{Gamma}\left(\tau_\phi ; \alpha_{\phi 0}, \zeta_{\phi 0}\right)\right)\right]+c^{e v} \\
\bar{c}_2^{\psi t} \equiv& \bar{c}_1^{\psi t}+\log \Gamma\left(\alpha_{\psi t}\right)-\alpha_{\psi t} log\left(\zeta_{\psi t}\right)
\end{aligned}
$$

- Now derive the non-standard moments:
  - From the Matrix Cookbook:

$$
E\left[\left(\beta^{\prime} f_t+r_t\right)\left(f_t^{\prime} \beta+r_t\right)\right]=Tr\left(f_t f_t^{\prime} \overline{\Lambda}_\beta^{-1}\right)+\left(\bar{\mu}_\beta^{\prime} f_t+r_t\right)\left(f_t^{\prime} \bar{\mu}_\beta+r_t\right)
$$

  - $E\left[x_t^2\right]$ is straight forward:

$$
E\left[x_t^2\right]=\frac{1}{\overline{\Lambda}_{x t}}+\bar{\mu}_{x t}^2
$$

  - In contrast, the expectation of $E[\nu]$ is complex and is computed through brute-force numerical integration. See Section 1.2.12 for details.

1.2.12 Derivation of $p(\nu)$

- The derivation of $p(\nu)$ is non-standard. No conjugate prior exists.
  - The prior gamma distribution is truncated from below by ν_{-}

- The conditional posterior:

$$\log p(\nu) = \frac{T\nu}{2}\log\frac{\nu}{2} - T\Gamma\left(\frac{\nu}{2}\right) + \sum_{t\in 1:T}\left[\left(\frac{\nu}{2}-1\right)\log\psi_t - \frac{\nu}{2}\psi_t\right] + (\alpha_{\nu 0}-1)\log\nu - \zeta_{\nu 0}\nu + c_1^{\nu}$$
$$= \left(\frac{T\nu}{2}+\alpha_{\nu 0}-1\right)\log\frac{\nu}{2} - T\log\Gamma\left(\frac{\nu}{2}\right) + \frac{\nu}{2}\eta_1 + c_2^{\nu}$$
$$p(\nu) = \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}+\alpha_{\nu 0}-1}\Gamma^{-T}\left(\frac{\nu}{2}\right)\exp\left(\frac{\nu\eta_1}{2}\right)\eta_2\exp c_3^{\nu}$$

$s.t.$

$$\eta_1 \equiv \sum_{t\in 1:T}(\log\psi_t - \psi_t) - 2\zeta_{\nu 0}$$
$$\eta_2 \equiv \left[\int_{\nu^-}^{\infty}\left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}+\alpha_{\nu 0}-1}\Gamma^{-T}\left(\frac{\nu}{2}\right)\exp\left(\frac{\nu\eta_1}{2}\right)d\nu\right]^{-1}$$
$$\begin{aligned}c_1^{\nu} \equiv&\alpha_{\nu 0}\log\zeta_{\nu 0} - \log\Gamma(\alpha_{\nu 0})\\ &+\log\left(MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)\right.\\ &\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\times\prod_{t=1}^{T}Gamma\left(\psi_t;\nu/2,\nu/2\right)\\ &\times\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\ &\times Gamma\left(\tau_x;\alpha_{y0},\zeta_{y0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\\ &\left.\left.\times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0}\right)\right)\right] + c^{ev}\end{aligned}$$
$$c_2^{\nu} \equiv c_1^{\nu} + (\alpha_{\nu 0}-1)\log 2 - \sum_{t\in 1:T}\log\psi_t$$
$$c_3^{\nu} \equiv c_2^{\nu} - \log\eta_2$$

  - Integration:
    * This is unnecessary in the Metropolis-Hastings procedure.
  - The lack of a conjugate prior necessitates a modified approach to MCMC sampling.
    * The project uses an independent flavor Metropolis-Hastings sampler to draw from the above conditional distribution.
    * Metropolis-Hastings procedures require a proposal distribution denoted as $r(\nu)$. The proposal must have fatter tails then the focal distribution. By default, the procedure uses the prior distribution as the proposal. This may be inefficient and warrant adjustment, particularly when the prior is diffuse.
    * For each iteration, the sampling procedure is as follows. Let ν^0 represent the previous iteration of ν. The goal is to select ν^1, the value of ν in the subsequent iteration.

    1. Propose a value of ν, denoted as ν', by drawing from the proposal distribution $r(\nu)$.

2. Evaluate

$$m\left(\nu^0,\nu'\right)=\frac{p\left(\nu'|rest'\right)r\left(\nu^0\right)}{p\left(\nu^0|rest'\right)r\left(\nu'\right)}$$

Note that the probability density using the previous value of ν is conditioned on the CURRENT value of the rest of the parameters.

3. If $m\left(\nu^0,\nu'\right)\geq 1$, accept the proposal by setting $\nu^1=\nu'$
4. If $m\left(\nu^0,\nu'\right)<1$:
   (a) Draw u where $u\sim Unif\,(0,1)$.
   (b) If $u>m$, reject the proposal and let $\nu^1=\nu^0$.
   (c) Otherwise, accept the proposal and let $\nu^1=\nu'$.

- The approximate unconditional posterior:

$$\begin{aligned}
\log q\left(\nu\right) \propto & E_{-\nu}\left[\frac{T\nu}{2}\log\left(\frac{\nu}{2}\right)-T\log\Gamma\left(\frac{\nu}{2}\right)+\sum_t\left(\left(\frac{\nu}{2}-1\right)\log\psi_t-\frac{\nu\psi_t}{2}\right)+\left(\alpha_{\nu 0}-1\right)\log\left(\nu\right)-\zeta_{\nu 0}\nu\right]+c_1^\nu \\
= & \left(\frac{T\nu}{2}+\alpha_{\nu 0}-1\right)\log\left(\frac{\nu}{2}\right)-T\log\Gamma\left(\frac{\nu}{2}\right)+\frac{\nu}{2}\eta_1+c_2^\nu \\
q\left(\nu\right) = & \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}+\alpha_{\nu 0}-1}\Gamma^{-T}\left(\frac{\nu}{2}\right)\eta_2\exp\left(\frac{\nu\eta_1}{2}\right)\exp c_3^\nu \\
& s.t. \\
c_1^\nu \equiv & E_{-\nu}\Bigg[\alpha_{\nu 0}\log\zeta_{\nu 0}-\log\Gamma\left(\alpha_{\nu 0}\right) \\
& +\log\Bigg(MN\left(y;\Phi x,\frac{1}{\tau_y}I\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right) \\
& \times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^\Delta,\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right) \\
& \times\prod_{k=1}^K Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right) \\
& \times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right) \\
& \times Gamma\left(\tau_\beta;\alpha_{\beta 0},\zeta_{\beta 0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi 0},\zeta_{\phi 0}\right)\Bigg)\Bigg] \\
c_2^\nu \equiv & c_1^\nu-\sum_t E\log\psi_t+\left(\alpha_{\nu 0}-1\right)\log\left(2\right) \\
c_3^\nu \equiv & c_2^\nu-\log\eta_2 \\
\overline{\eta}_1 = & \sum_t E\left[\log\psi_t-\psi_t\right]-2\zeta_{\nu 0} \\
\overline{\eta}_2 \equiv & \left(\int_{\nu^-}^{\infty}\left[\left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}+\alpha_{\nu 0}-1}\Gamma^{-T}\left(\frac{\nu}{2}\right)\exp\left(\frac{\nu\eta_1}{2}\right)\right]d\nu\right)^{-1}
\end{aligned}$$

- Brute-force integration gives the expectation of ν:

$$E[\nu] = \int_{\nu^-}^{\infty} \nu \times \left(\frac{\nu}{2}\right)^{\frac{T\nu}{2}+\alpha_{\nu 0}-1} \Gamma^{-T}\left(\frac{\nu}{2}\right)\eta_2 \exp\left(\frac{\nu\eta_1}{2}\right) d\nu$$

  - Compute using adaptive Gaussian-Konrod quadrature (quadgk)

- The definition of $\overline{\eta}_1$ depends on calculating $E\log(\psi_t)$:

$$E\log\psi_t = \digamma(\alpha_{\psi t}) - \log(\zeta_{\psi t})$$

- As a weakly informative prior, consider the following premises:
  - The first four moments of x_t most likely exist unconditionally, ($\nu > 4$, probably)
    - Furthermore, assume $\nu > 2$ always. This is akin to assuming that the mean and variance of the residuals of x_t exists and is finite. The support of the distribution is therefore bounded from below.
  - The PDF should smoothly decay to zero at zero to reflect the low probabilities of extremely pathological distributions.
  - The prior should reflect significant uncertainty around a wide range of values.
  - Conservatism- better to miss with a lower value of ν then a higher one. Formally this could be accounted for via the loss function, but as a first cut prefer more conservative estimates.

1.3 Predictive Distributions

1.3.1 Predictions of $y|F,\Theta$

- Conditional on the state and other variables, the observation equation is straight forward.
- The conditional independence of y further simplifies the math
- Let $y^u \subset y$ be the subset of y to be predicted and y^m be its complement.
  - Suppose there are S^u values to predict, such that y^u is $S^u \times 1$. Then let Φ^u and X_L^u be a subset of their complete forms, such that they represent $S^p \times T$ and $S^u \times (P+1)$ matrices containing only rows for unobserved values of y.

- Then the conditional distribution of the missing observed values of y follows as:

$$
\begin{aligned}
\log p\left(y^{u}\right) =& -\frac{\tau_y}{2}\left(y^{u}-\tilde{X}_L^{u}\phi-x_S^{u}\right)'\left(y^{u}-\tilde{X}_L^{u}\phi-x_S^{u}\right)+c_1^{yu}\\
&=\log\left(MN\left(y^{u};\mu_{yu},\Lambda_{yu}\right)\right)+c_2^{yu}\\
&s.t.\\
\mu_{yu}\equiv&\tilde{X}_L^{u}\phi+x_S^{u}\\
\Lambda_{yu}\equiv&I_{Su}\tau_y\\
c_1^{yu}\equiv&\left(\frac{S^{u}}{2}\right)\log\frac{\tau_y}{2\pi}+\log\Bigg(MN\left(y^{m};\Phi x^{m},\frac{1}{\tau_y}I\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x}\Psi^{-1}\right)\\
&\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^{\Delta},\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)\\
&\times\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right)\\
&\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\\
&\times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\Bigg)+c^{ev}\\
c_2^{yu}\equiv&c_1^{yu}-\left(\frac{S^{u}}{2}\right)\log\frac{\tau_y}{2\pi}
\end{aligned}
$$

- The approximate predictive follows the usual pattern:

$$
\begin{aligned}
\log q\left(y^u\right) =& E_{-y_u}\left[-\frac{\tau_y}{2}\left(y^u-\tilde{X}_L^u\phi-x_S^u\right)'\left(y^u-\tilde{X}_L^u\phi-\Delta t x_S^u\right)+c_1^{yu}\right]+\tilde{c}_1^{yu}\\
=& \log\left(MN\left(y^u;\mu_{yu},\Lambda_{yu}^{-1}\right)\right)+\tilde{c}_2^{yu}\\
\tilde{\mu}_y^u \equiv& E\left[\tilde{X}_L^u\right]\mu_\phi+\mu_{xs}^u\\
\tilde{\Lambda}_{yu} \equiv& I_{Su}E\left[\tau_y\right]\\
\tilde{c}_1^{yu} =& E_{-yu}\left[\left(\frac{S^u}{2}\right)\log\frac{\tau_y}{2\pi}+\log\left(MN\left(y^m;\Phi x^m,\frac{1}{\tau_y}I\right)\times MN\left(x;F\beta+r,\frac{1}{\tau_x\tau_y}\Psi^{-1}\right)\right.\right.\\
&\times MN\left(\phi;\phi_0,\frac{1}{\tau_y\tau_\phi}M_0^{-1}\right)\times MN\left(\beta;\beta_0+D^{-1}\beta_0^\Delta,\frac{1}{\tau_x\tau_y\tau_\beta}\left[DA_0D\right]^{-1}\right)\\
&\times\prod_{k=1}^{K}Bern\left(\gamma_k;\omega\right)\times Beta\left(\omega;\kappa_0,\delta_0\right)\\
&\times\prod_{t=1}^{T}Gamma\left(\psi;\nu/2,\nu/2\right)\times Gamma\left(\nu;\alpha_{\nu0},\zeta_{\nu0}\right)\\
&\times Gamma\left(\tau_x;\alpha_{x0},\zeta_{x0}\right)\times Gamma\left(\tau_y;\alpha_{y0},\zeta_{y0}\right)\\
&\left.\left.\times Gamma\left(\tau_\beta;\alpha_{\beta0},\zeta_{\beta0}\right)\times Gamma\left(\tau_\phi;\alpha_{\phi0},\zeta_{\phi0}\right)\right)\right]+c^{ev}\\
\tilde{c}_2^{yu} =& \tilde{c}_1^{yu}-\left(\frac{S^u}{2}\right)\log\frac{E\left[\tau_y\right]}{2\pi}+\frac{\left(\tilde{\mu}_y^u\right)'\tilde{\Lambda}_{yu}\tilde{\mu}_y^u}{2}\\
&-E\left[\frac{\tau_y}{2}\left(\tilde{X}_L^u\phi-\Delta t x_S^u\right)'\left(\tilde{X}_L^u\phi-\Delta t x_S^u\right)\right]
\end{aligned}
$$

1.4 Prediction of $y|F,\Theta_{-x}$ INCOMPLETE and UNTESTED

It is useful to integrate out x and derive the predictions based on the remaining parameters, given that x is unobservable.

$$
\begin{aligned}
\log p\left(x|rest\right) =& -\frac{\tau_y}{2}\left[\left(y-\Phi x\right)'\left(y-\Phi x\right)+\tau_x\left(\left(x-r\right)-F\beta\right)'\Psi\left(\left(x-r\right)-F\beta\right)\right]+c_1^x\\
=& -\frac{\tau_y}{2}\left[x'\Phi'\Phi x-x'\Phi'y-y'\Phi x+\tau_x x'\Psi x+\tau_x x'\Psi\left(r+F\beta\right)+\left(\rho r+F\beta\right)'\Psi x\tau_x\right]+c_2^x\\
=& -\frac{\tau_y}{2}\left[x'\left(\Phi'\Phi+\tau_x\Psi\right)x-x'\left(\Phi'y+\tau_x\Psi\left(r+T^{-1/2}F\beta\right)\right)-\left(y'\Phi+\tau_x\left(r+T^{-1/2}F\beta\right)'\Psi\right)x\right]+c_2^x\\
=& -\frac{1}{2}\left(x-\mu_x\right)'\Lambda_x\left(x-\mu_x\right)+c_3^x\\
=& \log MN\left(x;\mu_x,\Lambda_x^{-1}\right)+c_4^x\\
s.t.&\\
\Lambda_x \equiv& \Phi'\Phi+\tau_x\Psi\\
\mu_x \equiv& \tau_y\Lambda_x^{-1}\left(\Phi'y+\tau_x\Psi\left(r+T^{-1/2}F\beta\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
p\left(y|F,\Theta_{-x}\right) \propto & \int_x \exp\left(-\frac{\tau_y}{2}\left[\left(y-\Phi x\right)'\left(y-\Phi x\right)+\tau_x\left(\left(x-r\right)-F\beta\right)'\Psi\left(\left(x-r\right)-F\beta\right)\right]\right)dx\times C_1^{yMx} \\
= & \exp\left(-\frac{\tau_y}{2}y'y\right)\int_x \exp\Bigg(-\frac{\tau_y}{2}\Bigg[x'\Phi'\Phi x-x'\Phi'y-y'\Phi x \\
& +\tau_x x'\Psi x+\tau_x x'\Psi\left(r+F\beta\right)+\left(\rho r+F\beta\right)'\Psi x\tau_x\Bigg]\Bigg)dx\times C_2^{yMx} \\
= & \exp\left(-\frac{\tau_y}{2}y'y\right)\int_x \exp\Bigg(-\frac{\tau_y}{2}\Bigg[x'\left(\Phi'\Phi+\tau_x\Psi\right)x \\
& -x'\left(\Phi'y+\tau_x\Psi\left(r+F\beta\right)\right)-\left(y'\Phi+\tau_x\left(r+F\beta\right)'\Psi\right)x\Bigg]\Bigg)dx\times C_2^{yMx} \\
= & \exp\left(-\frac{\tau_y}{2}y'y+\frac{1}{2}\mu_x'\Lambda_x\mu_x\right)\int_x \exp\left(-\frac{1}{2}\left(x-\mu_x\right)'\Lambda_x\left(x-\mu_x\right)\right)dx\times C_2^{yMx} \\
= & \exp\left(-\frac{\tau_y}{2}y'y+\frac{1}{2}\mu_x'\Lambda_x\mu_x\right)\int_x MN\left(x;\mu_x,\Lambda_x^{-1}\right)dx\times C_3^{yMx} \\
& s.t. \\
\Lambda_x \equiv & \tau_y\left(\Phi'\Phi+\tau_x\Psi\right) \\
\mu_x \equiv & \tau_y\Lambda_x^{-1}\left(\Phi'y+\tau_x\Psi\left(r+F\beta\right)\right)
\end{aligned}
$$

This implies a predictive distribution given by:

$$
\begin{aligned}
\log p\left(y|F,\Theta_{-x}\right) \propto & -\frac{\tau_y}{2}y'y+\frac{\tau_y^2}{2}\left(\Phi'y+\tau_x\Psi\left(r+F\beta\right)\right)'\Lambda_x^{-1}\left(\Phi'y+\tau_x\Psi\left(r+F\beta\right)\right)+c_3^{yMx} \\
= & -\frac{\tau_y}{2}\left(y'y-\tau_y y'\Phi\Lambda_x^{-1}\Phi'y-\tau_y\tau_x y'\Phi\Lambda_x^{-1}\Psi\left(r+F\beta\right)-\tau_y\tau_x\left(r+F\beta\right)\Psi\Lambda_x^{-1}\Phi'y\right)+c_4^{yMx} \\
= & -\frac{1}{2}\left(y-\mu_{yMx}\right)'\Lambda_{yMx}\left(y-\mu_{yMx}\right)+c_4^{yMx} \\
= & \log MN\left(y;\mu_{yMx},\Lambda_{yMx}^{-1}\right)+c_5^{yMx} \\
& s.t. \\
\Lambda_{yMx} = & \tau_y\left(I_S-\tau_y\Phi\Lambda_x^{-1}\Phi'\right) \\
= & \tau_y\left(I_S-\Phi\left[\Phi'\Phi+\tau_x\Psi\right]^{-1}\Phi'\right) \\
\mu_{yMx} = & \tau_y^2\tau_x\Lambda_{yMx}^{-1}\Phi\Lambda_x^{-1}\Psi\left(r+F\beta\right) \\
= & \tau_y\tau_x\Lambda_{yMx}^{-1}\Phi\left[\Phi'\Phi+\tau_x\Psi\right]^{-1}\Psi\left(r+F\beta\right)
\end{aligned}
$$

Note because $\tau_x\Psi$ is strictly positive, the distribution should always exist. Also while precision increases with τ_y and τ_x as expected, the precision is bounded from above by τ_y. Further intuition is possible through manipulation of μ_{yMx} using the Woodbury matrix identity (Equation 156 in the MCB) and a Searle identity

(Equation 163 in the MCB).

$$\begin{aligned}
\mu_{yMx} =& \tau_x \left[I_S - \Phi \left[\Phi'\Phi + \tau_x \Psi \right]^{-1} \Phi' \right]^{-1} \Phi \left[\Phi'\Phi + \tau_x \Psi \right]^{-1} \Psi \left(r + F\beta \right) \\
=& \tau_x \left(I_S - \Phi \left(- \left[\Phi'\Phi + \tau_x \Psi \right] + \Phi'\Phi \right)^{-1} \Phi' \right) \Phi \left[\Phi'\Phi + \tau_x \Psi \right]^{-1} \Psi \left(r + F\beta \right) \\
=& \tau_x \left[I_S + \Phi \left(\tau_x \Psi \right)^{-1} \Phi' \right] \Phi \left[\Phi'\Phi + \tau_x \Psi \right]^{-1} \Psi \left(r + F\beta \right) \\
=& \tau_x \Phi \left[I_T + \left(\tau_x \Psi \right)^{-1} \Phi'\Phi \right] \left[\Phi'\Phi + \tau_x \Psi \right]^{-1} \Psi \left(r + F\beta \right) \\
=& \tau_x \Phi \left[I_T + \left(\tau_x \Psi \right)^{-1} \Phi'\Phi \right] \left[\Phi'\Phi \right]^{-1} \left[\left[\Phi'\Phi \right]^{-1} + \left[\tau_x \Psi \right]^{-1} \right]^{-1} \left[\tau_x \Psi \right]^{-1} \Psi \left(r + F\beta \right) \\
=& \Phi \left[\left[\Phi'\Phi \right]^{-1} + \left(\tau_x \Psi \right)^{-1} \right] \left[\left[\Phi'\Phi \right]^{-1} + \left[\tau_x \Psi \right]^{-1} \right]^{-1} \left(r + F\beta \right) \\
=& \Phi \left(r + F\beta \right)
\end{aligned}$$

This implies $E\left[\Phi \left(r + F\beta \right) \right]$ provides the mean of the predictive distribution, though the other moments of a distribution naively calculated $\Phi \left(r + F\beta \right)$ will not match the true predictive distribution.

Note this does not quite provide a recipe for scenario analysis, as Ψ is still indeterminate if the data are new. However it is possible that ψ_t is conditionally *iid*, even without conditioning on x, which would imply any ψ_t could be drawn without issue.

2 Supporting material

2.1 Additional model notes

- The core of the generative model follows in the style of George and Mccullock 1993/1997, Ishawaran and Rao 2006, and Rockova and George 2014. But there are significant differences from both of these:
  - This approach uses Variational Bayes (VB) for calculations as opposed to MCMC. Most VB papers pertaining to model selection use a delta spike instead of a normal distribution for spike. On the other hand, George and Mccullock recommend using a low variance normal spike if there are magnitudes of the coefficients which are not under consideration. This is the approach taken in this model, and it leads to some nice properties with respect to the posterior distributions. However, the spike-based approach of Carbonetto and Stephens 2012, among others, remains a viable back-up option.
  - This model includes an extra layer of latent variables, leading to complexity with respect to the measurement.

- This version uses a diagonal matrix among latent variables to account for heteroskedasticity. Wand et al 2011 (sec. 4.1) and Geweke 1993 show how this is can be equivalent to a t-distribution. I find it easier to think about this in terms of a mixture of normals, but the t-distribution version could be more efficient.

  - To show that diagonal heteroskedasticity is equivalent to a t-distribution, , marginalize out the heteroskedastic component of the variance. For instance, consider

$$y_i \sim N\left(\mu, \tau_y^{-1}\psi_i^{-1}\right)$$
$$\psi_i \sim Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\right)$$

    where all parameters except ψ_i and y_i are assumed known. Then:

$$y_i \sim ST\left[y_i; \mu, \tau_y^{-1}, \nu\right]$$

    Proof:

$$\begin{aligned}
T\left[y_i;\mu,\tau_y^{-1},\nu\right] \propto & \int_{\psi_i} N\left[y_i;\mu,\tau_y^{-1}\psi_i^{-1}\right] Gamma\left[\psi_i;\frac{\nu}{2},\frac{\nu}{2}\right] d\psi_i \\
=& C_1 \int_{\psi_i} \tau_y^{1/2}\psi_i^{1/2}\phi\left(\tau_y\psi_i\left(y_i-\mu\right)\right)\psi_i^{\frac{\nu-2}{2}} \exp\left[-\frac{\nu}{2}\psi_i\right] d\psi_i \\
=& C_1 \int_{\psi_i} \psi_i^{\frac{\nu-1}{2}} \exp\left[-\psi_i\left(\frac{\tau_y\left(y_i-\mu\right)^2}{2}+\frac{\nu}{2}\right)\right] d\psi_i \\
=& C_2 \int_{\psi_i} Gamma\left[\psi_i;\frac{\nu+1}{2},\frac{\tau_y\left(y_i-\mu\right)^2}{2}+\frac{\nu}{2}\right] d\psi_i \\
=& \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)}\tau_y^{1/2}\Gamma\left(\frac{\nu+1}{2}\right)\left(\frac{\tau_y\left(y_i-\mu\right)^2}{2}+\frac{\nu}{2}\right)^{\left(-\frac{\nu+1}{2}\right)} \\
=& \frac{\left(\frac{2}{\nu}\right)^{-\frac{\nu+1}{2}}}{\left(\frac{\nu}{2}\right)^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\tau_y^{1/2}\Gamma\left(\frac{\nu+1}{2}\right)\left(\frac{\tau_y\left(y_i-\mu\right)^2}{2}+\frac{\nu}{2}\right)^{\left(-\frac{\nu+1}{2}\right)} \\
=& \frac{\Gamma\left(\frac{\nu+1}{2}\right)\tau_y^{1/2}}{\left(\frac{\nu}{2}\right)^{1/2}\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\tau_y^{1/2}\left(y_i-\mu\right)^2}{\nu}+1\right)^{\left(-\frac{\nu+1}{2}\right)} \quad \checkmark \\
s.t.& \\
C_1 \equiv & \frac{\left(\frac{\nu}{2}\right)^{\nu/2}}{\Gamma\left(\frac{\nu}{2}\right)}\tau_y^{1/2} \\
C_2 \equiv & C_1 \times \Gamma\left(\frac{\nu+1}{2}\right)\left(\tau_y^{1/2}\frac{\left(y_i-\mu\right)^2}{2}+\frac{\nu}{2}\right)^{\left(-\frac{\nu+1}{2}\right)}
\end{aligned}$$

  - (The expectation and variance of the t-distribution are given by: $E\left[y_i\right] = \mu$ and $var\left(y_i\right) = \frac{\nu}{\tau_y(\nu-2)}\forall\nu > 2$ respectively)

- The accuracy of the VB framework can be an issue. Several approaches could serve to mitigate, but in any case checking the results with MCMC or another method may make sense.

  - Ormerod et al 2011 has accuracy results that are underwhelming, though the t-distribution does better than other models under consideration. Ormerod et al 2014 analyzes additional cases where accuracy suffers.

- Papers use a variety of mitigating techniques to avoid locking onto local optima.
  * See Ray and Szabo 2019, Ormerod et al 2017, and Rockova and George 2014 for methods for avoiding global optima.
  * Another option would be to start optimization points indicated by generalized ridge regressions, and hope this helps avoid local optima.

2.2 Derivation of Mean Field Variational Bayes

- Follow the derivation given in Blei et al 2018
- The KL divergence between an approximating function $q(\Theta)$ and the posterior given observations x denoted as $p(\Theta|x)$ is given by:

$$\begin{aligned}
KL\left(q\left(z\right)||p\left(z|x\right)\right) =& E_{\Theta}\left[\log q\left(\Theta\right)\right] - E_{\Theta}\left[\log p\left(\Theta|x\right)\right] \\
=& E_{\Theta}\left[\log q\left(\Theta\right)\right] - E_{\Theta}\left[\log \frac{p\left(\Theta,x\right)}{p\left(x\right)}\right] \\
=& E_{\Theta}\left[\log q\left(\Theta\right)\right] - E_{\Theta}\left[\log p\left(\Theta,x\right)\right] + \log p\left(x\right)
\end{aligned}$$

Note that $p(x)$ is assumed intractable.

  - The Evidence Lower Bound (ELBO) is defined as

$$\begin{aligned}
ELBO\left(q\right) =& -KL\left(q\left(\Theta\right)||p\left(\Theta|x\right)\right) + \log p\left(x\right) \\
=& E_{\Theta}\left[\log p\left(\Theta,x\right)\right] - E_{z}\left[\log q\left(\Theta\right)\right]
\end{aligned}$$

  Hence it is the entropy difference. Note that the term $\log p(x)$ is constant and thus maximizing the ELBO is the equivilent to minimizing the KL divergence.

    * As described in the name, the ELBO also provides a lower bound on the log evidence. Because the KL divergence is strictly positive,

$$\log p\left(x\right) \geq ELBO\left(q\right)$$

  - The maximizing the ELBO is equivalent to maximizing the expected likelihood and minimizing the variational distance to the prior. To see this, write the ELBO as a KL divergence between the approximating distribution and the prior:

$$\begin{aligned}
ELBO\left(q\right) =& E_{\Theta}\left[\log p\left(x|\Theta\right)\right] + E_{\Theta}\left[\log p\left(\Theta\right)\right] - E_{\Theta}\left[\log q\left(\Theta\right)\right] \\
=& E_{\Theta}\left[\log p\left(x|\Theta\right)\right] - KL\left(q\left(\Theta\right)||p\left(\Theta\right)\right)
\end{aligned}$$

- Mean field variational inference approximates the posterior as:

$$q\left(\Theta\right) = \prod_{j=1}^{M} q_j\left(\Theta_j\right)$$

where Θ_j represents one of the m parameter partitions.

- Coordinate Ascent Variational Inference maximizes the ELBO.
  - The ELBO for approximating function q_j is given by

$$\begin{aligned} ELBO\left(q_j\right) = & E_j\left[E_{-j}\left[\log p\left(\Theta, x\right)\right]\right] - E_j\left[\log q_j\left(\Theta_j\right)\right] + const \\ = & E_j\left[E_{-j}\left[\log p\left(\Theta_j, \Theta_{-j}, x\right)\right]\right] - E_j\left[\log q_j\left(\Theta_j\right)\right] + const \end{aligned}$$

  - This implies that the maximum ELBO is given by

$$\log q_j^*\left(\Theta_j\right) = E_{-j}\left[\log p\left(\Theta_j, \Theta_{-j}, x\right)\right] + const$$

  - The constant term implies that

$$q_j^*\left(\Theta_j\right) \propto \exp E_{-j}\left[\log p\left(\Theta_j, \Theta_{-j}, x\right)\right]$$

  - This suggests that iterative solutions to the above expression should each individually move closer towards a local optimum.

2.3 Unconditional approximate distribution- Additional implementation details

2.4 Implementation

- Without a closed form solution, the implementation consists of iteratively computing the moments until convergence. Local optimality is guaranteed, while global optimality is not.
  - See Table 2 and Table 3 for a summary of the auxiliary variables and moments.

Table 2: Summary of approximate posteriors and dependencies

This table provides the approximate posterior distributions as a function of auxiliary variables and their first-level dependencies. The precise formulas of the auxiliary variables can be found in the main text. Dependencies may include other auxiliary variables and/or moments from elsewhere in the model.

VB Posterior	Aux. Variable	Dimensions	Immediate Dependencies
$q(\phi) \sim MN\left(\phi; \mu_\phi, \Lambda_\phi^{-1}\right)$	Λ_ϕ	$P \times P$	$E\left[\tilde{X}_L' \tilde{X}_L\right]$, $E[\tau_y]$
$q(\phi) \sim MN\left(\phi; \mu_\phi, \Lambda_\phi^{-1}\right)$	μ_ϕ	$P \times 1$	$E\left[\tilde{X}_L' \tilde{y}\right]$, $E[\tau_y]$, Λ_ϕ
$q(x) \sim MN\left(x; \mu_x, \Lambda_x^{-1}\right)$	Λ_x	$T \times T$	$E[\Phi'\Phi]$, $E[\tau_y]$, $E[\tau_x]$, $E[\Psi]$
$q(x) \sim MN\left(x; \mu_x, \Lambda_x^{-1}\right)$	μ_x	$T \times 1$	$E[\Phi]$, μ_β, $E[\tau_y]$, $E[\tau_x]$, $E[\Psi]$, Λ_x
$q(\tau_y) \sim Gamma(\tau_y; \alpha_y, \zeta_y)$	α_y	Scalar	-
$q(\tau_y) \sim Gamma(\tau_y; \alpha_y, \zeta_y)$	ζ_y	Scalar	$E[\tilde{y}'\tilde{y}]$, $E\left[\phi' \tilde{X}_L' \tilde{X}_L \phi\right]$, $E\left[\tilde{y}' \tilde{X}_L\right]$, μ_ϕ, $E[\phi' M_0 \phi]$, $E[x'\Psi x$ $E[(\beta' F' + r')\,\Psi\,(F\beta + r)]$, $E[\Psi]$, μ_β, μ_x, $E[\beta' D A_0 D \beta]$, $E[D]$, $E[\tau_x]$
$q(\tau_x) \sim Gamma(\tau_x; \alpha_x, \zeta_x)$	α_x	Scalar	-
$q(\tau_x) \sim Gamma(\tau_x; \alpha_x, \zeta_x)$	ζ_x	Scalar	$E[x'\Psi x]$, $E[(\beta' F' + r')\,\Psi\,(F\beta + r)]$, $E[\Psi]$, μ_β, μ_x, $E[\beta' D A_0 D \beta]$, $E[D]$, $E[\tau_y]$
$q(g_y) \sim InvGamma\left(g_y; \alpha_y^g, \zeta_y^g\right)$	α_y^g	Scalar	-
$q(g_y) \sim InvGamma\left(g_y; \alpha_y^g, \zeta_y^g\right)$	ζ_y^g	Scalar	μ_ϕ, , $E[\tau_y]$,$E[\phi' M_0 \phi]$
$q(g_x) \sim InvGamma\left(g_x; \alpha_x^g, \zeta_x^g\right)$	α_x^g	Scalar	-
$q(g_x) \sim InvGamma\left(g_x; \alpha_x^g, \zeta_x^g\right)$	ζ_x^g	Scalar	μ_β, $E[\beta' D A_0 D \beta]$, $E[D]$, $E[\tau_y]$
$q(\beta) \sim MN\left(\beta; \mu_\beta, \Lambda_\beta^{-1}\right)$	Λ_β	$K \times K$	$E[D A_0 D]$, $E[\Psi]$, $E[\tau_y]$, $E[\tau_x]$
$q(\beta) \sim MN\left(\beta; \mu_\beta, \Lambda_\beta^{-1}\right)$	μ_β	$K \times 1$	μ_x, $E[D]$, $E[\Psi]$, $E[\tau_y]$, $E[\tau_x]$, Λ_β
$q(\gamma) \sim Bern(\gamma; p_\gamma)$	p_γ	Scalar	$E\left[\beta^2\right]$, $E[\log(1-\omega)]$, $E[\log(\omega)]$, μ_β, Λ_β,$E[D A_0 D]$,$E[\tau$ $E[\tau_x]$
$q(\omega) \sim Beta(\omega; \kappa, \delta)$	κ	Scalar	p_γ
$q(\omega) \sim Beta(\omega; \kappa, \delta)$	δ	Scalar	p_γ
$q(\psi_t) \sim Gamma(\psi_t; \alpha_{\psi t}, \zeta_{\psi t})$	$\alpha_{\psi t}$	Scalar	$E[\nu]$
$q(\psi_t) \sim Gamma(\psi_t; \alpha_{\psi t}, \zeta_{\psi t})$	$\zeta_{\psi t}$	Scalar	$E[\nu]$, $E\left[x_t^2\right]$, μ_{xt}, $E\left[(\beta' f_t + r_t)^2\right]$, μ_β, $E[\tau_y]$, $E[\tau_x]$
$q(\nu) \sim$ Non-standard (η_1, η_2)	η_1	Scalar	$E[\log\psi]$, $E[\psi]$
$q(\nu) \sim$ Non-standard (η_1, η_2)	η_2	Scalar	η_1

Table 3: Summary of approximate posterior moments and dependencies
This table provides the moments of the parameters as a function of auxiliary variables . The precise formulas of the auxiliary variables can be found in the main text. Dependencies may include other auxiliary variables and/or moments from elsewhere in the model. Definitions of quadratic forms and other complex transformations are in the text. Definitions for straight-forward first moments are readily available in terms of the respective distribution parameters.

Moment	Defined in Section	Dimensions	Immediate Dependencies
$E\left[\tilde{X}_L'\tilde{X}_L\right]$	1.2.1	$P \times P$	μ_x, Λ_x
$E\left[\tilde{X}_L'\tilde{y}\right]$	1.2.1	$P \times 1$	μ_x, Λ_x
$E\left[\tau_y\right]$	-	Scalar	α_y, ζ_y
$E\left[\Phi\right]$	-	$S \times T$	μ_ϕ
$E\left[\Phi'\Phi\right]$	1.2.2	$T \times T$	μ_ϕ, Λ_ϕ
$E\left[\tau_x\right]$	-	Scalar	α_x, ζ_x
$E\left[\Psi\right]$	-	$T \times T$ (Diagonal)	α_ψ, ζ_ψ
$E\left[g_y^{-1}\right]$	-	Scalar	α_y^g, ζ_y^g
$E\left[g_x^{-1}\right]$	-	Scalar	α_x^g, ζ_x^g
$E\left[\tilde{y}'\tilde{y}\right]$	1.2.3	Scalar	μ_x, Λ_x
$E\left[\phi'\tilde{X}_L'\tilde{X}_L\phi\right]$	1.2.3	Scalar	$E\left[\tilde{X}_L'\tilde{X}_L\right]$, μ_x, Λ_x
$E\left[\phi' M_0 \phi\right]$	1.2.3	Scalar	μ_ϕ, Λ_ϕ
$E\left[x'\Psi x\right]$	1.2.4	Scalar	μ_x, Λ_x,$E\left[\Psi\right]$
$E\left[\left(\beta' F' + r'\right)\Psi\left(F\beta + r\right)\right]$	1.2.4	Scalar	μ_β, Λ_β,$E\left[\Psi\right]$
$E\left[D\right]$	1.2.4	Scalar	p_γ
$E\left[DA_0D\right]$	1.2.4	$K \times K$	p_γ
$E\left[\beta' DA_0 D\beta\right]$	1.2.4	Scalar	μ_β, Λ_β, $E\left[DA_0D\right]$
$E\left[\beta^2\right]$	1.2.8	Scalar	μ_β, Λ_β
$E\left[\log\left(\omega\right)\right]$	1.2.8	Scalar	κ, δ
$E\left[\log\left(1-\omega\right)\right]$	1.2.8	Scalar	κ, δ
$E\left[x_t^2\right]$	1.2.11	Scalar	μ_x, Λ_x
$E\left[\left(\beta' f_t + r_t\right)^2\right]$	1.2.11	Scalar	μ_β, Λ_β
$E\log\left(\psi_t\right)$	1.2.12	Scalar	α_ψ, ζ_ψ
$E\left[\nu\right]$	1.2.12	Scalar	η_1, η_2

3 Appendix

3.1 Possible Revision

The goal is to squash the first layer of the hierarchy under an assumption of zero reporting variance.

Start by assuming 1:1 periodicity and a single value of y_t:

$$y_t = \left(x_t^L\right)' \phi + \tilde{\phi}_{P+1} x_t$$

where $x_t^L \equiv \left\{x_t L^{P-p+1}\right\} \forall p \in 1 : P$

The likelihood of y is given by:

$$\begin{aligned} \log p\left(y|rest\right) \propto &- \frac{\tau_x}{2}\left(y - \mu^y\right)' \Omega \left(y - \mu^y\right) \\ \mu^y \equiv &\Phi F \beta \end{aligned}$$

Suppose we have the first P shocks denoted as ε^-. Then conditional on the other parameters and no intermediate missing values of y, ε and therefore x are fully determined. Moreover, the likelihood is also determined by examining the likelihood of ε. This seems like a promising track to examine. It suggests the following algorithm:

1. Draw $y^m|rest$, the missing values of y conditional on all other parameters. Note that it may be possible to integrate out ε^-, so that this draw is from a standard conditional multivariate normal.

2. Draw from $\varepsilon^-|rest$ where the likelihood is based on the likelihood of the implied forward shocks.

Lots of details to work out.

3.2 Derivation of Expectations

- $E\left[\log X\right]$ s.t. $X \sim Beta\left(\eta, \delta\right)$:

$$\begin{aligned} E\left[\log X\right] =& \frac{1}{B\left(\eta,\delta\right)} \int_0^1 x^{\eta-1}\left(1-x\right)^{\delta-1} \log x dx \\ =& \frac{\partial}{\partial \eta} \int_0^1 x^{\eta-1}\left(1-x\right)^{\delta-1} dx \\ \text{(since } & \frac{\partial}{\partial\eta} x^{\eta} = x^{\eta} \frac{\partial}{\partial \eta} \log x^{\eta} = x^{\eta} \log x \text{)} \\ =& \frac{1}{B\left(\eta,\delta\right)} \frac{\partial}{\partial\eta} B\left(\eta,\delta\right) \\ =& \frac{1}{B\left(\eta,\delta\right)} \frac{\partial}{\partial\eta} \frac{\Gamma\left(\eta\right)\Gamma\left(\delta\right)}{\Gamma\left(\eta+\delta\right)} \\ =& \frac{\Gamma\left(\delta\right)}{B\left(\eta,\delta\right)} \left(\frac{\Gamma'\left(\eta\right)\Gamma\left(\eta+\delta\right) - \Gamma\left(\eta\right)\Gamma'\left(\eta+\delta\right)}{\Gamma^2\left(\eta+\delta\right)}\right) \\ =& \frac{}{B\left(\eta,\delta\right)} \left(B\left(\eta,\delta\right)\psi\left(\eta\right) - B\left(\eta,\delta\right)\psi\left(\eta+\delta\right)\right) \\ =& \digamma\left(\eta\right) - \digamma\left(\eta+\delta\right) \end{aligned}$$

where $\digamma\left(\cdot\right)$ is the digamma distribution.

- $E\left[\log\left(1-X\right)\right]$ s.t. $X \sim Beta\left(\eta,\delta\right)$:

$$\begin{aligned}
E\left[\log\left(1-X\right)\right] =& \frac{1}{B\left(\eta,\delta\right)}\int_{0}^{1}x^{\eta-1}\left(1-x\right)^{\delta-1}\log\left(1-x\right)dx\\
=& \frac{1}{B\left(\delta,\eta\right)}\int_{0}^{1}\left(1-u\right)^{\eta-1}u^{\delta-1}\log\left(u\right)dx\\
=& \digamma\left(\delta\right)-\digamma\left(\eta+\delta\right)\\
s.t.&\\
u \equiv& 1-x
\end{aligned}$$

3.3 Reconciliation of $q\left(\nu\right)$ with Wand et al 2011

- The derivation of $q\left(\nu\right)$ should be equivalent to the solution given in Wand et al 2011 (WOPF) after accounting for the following differences:
  - WOPF elect to use a uniform prior of $Unif\left(\nu_{min},\nu_{max}\right)$ instead of the gamma prior. A gamma prior seems more consistent with the rest of the model and has an advantage in not needing to truncate the support. However, for reconciliation purposes, this derivation will walk through the previous derivation using the Wand prior
  - WOPF parameterizes the variance instead of the precision (ψ_t^{-1} instead of ψ_t), and hence the prior and posterior variables at each point in time are inverse gamma distributions instead of gamma distributions. This difference is innocuous.
- Reconciling with WOPF equation 14:

$$\begin{aligned}
\log\tilde{q}\left(\nu\right) =& E_{-\nu}\left[\frac{T\nu}{2}\log\left(\frac{\nu}{2}\right)-T\log\Gamma\left(\frac{\nu}{2}\right)+\sum_{t}\left(\frac{\nu}{2}\log\psi_t-\frac{\nu\psi_t}{2}\right)\right]+\tilde{c}_1^{\nu}\\
=& \frac{T\nu}{2}\log\left(\frac{\nu}{2}\right)-T\log\Gamma\left(\frac{\nu}{2}\right)+\frac{\nu}{2}\tilde{\eta}_1\\
q\left(\nu\right) =& \exp\left(\frac{T\nu}{2}\log\left(\frac{\nu}{2}\right)-T\log\Gamma\left(\frac{\nu}{2}\right)+\frac{\nu}{2}\tilde{\eta}_1\right)\tilde{\eta}_2\\
s.t.&\\
\tilde{c}_1^{\nu} =& c_1^{\nu}-\sum_{t}\log\psi_t-\alpha_{\nu 0}\log\zeta_{\nu 0}+\log\Gamma\left(\alpha_{\nu 0}\right)+\frac{1}{\nu_{\max}-\nu_{\min}}\\
\tilde{\eta}_1 =& E\left[\sum_{t}\left(\log\psi_t-\psi_t\right)\right]\\
\tilde{\eta}_2 =& \left(\int_{\nu_{\min}}^{\nu_{\max}}\exp\left(\frac{T\nu}{2}\log\left(\frac{\nu}{2}\right)-T\log\Gamma\left(\frac{\nu}{2}\right)+\frac{\nu}{2}\tilde{\eta}_1\right)d\nu\right)^{-1}\\
=& F\left(0,\nu,-\tilde{\eta}_1,\nu_{\min},\nu_{\max}\right)
\end{aligned}$$

- Plugging in the property that ψ_t^{-1} equals the a_t from the Wand notation:

$$\begin{aligned}
-\eta_1 &= \sum_t \left(-E\left[\log \psi_t\right] + E\left[\psi_t\right]\right) \\
&= \sum_t \left(E\left[\log a_t\right] + E_t\left[\frac{1}{a_t}\right]\right) \\
&= C_1
\end{aligned}$$

- Thus using a uniform prior and accounting for the inverse gamma distribution leads to the same answer as WOPF.

3.4 Derivation of the precision a_{0k} given the expectation and variance of β and $p_{\gamma k}$

This is helpful for deriving sequential priors. First, consider the simpler scenario where $\beta_0^{\Delta} = 0$ and β_0 is matched to the expectation. This does not fully reflect the DGP, but it is a convenient and simple way to generate the prior:

$$\begin{aligned}
\beta_0 =& E\left[\beta_k\right] \\
Var\left(\beta_k\right) =& E\left[Var\left(\beta_k|\gamma_k\right)\right] + Var\left(E\left[\beta_k|\gamma_k\right]\right) \\
=& Var\left(\beta_k|\gamma_k = 1\right) p_k^{\gamma} + Var\left(\beta_k|\gamma_k = 0\right)\left(1 - p_k^{\gamma}\right) + Var\left(E\left[\beta_k|\gamma_k\right]\right) \\
=& \frac{p_k^{\gamma}}{a_{0k}} + \frac{\left(1 - p_k^{\gamma}\right) v^2}{a_{0k}}
\end{aligned}$$

Next, consider the more complex scenario where β_0 is given, and derive β_0^{Δ}

$$\begin{aligned}
E\left[\beta\right] &= \beta_0 + p_k^{\gamma}\beta_0^{\Delta} + v\beta_0^{\Delta}\left(1 - p_k^{\gamma}\right) \\
\implies \beta_0^{\Delta} &= \frac{E\left[\beta\right] - \beta_0}{p_k^{\gamma} + v * \left(1 - p_k^{\gamma}\right)}
\end{aligned}$$

Then apply the law of total variance to compute the implied a_{0k}:

$$\begin{aligned}
Var\left(\beta_k\right) =& E\left[Var\left(\beta_k|\gamma_k\right)\right] + Var\left(E\left[\beta_k|\gamma_k\right]\right) \\
=& Var\left(\beta_k|\gamma_k = 1\right) p_k^{\gamma} + Var\left(\beta_k|\gamma_k = 0\right)\left(1 - p_k^{\gamma}\right) + Var\left(E\left[\beta_k|\gamma_k\right]\right) \\
=& \frac{p_k^{\gamma}}{a_{0k}} + \frac{\left(1 - p_k^{\gamma}\right) v^2}{a_{0k}} + p_k^{\gamma}\left(\beta_0 + \beta_0^{\Delta}\right)^2 + \left(1 - p_k^{\gamma}\right)\left(\beta_0 + v\beta_0^{\Delta}\right)^2 - E\left[\beta_k\right]^2 \\
a_{0k} =& \frac{p_k^{\gamma} + \left(1 - p_k^{\gamma}\right) v^2}{Var\left(\beta_k\right) - p_k^{\gamma}\left(\beta_0 + \beta_0^{\Delta}\right)^2 - \left(1 - p_k^{\gamma}\right)\left(\beta_0 + v\beta_0^{\Delta}\right)^2 + E\left[\beta_k\right]^2}
\end{aligned}$$

3.5 Approximate moving average equivalence

- While the employment of measurement error may seem odd given that returns are generally considered factual, it is without loss of generality with respect to modeling a moving average process.

- Consider the sequential dynamics of y (if $\Delta t > 0$, model some values of y as unobserved).

$$y_t = \left(\Delta t - \sum_{p=1}^{P} \phi_p\right) x_t + \sum_{p=1}^{P} \phi_p L^{P-p} x_{t-1} + \epsilon_t$$

- To start, rescale the series towards the cannonical MA representation. Difference the new series:

$$z_t \equiv (y_t - \mu)\left(\Delta t - \sum_{p=1}^{P} \phi_p\right)^{-1}$$

$$z_t = \left(1 + \sum_{p=1}^{P} \varphi_{P-p+1} L^p\right) x_t + \varepsilon_t$$

- Assume the state variable x_t has an unconditional variance of σ_x^2. Then autocovariance is characterized as:

$$\gamma_0 = \left(1 + \sum_{p=1}^{P} \varphi_{P-p+1}^2\right) \sigma_x^2 + \sigma_\varepsilon^2$$

$$\gamma_1 = \sigma_x^2 \left(\varphi_P + \sum_{p=2}^{P} \varphi_{P-p+1}\varphi_{P-p+2}\right)$$

$$\gamma_2 = \sigma_x^2 \left(\varphi_{P-1} + \sum_{p=3}^{P} \varphi_{P-p+1}\varphi_{P-p+3}\right)$$

$$\gamma_s = \begin{cases} \left(1 + \sum_{p=1}^{P} \varphi_{P-p+1}^2\right) \sigma_x^2 + \sigma_\varepsilon^2 & s = 0 \\ \sigma_x^2 \left(\varphi_{P-s+1} + \sum_{p=s+1}^{P} \varphi_{P-p+1}\varphi_{P-p+1+s}\right) & 0 < s \leq P-1 \\ \sigma_x^2 \varphi_1 & s = P \\ 0 & P < s \end{cases}$$

- The Wold decomposition theorem states that for any covariance stationary process, the following form is equivalent up to the second moment:

$$z_t = \sum_{j=0}^{\infty} a_j \eta_{t-j}$$

As the autocovariance function is unique and terminates, the Wold representation is characterized as

$$z_t = \eta_t + \sum_{p=1}^{P} a_1 \eta_{t-p}$$

which is a moving average process.

– Matching the terms:

$$\sigma_\eta^2\left(1+\sum_{p=1}^{P}a_p^2\right) = \left(1+\sum_{p=1}^{P}\varphi_{P-p+1}^2\right)\sigma_x^2+\sigma_\varepsilon^2$$

$$\sigma_\eta^2\left(a_s+\sum_{p=s+1}^{P}a_pa_{p-s}\right) =\sigma_x^2\left[\left(\varphi_{P-s+1}+\sum_{p=s+1}^{P}\varphi_{P-p+1}\varphi_{P-p+1+s}\right)\right]$$

$$a_P\sigma_\eta^2 =\sigma_x^2\varphi_1$$

Future work could recover the Wold shocks (at least as stochastic variables), which would improve the MA fit. However, such recovery is of little practical benefit as the measurement error component is unforecastable and iid.

– The assumption that second moment exists is highly plausible but not without loss of generality. As the errors of x are t distributed, a necessary but not sufficient condition is $\nu > 2$.

3.6 Standard Proof of Metropolis-Hastings

Notation: Let $p(\theta|y)$ be the posterior pdf. The goal is to show $p(\theta|y)$ is the unique stationary distribution of the Markov Chain. Follow BDA 3rd pg 279, generalized for Metropolis-Hastings and with additional added emphasis on the detailed balance condition.

- Step 1: Show that the simulated sequence is a Markov Chain with a unique stationary distribution
  - Most of this step is true by assumption.
  - This is always true if the Markov chain is irreducible, aperiodic, and not transient.
    - Irreducible: This means any state is accessible from any other state. Irreducibility is a standard limitation of these algorithms that usually holds.
    - Aperiodic: This means that there is never a regular pattern such that for $i, 2i, \dots$ the transition probability is zero, and non-zero for all other iterations. This should always hold.
    - Transient: There are no states with finite probability that shift to zero probability. This should always hold.
  - Hence if $p(\theta|y)$ is a stationary distribution of the Markov Chain, it is the unique distribution and the proof is complete.
- Step 2: Show that $p(\theta|y)$ is a stationary distribution for the Markov Chain.
  - Let θ_a, θ_b represent any potential parameter draws in the support of p. Let $q(\theta_b|\theta_a)$ be any proposal distribution density with the same support as p[1]Following the Hastings extension of the Metropolis algorithm, do not assume that the distribution is symmetric. This is particularly important for the Independent Metropolis Hastings algorithm used in the model.

[1]

∗ The proposal distribution q must satisfy additional technical conditions, including a fat tail constraint. These are beyond the scope of this proof outline.

- Let T be the transition distribution as defined by the Metropolis-Hastings algorithm. In other words, T is the process for generating the Markov Chain. By definition, this is:

$$
\begin{aligned}
T\left(\theta_b|\theta_a\right) =& q\left(\theta_b|\theta_a\right)\min\left(1, \frac{p\left(\theta_b|y\right)}{p\left(\theta_a|y\right)} \times \frac{q\left(\theta_a|\theta_b\right)}{q\left(\theta_b|\theta_a\right)}\right) \\
T\left(\theta_a|\theta_b\right) =& q\left(\theta_a|\theta_b\right)\min\left(1, \frac{p\left(\theta_a|y\right)}{p\left(\theta_b|y\right)} \times \frac{q\left(\theta_b|\theta_a\right)}{q\left(\theta_a|\theta_b\right)}\right)
\end{aligned}
$$

- Assume the Markov Chain has detailed balance, that is, the probability of drawing θ_a from the stationary distribution and transitioning to θ_b is the same as the reverse. Then the balance assumption implies that $\pi\left(\theta|y\right)$ is the unique stationary distribution:

$$
\begin{aligned}
\pi\left(\theta_a|y\right) T\left(\theta_a|\theta_b\right) =& T\left(\theta_b|\theta_a\right)\pi\left(\theta_b|y\right) \\
\implies \frac{\pi\left(\theta_b|y\right)}{\pi\left(\theta_a|y\right)} =& \frac{T\left(\theta_b|\theta_a\right)}{T\left(\theta_a|\theta_b\right)} \\
=& \frac{q\left(\theta_b|\theta_a\right)}{q\left(\theta_a|\theta_b\right)} \frac{\min\left(1, \frac{p(\theta_b|y)}{p(\theta_a|y)} \times \frac{q(\theta_a|\theta_b)}{q(\theta_b|\theta_a)}\right)}{\min\left(1, \frac{p(\theta_a|y)}{p(\theta_b|y)} \times \frac{q(\theta_b|\theta_a)}{q(\theta_a|\theta_b)}\right)}
\end{aligned}
$$

- Suppose $\frac{p(\theta_b|y)}{p(\theta_a|y)} \times \frac{q(\theta_a|\theta_b)}{q(\theta_b|\theta_a)} \geq 1$ (the opposite condition follows identical logic). Then:

$$
\begin{aligned}
\frac{\pi\left(\theta_b|y\right)}{\pi\left(\theta_a|y\right)} =& \frac{q\left(\theta_b|\theta_a\right)}{q\left(\theta_a|\theta_b\right)} \frac{1}{\frac{p(\theta_a|y)}{p(\theta_b|y)} \times \frac{q(\theta_b|\theta_a)}{q(\theta_a|\theta_b)}} \\
=& \frac{p\left(\theta_b|y\right)}{p\left(\theta_a|y\right)}
\end{aligned}
$$

- This concludes the proof. Because the ratios are equal for all θ_a and θ_b, $\pi = p$, otherwise one of the distributions wouldn't be valid. Combined with step 1, the results show p is the unique stationary distribution.

• Special cases

- For Metropolis (but not Metropolis-Hastings), the distribution is assumed to by symmetric. Then:

$$
T\left(\theta_b|\theta_a\right) = q\left(\theta_b|\theta_a\right)\min\left(1, \frac{p\left(\theta_b|y\right)}{p\left(\theta_a|y\right)}\right)
$$

- For the Independent Metropolis Hastings (IMH) algorithm used in the model, q is an independent distribution with no dependency on the previous parameter draw, such that $q\left(\theta_b|\theta_a\right) = q\left(\theta_b\right)$. Then:

$$
T\left(\theta_b|\theta_a\right) = q\left(\theta_b\right)\min\left(1, \frac{p\left(\theta_b|y\right)}{p\left(\theta_a|y\right)} \times \frac{q\left(\theta_a\right)}{q\left(\theta_b\right)}\right)
$$

- For a conjugate prior where the conditional distribution is the proposal distribution, define the proposal distribution for parameter k as follows:

$$
q\left(\theta_b|\Theta^{-k}, \theta_a^k, y\right) \equiv p\left(\theta_b^k|\Theta^{-k}, \theta_a^k, y\right)
$$

Then:

$$
\begin{aligned}
T\left(\theta_b^k|\theta_a^k\right) =& p\left(\theta_b^k|\Theta^{-k},\theta_a^k,y\right)\min\left(1,\frac{p\left(\theta_b^k|y\right)}{p\left(\theta_a^k|y\right)}\times\frac{p\left(\theta_a^k|\Theta^{-k},\theta_b^k,y\right)}{p\left(\theta_b^k|\Theta^{-k},\theta_a^k,y\right)}\right)\\
=& p\left(\theta_b^k|\Theta^{-k},\theta_a^k,y\right)\min\left(1,\frac{p\left(\theta_a,\theta_b|\Theta^{-k},y\right)}{p\left(\theta_a,\theta_b|\Theta^{-k},y\right)}\right)\\
=& p\left(\theta_b^k|\Theta^{-k},\theta_a^k,y\right)
\end{aligned}
$$

which implies the proposal distribution is always accepted. Simlarly, allowing q to depend on Θ^{-k} does not change any of the other discussed proofs.