II. Vector autoregressions
A. Introduction
  1. VARs as forecasting models

Suppose we want to forecast $y_{1t}$ based on:

$y_{1,t-1}, y_{1,t-2}, \ldots, y_{1,t-p}$

$y_{2,t-1}, y_{2,t-2}, \ldots, y_{2,t-p}$

$\vdots$

$y_{n,t-1}, y_{n,t-2}, \ldots, y_{n,t-p}$

deterministic functions of $t$

$(1, t, \cos(\pi t/6), \text{seasonal dummies})$

Let $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$

$x_t = (1, y_{t-1}', y_{t-2}', \ldots, y_{t-p}')'$

$k = np + 1$
Suppose we consider linear forecast
\[ \hat{y}_{1|t-1} = \gamma_1' x_t \]
Best forecast within linear class:
value of \( \gamma_1 \) that minimizes
\[ E(y_{1t} - \gamma_1' x_t)^2 \]

Proposition: If \( y_t \) is covariance-stationary
and \( E(x_t x_t') \) is nonsingular, then optimal forecast uses
\[ \gamma_1^* = E(x_t x_t')^{-1} E(x_t y_t) \]

Definition: The optimal linear forecast,
\[ \hat{y}_{1|t-1} = \gamma_1^* x_t, \]
is called the “population linear projection” of \( y_{1t} \) on \( x_t \).
Proposition: If $y_t$ is stationary and ergodic, then

$$\hat{\gamma}_1^p \rightarrow \gamma_1$$

Proof: (Law of Large Numbers)

$$\hat{\gamma}_1 = \left( T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( T^{-1} \sum_{t=1}^{T} x_t y_{1t} \right)$$

$$\overset{p}{\underset{}{\rightarrow}} E(x_t x_t')^{-1} E(x_t y_{1t})$$

If form separate forecasting equation for each element of $y_t$ and collect in vector,

$$y_{1t} = \gamma_1' x_t + \epsilon_{1t}$$

$$\vdots$$

$$y_{nt} = \gamma_n' x_t + \epsilon_{nt}$$

$$y_t = \Gamma' x_t + \epsilon_t$$

result is called vector autoregression:

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$
Above results imply we can consistently estimate coefficients for VAR by OLS equation by equation

\[
\hat{\gamma}_1' = \left( \sum_{t=1}^{T} y_{1t}x_{1t}' \right) \left( \sum_{t=1}^{T} x_{1t}x_{1t}' \right)^{-1} \\
\vdots \\
\hat{\gamma}_n' = \left( \sum_{t=1}^{T} y_{nt}x_{nt}' \right) \left( \sum_{t=1}^{T} x_{nt}x_{nt}' \right)^{-1} \\
\hat{\Gamma}' = \left[ \hat{c}, \hat{\Phi}_1 \hat{\Phi}_2 \cdots \hat{\Phi}_p \right]
\]

II. Vector autoregressions

A. Introduction
   1. VARs as forecasting models
   2. Gaussian VARs as data-generating process
Consider the following process whereby the \((n \times 1)\) vector \(y_t\) might have been generated:

\[
y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \varepsilon_t
\]

\(\varepsilon_t \sim N(0, \Omega)\)


log likelihood:

\[
\mathcal{L} = \log p(y_1, y_2, \ldots, y_T | y_0, y_1, \ldots, y_{p+1}, \theta) \\
= \sum_{t=1}^{T} \log p(y_t | y_{t-1}, \ldots, y_{t-p}, \theta) \\
= -(Tn/2) \log(2\pi) - (T/2) \log|\Omega| \\
- (1/2) \sum_{t=1}^{T} \varepsilon_t \Omega^{-1} \varepsilon_t
\]

\(\varepsilon_t = y_t - c - \Phi_1 y_{t-1} - \Phi_2 y_{t-2} - \cdots - \Phi_p y_{t-p}\)

\(\theta = \) vector containing elements of \(c, \Phi_1, \Phi_2, \ldots, \Phi_p, \Omega\)

Classical results for VARs:
(1) The MLE of \(\Gamma\) is OLS equation by equation:

\[
\hat{\Gamma}' (n \times k) = \left( \sum_{t=1}^{T} y_t x_t' \right) \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1}
\]
(2) The MLE of $\Omega$ is average product of residuals:

$$\hat{\Omega} = T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

$$\hat{\varepsilon}_t = y_t - \hat{\Gamma}' x_t$$

(3) The asymptotic distribution of

$$\hat{\gamma} = \text{vec}(\hat{\Gamma}) = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_n)'$$

is given by

$$\sqrt{T} (\hat{\gamma} - \gamma) \overset{d}{\rightarrow} N(0, \Omega \otimes M)$$

$$M = \text{plim} \left( T^{-1} \sum_{t=1}^{T} x_t x_t' \right)^{-1}$$

$$\Omega \otimes M = \begin{bmatrix} \sigma_{11} M & \cdots & \sigma_{1n} M \\ \vdots & \ddots & \vdots \\ \sigma_{n1} M & \cdots & \sigma_{nn} M \end{bmatrix}$$

II. Vector autoregressions

A. Introduction
   1. VARs as forecasting models
   2. Gaussian VARs as data-generating process
   3. VARs as ad hoc dynamic structural models
Example:

\[ f_t = \text{federal funds rate} \]
\[ y_t = \text{output growth} \]
\[ \pi_t = \text{inflation} \]
\[ m_t = \text{money growth rate} \]

Represent Fed behavior by

\[ f_t = \alpha_0 + \alpha_1 y_t + \alpha_2 \pi_t + \alpha_3 f_{t-1} + \alpha_4 y_{t-1} + \alpha_5 \pi_{t-1} + \alpha_6 m_{t-1} + v_t \]

Fed responds to current output and inflation but not current money growth

\[ y_t = (f_t, y_t, \pi_t, m_t) \]
\[ B_0 y_t = k + B_1 y_{t-1} + v_t \]

\[ B_0 = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_2 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \]

\[ B_1 = \begin{bmatrix} \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \]
If \( \mathbf{v}_i \sim \text{i.i.d. } \mathcal{N}(\mathbf{0}, \mathbf{D}) \), log likelihood is
\[
\mathcal{L} = \log f(\mathbf{y}_1: \mathbf{y}_2: \ldots: \mathbf{y}_T | \mathbf{y}_0, \mathbf{y}_{-1}, \ldots, \mathbf{y}_{-p+1}, \theta) \\
= -(Tn/2) \log(2\pi) - (T/2) \log |\mathbf{D}| \\
+ T \log |\mathbf{B}_0| - (1/2) \sum_{i=1}^{T} \mathbf{v}_i \mathbf{D}^{-1} \mathbf{v}_i, \\
\mathbf{D} = E(\mathbf{v}_i \mathbf{v}_i^T) \\
\mathbf{v}_i = \mathbf{B}_0 \mathbf{y}_i - \mathbf{k} - \mathbf{B}_1 \mathbf{y}_{i-1} - \mathbf{B}_2 \mathbf{y}_{i-2} - \\
\ldots - \mathbf{B}_p \mathbf{y}_{i-p} \\
\theta = \text{vector containing elements of } \mathbf{k}, \mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p, \mathbf{D}
\]

If model is just-identified, the MLE’s
\( \hat{\mathbf{k}}, \hat{\mathbf{B}}_0, \hat{\mathbf{B}}_1, \ldots, \hat{\mathbf{B}}_p, \hat{\mathbf{D}} \)
are transformations of the VAR MLE’s
\( \hat{\mathbf{c}}, \hat{\Phi}_1, \hat{\Phi}_2, \ldots, \hat{\Phi}_p, \hat{\Omega} \)

II. Vector autoregressions

A. Introduction
B. Normal-Wishart priors for VARs
For univariate regression,
\[ y_t = \mathbf{x}_t'\mathbf{\beta} + \epsilon_t \]
\[ \epsilon_t \sim N(0, \sigma^2) \]
we used Normal-Gamma for natural conjugate prior:
\[ \mathbf{\beta}|\sigma^2 \sim N(m, \sigma^2\mathbf{M}) \]
\[ \sigma^{-2} \sim \Gamma(N, \lambda) \]

From asymptotic distribution of MLE,
\[ \sqrt{T}(\hat{\mathbf{\gamma}} - \mathbf{\gamma}) \xrightarrow{L} N(0, \mathbf{\Omega} \otimes \mathbf{M}) \]
\[ \mathbf{M} = \text{plim} \left( T^{-1} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1}, \]
we might guess that natural conjugate prior is of the form
\[ \mathbf{\gamma}|\mathbf{\Omega} \sim N(m, \mathbf{\Omega} \otimes \mathbf{M}) \]
\[ m = \text{prior guess for } \mathbf{\gamma} \]
\[ \mathbf{M} \text{ summarizes confidence} \]

Prior for \( \mathbf{\Omega} \):
univariate regression
\[ \sigma^2 = E(\epsilon_t^2) \]
\[ Z_i \sim N(0, \lambda^{-1}) \]
\[ W = (Z_1^2 + Z_2^2 + \cdots + Z_N^2) \]
\[ W \text{ has gamma distribution} \]
with parameters \( N, \lambda \)
\[ \sigma^{-2} \sim \Gamma(N, \lambda) \]
Vector autoregression:
\[ \Omega = E(\epsilon, \epsilon_i) \]
\[ z_t \sim N(0, \Lambda^{-1}) \]
\[ w = (z_t z_t' + z_{t+1} z_{t+1}' + \cdots + z_N z_N') \]
\[ w \] has Wishart distribution
with parameters \( N, \Lambda \)
\[ \Omega^{-1} \sim W(N, \Lambda) \]

\[ w \sim W(N, \Lambda) \Rightarrow \]
\[ p(w) = c |\Lambda|^{N/2} |w|^{(N-n-1)/2} \exp \left[ -\frac{1}{2} \text{tr}(w\Lambda) \right] \]
\[ c = \left[ 2^{Nn^2} \pi^{n(n-1)/4} \right] \prod_{j=1}^n \Gamma \left( \frac{N+1-j}{2} \right) \]
so prior takes the form
\[ p(\Omega^{-1}) \propto |\Omega|^{-(N-n-1)/2} \exp \left[ -\frac{1}{2} \text{tr}(\Omega^{-1}\Lambda) \right] \]

\[ y_t = \Gamma' x_t + \epsilon_t \]
\[ \gamma = \text{vec}(\Gamma) \]
first \( k \) components of \( \gamma \) = coefficients to explain \( y_{1t} \)
Prior for $\gamma$
univariate regression
$\beta|\sigma^2 \sim N(m, \sigma^2 \mathbf{M})$
vector autoregression

$\gamma|\Omega^{-1} \sim N\left(\begin{array}{c} \mathbf{m} \\ \Omega \otimes \mathbf{M} \end{array} \right)_{nk \times 1}$
$p(\gamma|\Omega^{-1}) \propto (2\pi)^{-nk/2}|\Omega^{-k/2}|$
$\exp\left[-\frac{1}{2} (\gamma - \mathbf{m})' (\Omega \otimes \mathbf{M})^{-1} (\gamma - \mathbf{m}) \right]$

After a lot of algebra, this can be rewritten as

$p(\gamma, \Omega^{-1}|\mathbf{Y}) \propto p(\gamma, \Omega^{-1}, \mathbf{Y})$
$= p(\mathbf{Y}|\gamma, \Omega^{-1})p(\gamma|\Omega^{-1})p(\Omega^{-1})$
$\propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2} \sum (y_t - \Gamma x_t)' \Omega^{-1} (y_t - \Gamma x_t) \right]$
$\Omega^{-k/2} \exp\left[-\frac{1}{2} (\gamma - \mathbf{m})' (\Omega \otimes \mathbf{M})^{-1} (\gamma - \mathbf{m}) \right]$
$\Omega^{-(N-n-1)/2} \exp\left[-\frac{1}{2} \text{tr}(\Omega^{-1} \Lambda^*) \right]$
i.e., $\gamma|\Omega^{-1}, \mathbf{Y} \sim N(\mathbf{m}^*, \Omega \otimes \mathbf{M}^*)$
$\Omega^{-1}|\mathbf{Y} \sim W(T+N, \Lambda^*)$
\[ y | \Omega^{-1}, Y \sim N(m^*, \Omega \otimes M^*) \]
\[ M^* = \left( M^{-1} + \sum_{t=1}^{T} x_t x_t' \right)^{-1} \]
\[ m^* = (I_n \otimes M^* M^{-1}) m + \left( I_n \otimes M^* \sum_{t=1}^{T} x_t x_t' \right) \hat{\gamma} \]
\[ \hat{\gamma} = \text{vec}(\hat{\Gamma}) \]
\[ \hat{\Gamma} = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( \sum_{t=1}^{T} x_t y_t' \right) \]
Diffuse prior: \( M^{-1} = 0 \)

\[ \gamma | \Omega^{-1}, Y \sim N \left( \frac{\hat{\gamma}}{\Omega} \otimes \left[ \sum_{t=1}^{T} x_t x_t' \right]^{-1} \right) \]
Estimate \( i \)th equation of VAR by OLS
\[ \hat{\gamma}_i = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( \sum_{t=1}^{T} x_t y_t' \right) \]
(corresponds to elements \( k(i-1) + 1 \) through \( ki \) of \( \hat{\gamma} \))
\[ \hat{\gamma}_i \] is posterior mean of \( \gamma_i \) and
\[ \sigma_i \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \] is posterior variance (conditional on \( \Omega \))

\[ \Omega^{-1} | Y \sim W(T + N, \Lambda^*) \]
\[ \Lambda^* = \Lambda + \hat{S} + Q \]
\[ \hat{S} = \sum_{t=1}^{T} \left( y_t - \hat{\Gamma} x_t \right) \left( y_t - \hat{\Gamma} x_t \right)' \]
\[ Q = V \left( \sum_{t=1}^{T} x_t x_t' \right) M^* M^{-1} V \]
\[ \text{vec} \left( \begin{bmatrix} V \\ \mathbf{k} \end{bmatrix} \right) = m - \hat{\gamma} \]
Diffuse prior: \( N = 0, M^{-1} = 0, \Lambda = 0 \)
Diffuse prior:
\[ \Omega^{-1}|Y \sim W(T; \hat{\mathbf{S}}) \]
\[ \gamma \Omega^{-1}, Y \sim N \left( \hat{\gamma}, \Omega \otimes \left[ \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right]^{-1} \right) \]

To generate a draw from the posterior distribution of \((\gamma, \Omega^{-1})\) with a diffuse prior:

1. Estimate \(i\)th equation of VAR by OLS for \(i = 1, 2, \ldots, n\)
   \[
   \hat{\gamma}_i = \left( \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \left( \sum_{t=1}^{T} \mathbf{x}_t y_{it} \right) 
   
   M^* = \left( \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1}
   
2. Calculate residual of \(i\)th equation and sum of outer products of residuals:
   \[
   \hat{\varepsilon}_{it} = y_{it} - \hat{\gamma}_i \mathbf{x}_i
   
   \hat{\mathbf{S}} = \sum_{t=1}^{T} \hat{\varepsilon}_i \hat{\varepsilon}_i'
   
   \]
(3) Generate an artificial sample of size $T$ of $(n \times 1)$ vector $z_t$ where $z_t \sim N(0, \hat{S}^{-1})$

and calculate the $(n \times n)$ matrix $w = \sum_{i=1}^{T} z_t z_t'$

(4) Set $\Omega = \mathbf{W}^{-1}$.

(5) Generate a draw for $\gamma$ from a $N(\tilde{\gamma}, \Omega \otimes \mathbf{M}^*)$ distribution.

The values of $\Omega$ from step (4) and $\gamma$ from step (5) represent a single draw from the posterior distribution. To generate a Monte Carlo sample of $D$ draws, repeat steps (3)-(5) $D$ times.

To get standard errors on impulse-response functions, for each draw calculate the impulse-response function implied by that value of $\gamma$. 
Usual procedure: $\Omega$ is fixed at $T^{-1}\hat{S}$ for all draws = asymptotic classical distribution or approximation to Bayesian distribution with diffuse prior.

Example of using a non-diffuse prior (Del Negro and Schorfheide, 2002). They use a log-linearization of a real business cycle model to get theoretical values for VAR parameters:

$$y_t = \Gamma'_0 x_t + \varepsilon_t$$
$$E(\varepsilon_t, \varepsilon_t') = \Omega_0$$

Suppose we had an artificial “sample” of observations $\{\tilde{y}_{i}\}_{i=p+1}$ and base $m$, $M$, $N$, and $\Lambda$ on a diffuse inference from this sample:
We could then use these values of $m$, $M$, $N$, and $\Lambda$ together with the actual observed data $\{y_t\}_{t-p+1}^T$ to form a posterior inference using the earlier formulas, where choice of $\tilde{T}$ would reflect how much weight to put on the “real business cycle prior.”

Actually, we can use the RBC’s implied values of $\Gamma_0$ and $\Omega_0$ not just to simulate a few observations, but we can use them to calculate analytically the moments:

$$E(\mathbf{x}, \mathbf{x}) = A_0$$

$$E(\mathbf{x}, \mathbf{y}) = B_0$$

$$E(\mathbf{y}, \mathbf{y}) = C_0$$

where $A_0$, $B_0$, and $C_0$ are functions of $\Gamma_0$ and $\Omega_0$. 

$$M = \left( \sum_{t=1}^{\tilde{T}} \mathbf{x}_t \mathbf{x}_t' \right)^{-1}$$

$$\tilde{T} = M \sum_{t=1}^{\tilde{T}} \mathbf{x}_t \mathbf{y}_t'$$

$$m = \text{vec}(\tilde{T})$$

$$N = \tilde{T}$$

$$\Lambda = \sum_{t=1}^{\tilde{T}} \left( \mathbf{y}_t - \tilde{T} \mathbf{x}_t \right) \left( \mathbf{y}_t - \tilde{T} \mathbf{x}_t \right)'$$
Del Negro and Schorfheide find that putting equal weights on prior and data ($\bar{T} = T$) results in substantially better forecasts than unrestricted VAR, and better than Minnesota prior for horizons greater than 1 quarter.

RBC = good simplification (shrinkage).

Problem with Normal-Wishart prior

$\gamma \Omega^{-1} \sim N\left( \begin{bmatrix} m \\ \Omega \otimes M \end{bmatrix} \right)$

$y_{1,t-1} = \text{first element of } x_t$

$y_{2,t-1} = \text{second element of } x_t$
\[ \gamma \Omega^{-1} \sim N \left( \begin{pmatrix} m_n \cdot & \Omega \otimes M \\ n \times n & k \times k \end{pmatrix} \right) \]

Confidence in \( \gamma_1 \) = coefficient relating \( y_{1r} \) to \( y_{1,1-1} \) is \( \sigma_{11}m_{11} \)

Confidence in \( \gamma_2 \) = coefficient relating \( y_{1r} \) to \( y_{2,1-1} \) is \( \sigma_{11}m_{22} \)

Confidence in \( \gamma_{k+1} \) = coefficient relating \( y_{2r} \) to \( y_{1,1-1} \) is \( \sigma_{22}m_{11} \)

Confidence in \( \gamma_{k+2} \) = coefficient relating \( y_{2r} \) to \( y_{2,1-1} \) is \( \sigma_{22}m_{22} \)

Problem: If \( \sigma_{11}m_{11} > \sigma_{11}m_{22} \) (pretty confident variable 2 doesn’t matter for variable 1), then must have \( \sigma_{22}m_{11} > \sigma_{22}m_{22} \) (think variable 2 doesn’t matter for variable 2 either)
Ways to get around this problem:
(1) Assume that $\Omega$ is diagonal. Then single-equation methods are equivalent to full-system inference, use different $M$, for each equation.

(2) Drop natural conjugates, turn to numerical Bayesian methods.

II. Vector autoregressions
A. Introduction
B. Normal-Wishart priors for VARs
C. Bayesian analysis of structural VARs
Consider structural VAR:

\[
\begin{align*}
A'_{n	imes n} y'_{t} &= B'_{n	imes k} x_{t} + v_{t} \\
x_{t} &= (y'_{t-1}, y'_{t-2}, \ldots, y'_{t-p}, 1) \\
k &= np + 1 \\
v_{t}|y_{t-1}, y_{t-2}, \ldots, y_{t-p+1} \sim N(0, I_{n})
\end{align*}
\]

\[
p(y_{t}|y_{t-1}, y_{t-2}, \ldots, y_{t-p+1}) = p(v_{t}|y_{t-1}, y_{t-2}, \ldots, y_{t-p+1}) \left| \frac{\partial y_{t}}{\partial v_{t}} \right|^{-1}
\]

\[
\frac{\partial y_{t}}{\partial v_{t}} = (A')^{-1}
\]

\[
p(y_{1}, \ldots, y_{t}|A, B; y_{0}, y_{1}, \ldots, y_{p+1}) = (2\pi)^{-T/2}|A|^{-T}
\]

\[
\exp\left[-\frac{1}{2} \sum_{t=1}^{T} \left( A y_{t} - B x_{t} \right)' \left( A y_{t} - B x_{t} \right)\right]
\]

Take transpose

\[y_{t} A = x_{t} B + v_{t}\]

and stack rows on top of each other for \(t = 1, 2, \ldots, T:\)

\[
\begin{bmatrix} Y_{T \times n} \\ A_{n \times n} \end{bmatrix} = \begin{bmatrix} X_{T \times k} \\ B_{k \times n} \end{bmatrix} + \begin{bmatrix} V_{T \times n} \end{bmatrix}
\]
The vec operator stacks the columns of a \((k \times n)\) matrix on top of each other, from left to right, to form a \((kn \times 1)\) vector:

\[
\text{vec}(B) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = b
\]

So first \(k\) elements of the \((nk \times 1)\) vector \(b\) correspond to the coefficients on lags for the first equation in the VAR.
Likewise define
\[ \tilde{y} \equiv \text{vec}(YA) \]
\[ = (I_n \otimes Y)a \]
for \( a \equiv \text{vec}(A) \). First \( n \) elements of \( a \)
correspond to coefficients on contemporaneous variables in first
equation of VAR.

\[ \begin{pmatrix} Y & A \\ T \times n & n \times n \end{pmatrix} = \begin{pmatrix} X & B + V \\ T \times k & k \times n \end{pmatrix} \]
Taking vec of full system,
\[ \tilde{y} = \tilde{X}b + \tilde{v} \]
where \( \tilde{v} \sim N(0, I_n) \). Note that conditional on \( a \), this is a classical regression model
with unit variance for the residual.

\[
p(Y|a, b) = (2\pi)^{-Tn/2}|A|^T \exp\left[-\frac{1}{2} (\tilde{y} - \tilde{X}b)'(\tilde{y} - \tilde{X}b) \right] \\
b|a \sim N(m(a), M(a)) \\
p(b|a) = (2\pi)^{-nk/2}|M(a)|^{-1/2} \exp \left\{ -\frac{1}{2} [b - m(a)]' [M(a)]^{-1}[b - m(a)] \right\} \\
p(a) \text{ arbitrary} \]
\[ p(b, a|Y) = p(b|a, Y)p(a|Y) \]
\[ b|a, Y \sim \mathcal{N}(m^*(a), M^*(a)) \]
\[ M^*(a) = \left\{ [M(a)]^{-1} + \bar{X}' \bar{X} \right\}^{-1} \]
\[ \bar{X} = (I_n \otimes X) \]
\[ M^*(a) = \left\{ [M(a)]^{-1} + [I_n \otimes X'X] \right\}^{-1} \]

\[ b|a, Y \sim \mathcal{N}(m^*(a), M^*(a)) \]
\[ m^*(a) = M^*(a) \left\{ [M(a)]^{-1} m(a) + \bar{X}' \bar{y} \right\} \]
\[ \bar{X}' \bar{y} = (I_n \otimes X')(I_n \otimes Y)a \]
\[ = (I_n \otimes X'Y)a \]

Still, we need to “invert” \((nk \times nk)\) matrix
\[ \left\{ [M(a)]^{-1} + [I_n \otimes X'X] \right\}^{-1} \]
Suppose our prior for equation $i$ takes the form 

$$b_i | a \sim \mathcal{N}(m_i(a), M_i(a))$$

for $M_i(a)$ a $(k \times k)$ matrix, with priors independent across equations.
Specification of $p(b|a)$:
expect each series to behave like a random walk.

Structural equations:
$$Y_{T 	imes n} = X_{T 	imes k}^{n 	imes k} B + V_{T 	imes n}$$
Reduced form:
$$Y_{T 	imes n} = X_{T 	imes k}^{n 	imes k} \Pi + E_{T 	imes n}$$
$$\Pi = BA^{-1}$$
$$E = VA^{-1}$$
Random walk:

\[
\Pi_{k \times n} = \begin{bmatrix}
I_n \\
0 \\
\vdots \\
0
\end{bmatrix} = B_{k \times n} A^{-1}_{n \times n}
\]

\[
E(B|A) = \begin{bmatrix}
A \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

\[
b_i | a \sim N(m_i(a), M_i(a))
\]

\[
m_i(a) = \begin{bmatrix}
a_i \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

Let \( m_i^{(r)} \) denote the \( r \)th element of this vector (prior expectation of \( r \)th element of \( b_i \)) and \( M_i^{(r)} \) its variance:

\[
b_i^{(r)} | a \sim N(m_i^{(r)}, M_i^{(r)})
\]

with priors across coefficients taken to be independent.
Multiplying the likelihood by the prior density
\[ p(b_i^{(r)}|a) = (2\pi)^{-1/2}[M_i^{(r)}]^{-1/2} \exp\left[-\frac{(b_i^{(r)}-m_i^{(r)})^2}{2M_i^{(r)}}\right] \]
is numerically identical to acting as if I had observed a variable \( q_i^{(r)} \) from the system
\[ q_i^{(r)} = b_i^{(r)}/\sqrt{M_i^{(r)}} + v_i^{(r)} \]
where the observed value of \( q_i^{(r)} \) is \( m_i^{(r)}/\sqrt{M_i^{(r)}} \) and \( v_i^{(r)} \sim N(0, 1) \).

Or, since \( m_i^{(r)} \) is the \( r \)th element of \( a_i \) if \( r \leq n \) and is zero otherwise, this is numerically identical to having observed the \((n \times 1)\) vector \( y_i^{(r)} \) and \((k \times 1)\) vector \( x_i^{(r)} \) from the following system:

\[
[y_i^{(r)}]a_i = [x_i^{(r)}]b_i + v_i^{(r)}
\]

\[ y_i^{(r)} = \begin{cases} 
  e_r(n)/\sqrt{M_i^{(r)}} & \text{if } r \leq n \\
  0 & \text{otherwise}
\end{cases} \]

\[ x_i^{(r)} = e_r(k)/\sqrt{M_i^{(r)}} \]

for \( e_r(n) \) the \( r \)th column of \( I_n \)  
\( e_r(k) \) the \( r \)th column of \( I_k \)
Stack these “dummy observations” for $r = 1, 2, \ldots, k$ in matrices

\[
Y_{id}^{k \times n} = \begin{bmatrix}
[ y_i^{(1)} ]'
\vdots
[ y_i^{(k)} ]'
\end{bmatrix}
\quad
X_{id}^{k \times k} = \begin{bmatrix}
[ x_i^{(1)} ]'
\vdots
[ x_i^{(k)} ]'
\end{bmatrix}
\]

The posterior $p(b_i|a, Y)$ is then:

$b_i|a, Y \sim N(m_i^*(a), M_i^*(a))$

$M_i^*(a) = [M_i(a)]^{-1} + X'X^{-1}$

$= (X_{id}'X_{id} + X'X)^{-1}$

$m_i^*(a) = M_i^*(a) \{[M_i(a)]^{-1}m_i(a) + X'Y_{id}\}$

$= (X_{id}'X_{id} + X'X)^{-1}(X_{id}'Y_{id} + X'Y)a_i$

$b_i^{(r)}|a \sim N(m_i^{(r)}, M_i^{(r)})$

Remaining question: how choose $M_i^{(r)}$.

Let $j(r)$ denote which variable the $r$th coefficient refers to ($j = 1, 2, \ldots, n$) and $((r)$ its lag ($i = 1, 2, \ldots, p$)

i.e., $b_i^{(r)}$ is the coefficient relating $y_{j,a}$ to $y_{j,t-i}$
(1) If units in which \( y_{jt} \) is measured are doubled, value of \( b_i^{(r)} \) is cut in half
\[ \Rightarrow \text{make } \sqrt{M_i^{(r)}} \text{ inversely proportional to } \sigma_j, \text{ the standard deviation of univariate autoregression for } y_{jt} \]

(2) Have more confidence in zero priors for bigger \( \ell \)
\[ \Rightarrow \text{make } \sqrt{M_i^{(r)}} \text{ inversely proportional to } \ell^{\lambda_3} \]
\((\lambda_3 > 0)\)

\[ b_i^{(r)} | \alpha \sim N(m_i^{(r)}, M_i^{(r)}) \]
\[ \sqrt{M_i^{(r)}} = \left\{ \begin{array}{ll}
\frac{\lambda_0 \lambda_1}{\sigma_{m_i^{(r)}}^{(r)}} & \text{for } r = 1, 2, \ldots, k - 1 \\
\lambda_0 \lambda_4 & \text{for } r = k
\end{array} \right. \]
where \( \lambda_0 \) controls tightness of prior for \( \alpha \) and \( \lambda_1 \) controls tightness of random walk prior
Distribution of $\mathbf{a}$

$$p(\mathbf{a}, \mathbf{b}| \mathbf{Y}) = p(\mathbf{b}| \mathbf{a}, \mathbf{Y})p(\mathbf{a}| \mathbf{Y})$$

$\mathbf{b}| \mathbf{Y} \sim N(\mathbf{m}^*(\mathbf{a}), \mathbf{M}^*(\mathbf{a}))$

$$p(\mathbf{a}| \mathbf{Y}) \propto p(\mathbf{a})|\mathbf{A}|^T$$

$$|I_n + (I_n \otimes \mathbf{X})\mathbf{M}(\mathbf{a})(I_n \otimes \mathbf{X})|^{-1/2}$$

$$\exp \left\{ -\frac{1}{2} [\mathbf{a}'(I_n \otimes \mathbf{Y}' \mathbf{Y})\mathbf{a} +$$

$$+ \mathbf{m}(\mathbf{a})'\mathbf{M}(\mathbf{a})^{-1}\mathbf{m}(\mathbf{a})$$

$$- \mathbf{m}^*(\mathbf{a})'(\mathbf{M}^*(\mathbf{a}))^{-1}\mathbf{m}^*(\mathbf{a})] \right\}$$

What is this distribution $p(\mathbf{a}| \mathbf{Y})$?
Use numerical methods.

Example: Sims-Zha, distribution of $p(\mathbf{a}| \mathbf{Y})$:

$$p(\mathbf{a}| \mathbf{Y}) \propto p(\mathbf{a})|\mathbf{A}|^T$$

$$|I_n + (I_n \otimes \mathbf{X})\mathbf{M}(\mathbf{a})(I_n \otimes \mathbf{X})|^{-1/2}$$

$$\exp \left\{ -\frac{1}{2} [\mathbf{a}'(I_n \otimes \mathbf{Y}' \mathbf{Y})\mathbf{a} +$$

$$+ \mathbf{m}(\mathbf{a})'\mathbf{M}(\mathbf{a})^{-1}\mathbf{m}(\mathbf{a})$$

$$- \mathbf{m}^*(\mathbf{a})'(\mathbf{M}^*(\mathbf{a}))^{-1}\mathbf{m}^*(\mathbf{a})] \right\}$$

$$\equiv q(\mathbf{a})$$
Importance density $g(a)$ should be similar to $q(a)$ but with fatter tails.

1. Find
   
   $a_0 = \arg \max_a q(a)$

   
   $H_0 = -\frac{\partial^2 \log q(a)}{\partial a \partial a'} \bigg|_{a=a_0}$

(2) Let $g(a)$ be $n^2$-dimensional Student t-distribution centered at $a_0$ with scale matrix $H_0^{-1}$ and 9 degrees of freedom.

(3) Generate $j$th draw $a^{(j)}$ from $g(a)$.

E.g., generate

$u^{(j)} \sim N(0, I_{n^2})$

$v^{(j)} = P_0^{-1} u^{(j)} + a_0$

$P_0$ is Cholesky factor of $H_0$

$e^{(j)} \sim N(0, I_{9})$

$v^{(j)} = (1/9)[e^{(j)}][e^{(j)}]$

$a^{(j)} = v^{(j)}/\sqrt{v^{(j)}}$
(4) Calculate

\[ g(a^{(j)}) = c \left\{ 1 + |a^{(j)} - a_0| H_0^{-1} |a^{(j)} - a_0| \right\}^{-\left(\sigma^2 + 9\right)/2} \]

choosing \( c \) so that \( \max_{j=1,\ldots,D} g(a^{(j)}) \) is same scale as \( \max_{j=1,\ldots,D} q(a^{(j)}) \).

(5) Calculate the weight

\[ \omega^{(j)} = \frac{g(a^{(j)})}{q(a^{(j)})} \]

(6) Repeat steps (3)-(5) for \( j = 1, 2, \ldots, D \).

(7) The sample \( a^{(1)}, \ldots, a^{(D)} \), when weighted by \( K^{-1} \omega^{(1)}, \ldots, K^{-1} \omega^{(D)} \) for \( K = \omega^{(1)} + \cdots + \omega^{(D)} \), is a sample from \( p(a|Y) \), e.g.,

\[ \hat{E}(a|Y) = \frac{\sum_{j=1}^{D} \omega^{(j)} a^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}} \]
(8) For each \( j \), generate \( b^{(j)} \) from the \( N(m^*(a^{(j)}), M^*(a^{(j)}) \) distribution and weight with \( K^{-1} \omega^{(j)} \) for a draw from \( p(b|Y) \), e.g.

\[
\hat{E}(b|Y) = \frac{\sum_{j=1}^{D} \omega^{(j)} b^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}}
\]