ECON5120 Bayesian Data Analysis

Week 10: Time Series

Santiago Montoya-Blandón

University of Glasgow
Adam Smith Business School

March 19, 2025

Lecture outline

1. Multivariate models

2. VAR Models

Models for many variables

- We will begin with a general view at *multivariate* time series methods
- We then get to more "interesting" models such as VARs (Vector Autoregressions)
- Econometrically, the results for the linear regression model are similar to the results in the multivariate case
- The only difference is that now our data x_t, y_t, z_t will be vectors, not scalars
- Dependent y_t and exogenous x_t. Time dimension: $t = 1, \ldots, T$
- The likelihood function is multivariate, and we need to use prior distributions which are multivariate

Question for you

What do you remember about multivariate distributions?

Multivariate distributions

- Multivariate normal: y_t is $p \times 1$ for time series observations $t = 1, \ldots, T$
- Define $Y = [y_1^{'}, y_2^{'}, \ldots, y_{T-1}^{'}, y_T^{'}]^{'}$ a $T \times p$ matrix of all variables
- Y will contain all observations for say GDP, inflation, interest rate, so that $p = 3$

$$y \sim \mathcal{N}(\mu, \Sigma)$$
$$p(y; \mu, \Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{(y-\mu)' \Sigma^{-1} (y-\mu)}$$

- μ is $p \times 1$ and Σ is a $p \times p$ positive definite matrix

Multivariate likelihood

- Assume y is a vector of p asset returns.
- In portfolio analysis we want to estimate the mean and variance of multivariate returns in order to construct our portfolio weights. Thus our model likelihood is:

$$y \sim \mathcal{N}(\mu, \Sigma)$$

- Maximum Likelihood estimators of μ and Σ are simply their sample quantities

$$\mu_{\text{MLE}} = \frac{1}{T}\sum_{t=1}^{T} y_t$$
$$\Sigma_{\text{MLE}} = (y - \mu_{\text{MLE}})(y - \mu_{\text{MLE}})'/(T - p)$$

Multivariate likelihood

- We know from practice these estimators are not good for large p/small T
- For example, in empirical portfolio analysis, portfolio weights are inefficiently estimated
- Bayesian theory provides several shrinkage estimators for such a problem
- We can generalize the univariate analysis by considering conjugate priors for μ and Σ
- Conjugate priors are ones which allow easy computation of the posterior
- Nevertheless, recall Bayes' rule is the same:

$$posterior \propto likelihood \times prior \tag{1}$$

Unknown multivariate mean and covariance

- Conjugate priors for μ and Σ are

$$\mu \sim \mathcal{N}(\underline{\mu}, \underline{V})$$
$$\Sigma \sim \text{IW}(\underline{v}, \underline{S})$$

- $\mathcal{N}(0, V)$ is a p-variate Normal distribution, and $\text{IW}(v, S)$ is the inverse Wishart distribution
- IW distribution is a multivariate extension of the inverse Gamma
- Under these priors, we have the following posteriors

$$\mu \sim \mathcal{N}(\overline{\mu}, \overline{V})$$
$$\Sigma \sim \text{IW}(\overline{v}, \overline{S})$$

- Ignore exact expressions for the posterior hyperparameters $(\overline{\mu}, \overline{V}, \overline{v}, \overline{S})$

Multivariate regression: Standard form

- Assume now that at time t we have a regression model with k exogenous variables

$$y_t = B^\top x_t + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \Sigma)$$

- Using multivariate notation, this model can be written as the **standard form**:

$$Y = XB + e$$

- X is $T \times k$ and B is $k \times p$, so that XB is $T \times p$ (which conforms with the dimension of Y), and e is $T \times p$
- In practice, this formulation is not very flexible because the same X applies in all equations

Standard form

- For example we have for $p = 2$ and $k = 3$

$$y_{t1} = \beta_{11}x_{1t} + \beta_{12}x_{2t} + \beta_{13}x_{3t} + e_{1t}$$
$$y_{t2} = \beta_{21}x_{1t} + \beta_{22}x_{2t} + \beta_{23}x_{3t} + e_{2t}$$

- Using vector notation:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

- Using matrix notation:

$$\begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \vdots & \vdots \\ y_{1T} & y_{2T} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots \\ x_{1T} & x_{2T} & x_{3T} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \end{bmatrix} + \begin{bmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \\ \vdots & \vdots \\ e_{1T} & e_{2T} \end{bmatrix}$$

Natural conjugate prior

- As in the univariate regression model, here we can define the **natural** conjugate priors:

$$b|\Sigma \sim \mathcal{N}(\underline{b}, \Sigma \otimes \underline{V})$$
$$\Sigma \sim \text{IW}(\underline{v}, \underline{S})$$

- $b = \text{vec}(B)$ is a $kp \times 1$ vector of regression coefficients
- The natural conjugate prior for B (or $b = \text{vec}(B)$) is conditional on Σ
- $\underline{V}$ is $k \times k$, so that $\Sigma \otimes \underline{V}$ is $kp \times kp$
- A good thing is that for this prior the joint posterior can be computed analytically (posterior is N/IW)

Problem with natural conjugate prior

- Downside is that the prior covariance of the coefficients in any two equations is proportional to one another i.e., $\underline{V}$ applies in every equation
- For example, let x_t to be lags of y_t: i.e. $x_t = [y_{t-1}, \ldots, y_{t-k}]$
- In practice, we might want different lags of y_t to enter different equations
- Not possible to implement with a natural conjugate prior

The Kronecker form

- Our model is

$$Y = XB + e$$

- Can be written as the observationally equivalent model in terms of the coefficients $b = \text{vec}(B)$ with Σ now $Tp \times Tp$

$$\text{vec}(Y) = (I \otimes X)\,\text{vec}(B) + \text{vec}(e)$$
$$y = Zb + \epsilon, \epsilon \sim N(0, \Sigma)$$

- It is instructive to compare the Least Squares estimators of b and B:

$$b_{\text{LS}} = (Z^\top Z)^{-1}(Z^\top y)$$
$$B_{\text{LS}} = (X^\top X)^{-1}(X^\top Y)$$

- This formulation is flexible and allows for covariates to enter each regression separately as defined by the analyst

Independent prior

- It is better to use the independent prior with $\underline{V}$ now of dimension $kp \times kp$

$$b \sim \mathcal{N}(\underline{b}, \underline{V})$$
$$\Sigma \sim \text{IW}(\underline{v}, \underline{S})$$

- Can now select a prior variance for each predictor in each equation independently
- Joint posterior is not tractable, but conditional posteriors are N/IW:

$$b|\Sigma, Y, X \sim \mathcal{N}(\overline{b}, \overline{V})$$
$$\Sigma|b, Y, X \sim \text{IW}(\overline{v}, \overline{S})$$

$$\overline{V} = (\underline{V}^{-1} + Z^\top \Sigma^{-1} Z)^{-1}, \qquad \overline{b} = \overline{V}(\underline{V}^{-1}\underline{b} + Z^\top \Sigma^{-1} y)$$
$$\overline{v} = T + \underline{v}, \qquad \overline{S} = \underline{S} + (y - Zb)(y - Zb)^\top$$

Lecture outline

1. Multivariate models

2. VAR Models

Vector autoregressions

- Vector autoregressions (VAR) are linear time series models designed to capture joint dynamics in multiple time series
- At first glance, they appear straightforward multivariate extensions of univariate models
- They turn out to be one of the key empirical tools in macro
- The central bank controls the short-term interest rate
- The government controls taxes and fiscal policy
- What are the dynamic responses of outcome variables when the controlled variables change?

Vector autoregressions

- Sims (1980) Macroeconomics and Reality, Econometrica, suggested VARs as a replacement for large-scale macroeconometric models of the 60s which imposed unreasonable economic restrictions
- In general the "Minnesota revolution" of the 70s-80s (incl Bayesians such as: Tom Sargent, Chris Sims, Bob Littermann, Todd Doan, John Geweke) has shaped macroeconomics up to today
- Here we focus on estimation of reduced-form models

Bayesian VARs

- One way of writing VAR(p) model:

$$y_t = A_0 + \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t$$

- y_t is an $M \times 1$ vector of variables of interest
- ϵ_t is an $M \times 1$ vector of disturbances
- A_0 is an $M \times 1$ vector of intercepts
- A_j is an $M \times M$ matrix of coefficients
- $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$, Σ define contemporaneous correlation
- Exogenous variables, or deterministic terms (e.g. time trends) can be added easily (do not add to save on notation)

Bayesian VARs

- Note that

$$y_t = A_0 + \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t$$
$$= A^\top x_t + \epsilon_t$$

- $x_t = (1, y_{t-1}^\top, y_{t-2}^\top, \ldots, y_{t-p}^\top)^\top$ and $A = [A_0^\top, A_1^\top, \ldots, A_p^\top]^\top$ is a $(Mp+1) \times M$ matrix
- Same formulation from the previous section!
- In the notation from the previous section, we have that p is now M and k is now $Mp+1$

Bayesian VARs

- VAR in matrix notation: Y and E be $T \times M$ matrices placing the T observations on each variable in columns next to one another
- Then we can write the VAR as

$$Y = XA + E$$

- In first VAR, α is $kM \times 1$ vector of VAR coefficients, here A is $k \times M$
- Here, we cannot impose restrictions "manually" as the same X applies to all equations
- We will use both notations below. Additional notation is needed to deal with VARs with time-varying parameters (TVP-VAR) ▸ 3rd way of writing the VAR

VAR Likelihood function

- Given the standard way of writing the VAR:

$$p(Y|A,\Sigma,Y_{1-p:0}) \propto |\Sigma|^{-T/2} exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}\widehat{S}\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2}\text{tr}\left[\Sigma^{-1}(A-\widehat{A})^{\top}X^{\top}X(A-\widehat{A})\right]\right\}$$

- $\widehat{A} = (X^{\top}X)^{-1}X^{\top}Y$ and $\widehat{S} = (Y - X\widehat{A})^{\top}(Y - X\widehat{A})$
- Non-informative, improper prior $p(A,\Sigma) \propto |\Sigma|^{-M+1/2}$ leads to

$$p(A,\Sigma|Y) \sim MNIW(\widehat{A},(X^{\top}X)^{-1},\widehat{S},T-K)$$

- We can use Monte Carlo to obtain draws from $p(A,\Sigma|Y)$:
  1. Draw $\Sigma^{(s)}$ from an $\text{IW}(\widehat{S},T-K)$
  2. Draw $A^{(s)}$ from a $MN(\widehat{A},\Sigma^{(s)}\otimes(X^{\top}X)^{-1})$

Natural conjugate prior

- Since the likelihood has a Multivariate Normal-Inverse Wishart structure, the natural conjugate prior is also MNIW

$$\alpha := \text{vec}(A)|\Sigma \sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})$$
$$\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$$

- W denotes the Wishart distribution with its respective prior hyperparameters
- Note: $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$ is equivalent to $\Sigma \sim \text{IW}(\underline{S}, \underline{\nu})$
- Alternatively, you might see the following notation in terms of the matrix A (rather than the vector α):

$$A \sim MN(\underline{\alpha}, \Sigma, \underline{V})$$

- This is known as the Matrix-Variate Normal distribution
- In matrix programming languages, it is more numerically stable to stick with the Normal distribution for a vector, rather than the equivalent matrix form

Natural conjugate prior

- Posterior has analytical form

$$p(\alpha|\Sigma, y) \sim \mathcal{N}\left(\overline{\alpha}, \Sigma \otimes \overline{V}\right)$$
$$p\left(\Sigma^{-1}|y\right) \sim W\left(\overline{S}^{-1}, \overline{v}\right)$$

- Posterior hyperparameters

$$\overline{V} = \left[\underline{V}^{-1} + X^{\top}X\right]^{-1}$$
$$\overline{A} = \overline{V}\left[\underline{V}^{-1}\underline{A} + X^{\top}Y\right]$$
$$\overline{S} = \hat{S} + \underline{S} + \hat{A}^{\top}X^{\top}X\hat{A} + \underline{A}^{\top}\underline{V}\underline{A} - \overline{A}^{\top}\overline{V}^{-1}\overline{A}$$
$$\overline{v} = T + \underline{v}$$

Natural conjugate prior

- Remember: in regression model joint posterior for (β, h) was Normal-Gamma, but marginal posterior for β had t-distribution
- Same thing happens with VAR coefficients
- Marginal posterior for α is multivariate t-distribution
- Posterior mean is the OLS estimate, $\hat{\alpha}$, as the prior becomes non-informative
- Degrees of freedom parameter is $\overline{v}$
- Posterior covariance matrix:

$$\text{Var}(\alpha \mid y) = \frac{1}{\overline{v} - M - 1}(\overline{S} \otimes \overline{V}) \tag{2}$$

- Posterior inference can be done using (analytical) properties of multivariate t-distribution

Problems with natural conjugate prior

- Natural conjugate prior has great advantage of analytical results, but has some problems which make it rarely used in practice.
- To make problems concrete consider a macro example: The VAR involves variables such as output growth and the growth in the money supply
- Researcher wants to impose the neutrality of money.
- Implies: coefficients on the lagged money growth variables in the output growth equation are zero (but coefficients of lagged money growth in other equations would not be zero)

Problems with natural conjugate prior

- Cannot flexibly impose neutrality of money restriction through the prior.
- Cannot set prior variance to very small value
- To see why, let individual elements of Σ be σ_{ij}.
- Prior covariance matrix has form $\Sigma \otimes \underline{V}$
- This implies prior covariance of coefficients in equation i is $\sigma_{ii}\underline{V}$.
- Thus prior covariance of the coefficients in any two equations must be proportional to one another.
- So can impose coefficients on lagged money growth to be zero in ALL equations, but cannot flexibly do so in equation by equation.

VAR in Kronecker form

- Another way to write the model: Let $y = \left(y_1^\top, y_2^\top, \ldots, y_T^\top\right)^\top$ be $MT \times 1$ vector, and ε stacked conformably

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix} = \begin{bmatrix} 1 & y_0 & \cdots & y_{-p+1} \\ 1 & y_1 & \cdots & y_{-p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & y_{T-2} & \cdots & y_{T-p-1} \\ 1 & y_{T-1} & \cdots & y_{T-p} \end{bmatrix}$$

- $k = 1 + Mp$ is number of coefficients in each equation of the VAR and X is a $T \times k$ matrix
- The VAR can be written as, with $\alpha = \text{vec}(A)$,

$$y = (I_M \otimes X)\alpha + \varepsilon$$

- $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_M)$ Jump to an example of $I_M \otimes X$

VAR in Kronecker form

- Some economic theory might imply some variable has no effect on a certain variable
- e.g., money neutrality: money growth does not affect GDP
- We can remove the money growth from the GDP equation manually, in $I_M \otimes X$
- Or we could use a shrinkage prior

Independent Normal-Wishart prior

For comparison, in case 1 we had $W = I_M \otimes X$ which was

$$W = \begin{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} & 0 & \cdots & 0 \\ 0 & \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \end{bmatrix}$$

Independent Normal-Wishart prior

- Define a MNIW without conditioning α on Σ
- Allow possibly different lags to enter different equations
- Let's write VAR in a third way:

$$y_t = (I_M \otimes x_t^\top)\alpha + \epsilon_t$$

- In matrix form (where the vectorization is done on the rows of Y and E):

$$\text{vec}(Y^\top) = Z\alpha + \text{vec}(E^\top)$$

- If we define $z_t = I_M \otimes x_t^\top$ then Z is

$$Z = \begin{bmatrix} I_M \otimes x_1 \\ I_M \otimes x_2 \\ \vdots \\ I_M \otimes x_T \end{bmatrix}$$

Independent Normal-Wishart prior

- Working with VAR

$$y_t = (I_M \otimes x_t)\alpha + \epsilon_t$$

- The Independent Normal-Wishart prior is

$$\begin{aligned} p(\alpha, \Sigma) &= p(\alpha)p(\Sigma) \\ p(\alpha) &\sim \mathcal{N}(\underline{\alpha}, \underline{V}_\alpha) \\ p(\Sigma^{-1}) &\sim W(\underline{S}^{-1}, \underline{\nu}) \end{aligned}$$

- Joint posterior $p(\alpha, \Sigma|Y)$ is not of a known form
- No worries since $p(\alpha|\Sigma, Y)$ is Normal, $p(\Sigma|\alpha, Y)$ is IW $\Rightarrow$ This calls for a Gibbs sampling scheme!

Independent Normal-Wishart prior

For $s = 1, 2, \ldots, S$ draw from the densities

$$\alpha|\Sigma, Y \sim \mathcal{N}(\overline{\alpha}, \overline{V}_\alpha)$$

$$\Sigma^{-1}|\alpha, Y \sim W(\overline{S}^{-1}, \overline{\nu})$$

where

$$\overline{V}_\alpha = \left(\underline{V}_\alpha^{-1} + \sum_{t=1}^{T} z_t^\top \Sigma^{-1} z_t\right)^{-1}$$

$$\overline{\alpha} = \overline{V}_\alpha \left(\underline{V}_\alpha^{-1}\underline{\alpha} + \sum_{t=1}^{T} z_t^\top \Sigma^{-1} y_t\right)$$

$$\overline{S} = \underline{S} + \sum_{t=1}^{T} (y_t - z_t\alpha)^\top (y_t - z_t\alpha)$$

$$\overline{\nu} = \underline{\nu} + T$$

Independent Normal-Wishart prior

- Independent Normal-Wishart is a good start for allowing different equations to have different explanatory variables (lags of y_t)
- E.g. we can introduce LASSO shrinkage priors, model selection priors (see Korobilis, 2013, JAE)
- Additionally, independent Normal-Wishart priors can become noninformative in the limit:
- The conditional posterior of Σ becomes equivalent to ML when $\underline{S} = \underline{\nu} \to 0$
- The conditional posterior of α becomes equivalent to ML when $\overline{V}_\alpha \to \infty$ and $\underline{\alpha} = c$ with c finite constant (or simply $c = 0$)

The importance of shrinkage in VARs

$$y_t = z_t\alpha + \epsilon_t$$

- α has $kM = M \times (1 + Mp)$ elements, Σ has $M(M+1)/2$ elements
- A 3-variable VAR with 4 lags has $M = 3$, $kM = 39$, $M(M+1)/2 = 6$
- A 9-variable VAR with 4 lags has $M = 9$, $kM = 333$, $M(M+1)/2 = 45 \rightarrow$ It becomes evident that α explodes as the dimension of the VAR increases
- By using Bayesian shrinkage, we can shrink specific elements of α to zero, thus saving degrees of freedom
- E.g. Banbura, Giannone, Reichlin (2010), JAE, use a VAR with 130 variables and 13 lags: α has more than 200,000 elements but estimation is feasible!

The Minnesota prior

- "Minnesota prior" is the first attempt for shrinkage in VARs
- Developed in U of Minnesota in the '70s by Bob Litterman (as in "Black-Litterman" portfolio model of the '90s)
- Originally developed using the natural conjugate prior
- Here we present the modern, Gibbs-sampler version based on the independent Normal-Wishart prior with $\underline{V}_{MIN}$ a diagonal matrix

$$\alpha \sim N(0, \underline{V}_{MIN})$$

- We have lags $r = 1, \ldots, p$, equations $i = 1, \ldots, M$, and variables $j = 1, \ldots, p + 1$

The Minnesota prior

$$\underline{V}_{kk,MIN} = \begin{cases} \underline{a}_1/r^2 & \text{on coeff of variable } i, \text{ eq. } i \\ \underline{a}_2/r^2 \times s_i/s_j & \text{on coeff of variable } j, \text{ eq. } i \\ \underline{a}_3 \times s_i & \text{on intercept/exogenous} \end{cases}$$

- $\underline{a}_1, \underline{a}_2, \underline{a}_3$ are preselected constants
- s_i is the variance of the residual of the AR(p) model on variable Y_i
- $r = 1, \ldots, p$ so that as $r \uparrow$, $\underline{V}_{kk,MIN} \downarrow$
  $\rightarrow$ Penalize more distant lags
- a_3 is "large", a_1 is "small", a_2 is even smaller
- Constants need not be shrunk; own lags are important; lags of other variables less important (a-priori)
- Thus, as $\underline{V}_{kk,MIN} \rightarrow 0$, $\alpha_k \rightarrow 0$ which is the prior mean, otherwise it is unrestricted

Theil's mixed estimation

- Typical Bayesian inference treats likelihood and prior as being two distinct elements that combined give us the posterior
- A different point of view can be obtained if we mix prior and likelihood together
- We can do so by creating dummy observations generated from our prior and staking them to our observed data
- For special classes of priors we can just do things analytically
- Assume $x^* = [x, \tilde{x}]$ and $y^* = [y, \tilde{y}]$, where $\tilde{x}, \tilde{y}$ are data from prior.
- Then OLS estimate using x^*, y^* is equivalent to Bayes inference under the prior that generated $\tilde{x}, \tilde{y}$

Example: Minnesota-type prior

- Banbura, Giannone, Reichlin (2010) use this approach combined with a Minnesota-type prior
- Consider the VAR (using their notation) where $x_t = (1, y_{t-1}, \ldots, y_{t-p})$, and $\varepsilon_t \sim N(0, \Psi)$

$$y_t = c + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t$$
$$y_t = B x_t + \varepsilon_t$$

- The priors that BGR(2010) set are of natural conjugate form

$$\text{vec}(B)|\Psi \sim N(\text{vec}(B_0), \Psi \otimes \Omega_0), \qquad \Psi \sim \text{IW}(S_0, a_0)$$

- And the prior moments are of the Minnesota type

$$\mathbb{E}[(A_k)_{ij}] = \begin{cases} \delta_i & i = j, k = 1 \\ 0 & \text{otherwise} \end{cases}, \text{Var}((A_k)_{ij}) = \begin{cases} \lambda^2/k^2 & j = 1 \\ (\lambda^2/k^2)(\sigma_i^2/\sigma_j^2) & \text{otherwise} \end{cases}$$

Example: Minnesota-type prior

- The way to represent the Minnesota prior moments using dummy variable representation of the natural conjugate prior is through the following data:

$$Y_d = \begin{bmatrix} 0_{1\times M} \\ \operatorname{diag}(\delta_1\sigma_1, \ldots, \delta_M\sigma_M)/\lambda \\ 0_{M(p-1)\times M} \\ \vdots \\ \operatorname{diag}(\sigma_1, \ldots, \sigma_M) \\ \vdots \end{bmatrix}, \quad X_d = \begin{bmatrix} 0_{1\times Mp} & \varepsilon \\ J_p \otimes \operatorname{diag}(\sigma_1, \ldots, \sigma_M)/\lambda & 0_{Mp\times 1} \\ \vdots & \vdots \\ 0_{M\times mp} & 0_{Mp\times 1} \\ \vdots & \vdots \end{bmatrix}$$

- $J_p = \operatorname{diag}(1, 2, 3, \ldots, p)$
- First block is uninformative prior on intercepts (ε is a very small number)
- Second block of dummies imposes prior beliefs on the autoregressive coefficients
- Third block implements the prior for the covariance matrix

Example: Minnesota-type prior

- As long as we know the prior hyperparameters, we can easily construct the data X_d, Y_d
- For data in log-levels we set $\delta_i = 1$
- σ_i is obtained from univariate autoregressions, before estimation of VAR
- Regression model augmented with dummies is

$$Y^\star = X^\star B + E$$

- We can use a diffuse prior (or simply OLS) on this augmented VAR
- $\tilde{B} = \left(X^{\star\top} X^\star\right)^{-1} \left(X^{\star\top} Y^\star\right)$ and $\tilde{\Psi} = (Y^\star - X^\star \tilde{B})^\top (Y^\star - X^\star \tilde{B})$
- These quantities incorporate now the Minnesota prior restrictions

Other examples

- This idea of mixed estimation can be used in other settings
- Example, we will see later how to use it to impose DSGE priors on a VAR
- **Controlling for trends:** If we assume that the average value of $y_{i,t}$ (say $\overline{y}_i$) is a good predictor of y_t in equation i, then we can construct the following dummies

$$Y^d = \text{diag}\left(\frac{\overline{y}}{\mu}\right)$$
$$X^d = \left[0_{M\times 1}, Y^d, Y^d, \ldots, Y^d\right]$$

- When $\mu \to \infty$ the prior becomes uninfomrative
- When $\mu \to 0$ prior implies the presence of a unit root in each equation and rules out cointegration.
- This is the "sum-of-coefficients" prior

Other examples

- **Controlling for seasonality:** In the previous prior replace with the average value of y_{t-j}
- **Controlling for cointegration:** The previous prior is not consistent with cointegration in the limit
- Solution is the "dummy-initial-observation" prior:

$$Y^d = \begin{pmatrix} \overline{y} \\ \overline{\delta} \end{pmatrix}$$
$$X^d = \left[1/\delta, Y^d, Y^d, \ldots, Y^d\right]$$

- As $\delta \to 0$ all the variables of the VAR are forced to be at their unconditional mean.
- The system is characterized by the presence of an unspecified number of unit roots without drift