The Minnesota Prior

November 19, 2019

Its properties

- It is symmetric in variables except for scale factors, implying it has no economic content.
- It is conjugate: It can be implemented entirely through dummy observations.
- It shrinks toward unit roots, which dampens the tendency of VAR estimates to imply unrealistic long-run forecasting power from initial conditions of the sample.

The Minnesota Prior: notation

$$y(t) = c + \sum_{s=1}^{k} B_s y(t-s) + \varepsilon(t)$$

with data for $t = 1, \ldots, T$ and $\varepsilon(t) \mid \{y(t-s), s \geq 1; B, c, \Sigma\} \sim N(0, \Sigma)$.

MN prior: single-unit-root prior

We introduce data for the artificial date t^* in which

$$\underset{n\times 1}{y(t^*)} = y(t^*-1) = \ldots = y(t^*-k) = \underset{n\times 1}{\bar{y}}\ \lambda$$

and the vector of 1's that corresponds to the constant term in the data matrix is set to λ in the t^* observation. The vector $\bar{y}$ is usually set to the sample mean of the initial conditions, i.e.

$$\bar{y} = \frac{1}{k}\sum_{s=1}^{k} y(1-k)\,.$$

Regression equation form

$$\bar{y}\lambda = \left(\sum_{s=1}^{k} B_s\right)\bar{y}\lambda + c\lambda + \varepsilon(t^*)$$

$$(I - B(1))\bar{y} = c + \varepsilon(t^*)\,.$$

Thus the prior is centered on a part of the parameter space where either $c = 0$ and the system contains a unit root with $\bar{y}$ as its eigenvector, or $c \neq 0$, y is stationary, and $y(0)$ is close to the model's implied population mean.

Connecting beliefs about c and B

One may want to favor the unit-root part of this region. Can do so by omitting the "constant" from the dummy observation, by adding a dummy observation that asserts a prior directly on c, by adding a dummy observation that omits the constant in addition to the one that includes it, or by including the multiple-unit-root dummy observations.

It is a good property of the single-unit-root dummy that it makes beliefs about c connect to beliefs about the presence of non-stationarity.

Unit roots convert constants into polynomial trends.

MN prior: multiple unit roots

$$y(t_j^*) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \bar{y}_j\lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix} = y(t_j^* - 1) = \ldots = y(t_j^* - k)\,,$$

with the $\bar{y}_j$ in the bracketed vector occurring in the j'th position, and the column of the data matrix corresponding to c set to 0.

Regression equation form

This j'th dummy observation reads, for the j'th equation,

$$\lambda \bar{y}_j = \sum_s B_{jjs} \bar{y}_j \lambda + \varepsilon_j(t_j^*)$$

and for equation $k \neq j$,

$$0 = \sum_s B_{kjs} \bar{y}_j + \varepsilon_k(t_j^*) \, .$$

It is easy to see that this prior favors B's whose off-diagonal elements are small, and also B's that are close to putting a unit root in the $B_{jj}(L)$ polynomial for each j.

MN prior: the original

- Postulates a separate set of dummy observations for each equation.
- Independent distributions for all coefficients, with the coefficient

$$B_{jks} \sim N(\mu_{jks}, \pi_1^{-1}\pi_2^{1-\delta(j,k)}s^{-\pi_3})$$

  and $\mu_{jks} = 0$, except $\mu_{jj1} = 1$ for own first lags.
- With $\pi_2 < 1$ coefficients on own lags ($B_{jj}(s)$, $s = 1, \ldots, k$) are likely to be larger in absolute value than coefficients on other variables.
- With $\pi_3 > 0$ coefficients on more distant lags are likely to be smaller.

MN prior: the original

- $\pi_2 \neq 1$ (own and cross lag effects with different prior variances) makes this prior non-conjugate: It cannot be implemented with system-wide dummy observations, and equation-by-equation OLS is therefore not MLE.

- $\pi_2 = 1$ is now the most common choice. Older literature used $\pi_2 < 1$, but also used inefficient equation-by-equation estimation that undermined likelihood-based interpretation of results.

System dummy observations for $\pi_2 = 1$ MN prior

If t^*_{ks} is the artificial date for the dummy observation applying to the s'th lag of the k'th variable, one of these dummy observations has the form

$$y(t^*_{k1}) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \pi_1\sigma_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} = y(t^*_{k1} - 1)\,, \qquad y(t^*_{ks} - s) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \pi_1 s^{\pi_3}\sigma_k \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$y(t^*_{ks} - v) = 0 \text{ for } v \neq s > 1 \text{ and for } v > s = 1\,.$$

Setting σ_k

The parameter σ_k is a measure of the degree of variability in the k'th variable. It has most commonly, again somewhat inconsistently, been set by estimating a low-order univariate AR regression for each variable and taking σ_k to be the standard deviation of the residual of that regression. It could also be set as the sample standard deviation of the initial conditions for each variable, or on the basis of rough a priori reasoning about the likely scale of variation in each variable. Usually only the order of magnitude matters.

Variance dummy observations

These express a belief about the expected size of residuals. w reflects the strength of belief that residuals are small, and σ_k matches the scale of variation expected. w can be zero, in which case these dummy observations disappear, though the σ_k parameters are still needed in other parts of the prior. These dummy observations take the form

$$y(t_k^*) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ w\sigma_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad y(t_k^* - s) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{for } s > 0\,.$$

Del Negro and Schorfheide

- Use these same dummy observations, except center at the structural model.
- Apply them to the VARMA form.
- The above is not actually the original DN&S proposal.
- The original: Use the model to produce an implied $N(\mu, \Sigma)$ form for the posterior on the coefficients of a k'th order VAR fit to data generated from the structural model.