Multivariate predictors

Economic Modeling Unit, International Monetary Fund

November 2009

1 Bayesian Vector Autoregressions (BVARs)

Let $Y_t = (y_{1,t}, y_{2,t}, \ldots, y_{N,t})'$ be a set of time series with a reduced-form VAR(p) representation:

$$Y_t = c + \sum_{k=1}^{p} B_k Y_{t-k} + u_t \tag{1}$$

where $c = (c_1, \ldots, c_N)'$ is an n-dimensional vector of constants, B_k is an $N \times N$ autoregressive matrix, and u_t is an N-dimensional white noise process with covariance matrix $\mathrm{E} u_t u_t' = \Psi$.

The Litterman (1986) prior, often referred to as the Minnesota prior, shrinks the diagonal elements of B_1 towards one and the other coefficients $(B_1, \ldots, B_p)$ towards zero:

$$Y_t = c + Y_{t-1} + u_t \tag{2}$$

The moments for the prior distribution of the coefficients are:

$$\mathrm{E}[(B_k)_{ij}] = \begin{cases} \delta_i, & j = i, k = 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad \mathrm{V}[(B_k)_{ij}] = \left(\frac{1}{\mu_1} \frac{1}{k^{\lambda}} \frac{\sigma_i}{\sigma_j} \right)^2 \tag{3}$$

The Minnesota prior thus embodies the belief that more recent lags provide more useful information than more distant ones. The coefficients $B_1, \ldots, B_p$ are assumed to be independent and normally distributed. Here, it is assumed that the covariance matrix of the residuals Ψ has an inverse Wishart prior distribution, following Sims and Zha (1998).[1] The prior on the intercept is diffuse. Note that the random walk prior, $\delta_i = 1$ for all i, reflects a belief that all the variables are highly persistent. However, the researcher can also incorporate priors where some variables are characterised by a degree of mean-reversion, $0 \leq \delta < 1$.

The overall tightness of the prior distribution around δ_i is governed by the hyperparameter μ_1: $\mu_1 = \infty$ imposes the prior exactly so that the data do not inform the parameter estimates, and $\lambda = 0$ removes the influence of the prior altogether. The factor $1/k^{\lambda}$ is the rate at which the prior standard deviation decreases with the lag length of the VAR, and σ_i/σ_j accounts for the different scale and variability of the data.

The sums of coefficients prior of Doan, Litterman and Sims (1984) is a modification of the Minnesota prior that is motivated by the frequent practice of

[1] Litterman's original assumption that the residual covariance matrix is fixed and diagonal has been removed, and the hyper-parameter λ has also been added.

specifying a VAR in first differences. The sums of coefficients prior is best described by writing the VAR in error correction form:

$$\Delta Y_t = c - (I_N - B_1 - \ldots - B_p)Y_{t-1} + C_1 \Delta Y_{t-1} + \ldots + C_{p-1}\Delta Y_{t-p+1} + u_t \quad (4)$$

The sums of coefficients prior shrinks $(I_N - B_1 - \ldots - B_p)$ towards zero, where a hyperparameter μ_2 controls the degree of shrinkage. As $\mu_2 \to \infty$ the VAR will increasingly satisfy the prior, while lower values of μ_2 will loosen the prior until, when $\mu_2 = 0$, the prior has no influence on VAR estimates. The sums of coefficients restriction implies that there are as many stochastic trends in the VAR as there are $I(1)$ variables. Sims and Zha (1998) discuss a prior that makes some allowance for stable, long-run cointegrating relationships amongst the variables in the system. This 'co-persistence' prior is governed by the hyperparameter μ_3. As $\mu_3 \to \infty$, the VAR will increasingly satisfy the prior, while as $\mu_3 \to 0$ there will be increasingly more stochastic trends in the system.

The Litterman, sums of coefficients, and co-persistence priors can be implemented using dummy observations.

Writing the VAR in matrix notation yields:

$$Y = XB + U \quad (5)$$

where $Y = (y_1, \ldots, y_T)'$, $X = (X_1, \ldots, X_T)'$, $X_t = (Y'_{t-1}, \ldots, Y'_{t-p}, 1)$, $U = (u_1, \ldots, u_T)'$, and $B = (B_1, \ldots, B_p, c)'$ is the $k \times N$ matrix of coefficients with $k = Np + 1$. The form of the prior is then:

$$\Psi \sim iW(S_0, \alpha_0) \quad \text{and} \quad B|\Psi \sim N(B_0, \Psi \otimes \Omega_0) \quad (6)$$

where the parameters B_0, Ω_0, S_0, and α_0 satisfy the prior expectations for B and Ψ.

The modified Litterman prior is implemented by adding dummy observations to the system. It can be shown that adding T_d dummy observations Y_d and X_d is equivalent to imposing the Inverse-Wishart prior with $B_0 = (X'_d X_d)^{-1} X'_d Y_d$, $\Omega = (X'_d X_d)^{-1}$, $S_0 = (Y_d - X_d B_0)'(Y_d - X_d B_0)$, and $\alpha_0 = T_d - k - N - 1$.

Augmenting the system with dummy observations yields:

$$Y^* = X^* B + U^* \quad (7)$$

where $Y^* = (Y', Y'_d)'$, $X^* = (X', X'_d)'$ and $U^* = (U', U'_d)'$. After adding the diffuse prior $\Psi \propto |\Psi|^{-(N+3)/2}$, which ensures the existence of the prior expectation of Ψ, the posterior has the form:

$$\Psi|Y \sim iW(\hat{\Sigma}, T_d + 2 + T - k) \;\; \text{and} \;\; B|\Psi, Y \sim N(\hat{B}, \Psi \otimes (X^{*\prime} X^*)^{-1}) \quad (8)$$

where $\hat{B} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$ and $\hat{\Sigma} = (Y^* - X^*\hat{B})'(Y^* - X^*\hat{B})$. Thus, the posterior expectation of the parameters coincide with the OLS estimates of the dummy-augmented system. As the prior is loosened, the posterior parameter estimates will tend to the OLS estimates from the original, un-augmented system.

1.1 Conditional forecasting

The reduced-form VAR can be written as::

$$Y_t = c + \sum_{k=1}^{p} B_k Y_{t-k} + A_0^{-1}\epsilon_t \tag{9}$$

where the relationships between the reduced-form parameters and the the structural parameters are $c = A_0^{-1}C$, $B_k = A_0^{-1}A_k$, with $u_t = A_0^{-1}\epsilon_t$. Given data up to time T, the h-step out-of-sample forecast at time T can then be decomposed:

$$Y_{T+h} = D + \sum_{j=1}^{h} M_{h-j}\epsilon_{T+j}, \qquad h = 1, 2, \ldots \tag{10}$$

where:

$$\begin{aligned}
&M_0 = A_0^{-1} \\
&M_i = \textstyle\sum_{j=1}^{i} B_j M_{i-j}, \qquad i = 1, 2, \ldots \\
\\
&B_j = 0 \quad \text{ for } \quad j > p
\end{aligned}$$

This forecast decomposition consists of two parts. The first term, D, includes the initial conditions and produces dynamic forecasts in the absence of shocks, while the second term is the dynamic impact of future structural shocks. Future shocks impact on the variables in the VAR through the matrix of impulse response M_i. A conditional forecast is then defined to be when constraints are imposed on future values of variables and/or shocks.

Conditional forecasts are constructed on the basis of imposing future values for some variables (or, equivalently, for future reduced-form shocks). Doan, Litterman and Sims (1984) show that a unique and optimal (in the least squares sense) vector of forecast errors that satisfy the constraints on the forecasts is given by:

$$\epsilon = R'(RR')^{-1}r \tag{11}$$

where R is a $q \times k$ stacked matrix from the impulse responses $M_{h-h_n}(.,j)$, ϵ is a $k \times 1$ vector correspondingly stacked from ϵ_{t+h_n}, and r is a $q \times 1$ vector of constraints, where k is the total number of future shocks, q is the number of constraints, and $h_n = 1, \ldots, h$.

Waggoner and Zha (1999) show that with conditions imposed on future variables (or reduced form shocks) the forecast distribution is invariant to orthonormal transformation of the system.

2 Dynamic factor model

The dynamic factor model assumes that a panel of data can be decomposed into two orthogonal unobserved components: the common component and the idiosyncatic component. The common component captures the bulk of the covariation between the series in the panel and is driven by a handful of shocks, while the idiosyncatic components are local and affect a only a limited number of series.

2.1 Specification

The following specification is used in Doz and others (2007), Gianonne and others (2005,2008), and Matheson (2009).

$$X_t = \Lambda F_t + \varepsilon_t \tag{12}$$
$$F_t = A F_{t-1} + B u_t \tag{13}$$

where:

$X_t = (x_{1t}, \ldots, x_{nt})'$ is an $(n \times 1)$ stationary process

$F_t = (f_{1t}, ..., f_{rt})'$ is a $(r \times 1)$ stationary process (static factors)

Λ is an $(n \times r)$ matrix of factor loadings (ΛF_t is the common component)

$\varepsilon_t = (\varepsilon_{1t}, ..., \varepsilon_{nt})'$ is a $(n \times 1)$ stationary process (idiosyncratic component), with $\mathrm{E}(\varepsilon_t \varepsilon_t') = \Psi$

$u_t = (u_{1t}, ..., u_{qt})'$ is a $(q \times 1)$ stationary process (dynamic factors), with u_t $WN(0, I_q)$

A is an $r \times r$ matrix with all roots $\det(I_r - A_z)$ inside the unit circle

B is a $r \times q$ matrix of rank q

There are two key hyper-parameters that need to be determined: the number of static factors r and the number of dynamic factors q, where $r \geq q$. While formal statistical criteria can be used to determine r and q, these parameters can also be chosen using rules of thumb.[2]

2.2 Two-step estimation procedure

The two-step procedure for estimating the dynamic factor model is detailed in Gianonne and others (2007).

Steps:

1. Estimate the static factors and then run a VAR on the static factors.
2. Re-estimate the factors using the Kalman filter.

Note that when there are missing observations, Step 1 occurs only on the part of the sample where all series have the same number of observations. The Kalman filter used in Step 2 allows us to 'back out' (predict) any observations missing at the beginning of Step 1.

References

Bai, J. and S. Ng (2002). "Determining the number of factors in approximate factor models," Econometrica, 70(1), 135–172.

Bai, J. and S. Ng (2007). "Determining the number of primitive shocks in factor models," Journal of Business and Economic Statistics, 25(1), 52–60.

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projections using realistic prior distributions. Econometric Reviews, 3, 1- 100.

Doz, C, D Gianonne, and L Reichlin (2007), " A two-step estimator for large

[2]See Bai and Ng (2002) and Bai and Ng (2007) for information criteria for determining r and q. Gianonne and others (2005) choose the number of factors based on the proportion of the variation in the panel explained by the factors.

approximate dynamic factor models based on Kalman filtering," Centre for Economic Policy Research (CEPR), Discussion paper, 6043.

Gianonne, D, L Reichlin, and L Sala (2005),"Monetary policy in real time," in NBER Macroeconomics Annual 2004, eds M Gertler and K Rogoff, MIT Press, Cambridge, Mass.

Gianonne, D, L Reichlin, and D Small (2008), "Nowcasting GDP and inflation: The real-time informational content of macroeconomic data releases," Journal of Monetary Economics, 55(4), 665–676.

Litterman, R. (1986). "Forecasting with Bayesian vector autoregressions - five years of experience." Journal of Business and Economic Statistics, 4, 2538.

Matheson, T.D (2009), " An analysis of the informational content of New Zealand data releases: The importance of business opinion surveys," Economic Modelling, forthcoming.

Sims, C. A. and T. Zha (1998). Bayesian methods for dynamic multivariate analysis. International Economic Review 39 (4), 949968.

Waggoner, D. F. and T. Zha (1999, November). Conditional forecasts in dynamic multivariate models. The Review of Economics and Statistics 81 (4), 639651.