A MINNESOTA-STYLE PRIOR FOR VAR'S

CHRISTOPHER A. SIMS

ABSTRACT. Minnesota-style priors for multivariate time series models m treat interactions among variables symmetrically. They express beliefs about time series properties without reference to substantive knowledge about the variables. The version described here is implemented in the `varprior()` function of the VARmodels R package.

The prior is for the parameters of a reduced-form VAR model,

$$y_t = B(L)y_t + Cx_t + \varepsilon_t \,. \tag{1}$$

where y_t is an n-dimensional vector of time series, $B(L)$ is an $n \times n$ matrix polynomial of finite order k in strictly positive powers of the lag operator L, C is a vector of constants, and

$$\varepsilon_t \mid \{y_s, s < t\} \sim N(0, \Sigma) \,. \tag{2}$$

Though the prior is a (possibly improper) normal-inverse-Wishart distribution on the coefficients in $B(L)$ and the elements of Σ, it is built up from "dummy observations" that have understandable interpretations and that make it natural to vary the properties of the prior by changing subsets of the dummy observations. This is in the spirit of Henri Theil's (1961) "mixed estimation".

I. THE MEAN FOR B AND C

All the dummy observations are consistent with the same prior mean on $B(L)$ and C. All the coefficients B_{ijs} for $i \neq j$ have prior mean 0. The prior means of own-lag coefficients B_{iis} may be non-zero. The most common choice in applied work has been to make the prior mean of B_{ii1} 1.0 for all i, which centers the prior at independent discrete time random walks.

However, much economic data (e.g. national income and product accounts) is in principle time averaged over the time unit. A Wiener process in continuous time, averaged over the time unit, delivers after discrete time sampling not a discrete time random walk, but an ARMA(1,1) that is very close to the second order AR with coefficients 1.25, -.25, and this works better than the usual simple discrete time random walk as a prior mean in many cases.

Date: August 18, 2021.

With seasonal data one could consider a prior mean (for quarterly data) of $(1, 0, 0, 1, -1)$ for the own lag coefficients, for example. For stationary data a mean of zero might be appropriate.

When the sum of prior mean own-lag coefficients is not one, the persistence dummy observations described below can imply a non-zero prior mean for the constant. Otherwise, the prior means of all the elements of C are zero.

In the VARmodels package, the own lag prior means are specified in the parameter `OwnLagMeans`, and may vary across equations.

II. Single-coefficient dummy observations

Each dummy observation can be thought of as a $(k + 1) \times (n + m)$ block of data, where m is the number of x variables, including the constant. In the single-coefficient dummy observations, just one right-hand-side variable (lagged y or current x) is non-zero, and the left-hand-side variables (current y) are set to their prior expected values given the prior mean of the coefficient on the non-zero right-hand-side variable and the value of the dummy observation on that variable.

These dummy observations need to be scaled by the degrees of variability of the variables whose coefficients they match. The prior includes a vector of parameters of length n called `sig`, which we'll call σ in these notes. σ should consist of prior estimates of the standard deviations of disturbances in in the corresponding variables. There is a similar parameter for the variability of the x variables, called `xsig` in VARmodels. If the only x is the constant, then the default `xsig=NULL` is sensible. It implies that the mean of the constant is determined by the persistence (λ, μ) dummies below. If there are non-constant x's, non-null `xsig` is probably required to deliver a proper prior. In that case to leave determination of the prior on the constant to λ, μ, set the last element of `xsig` (which always corresponds to the constant) to zero.

A single parameter, τ here and `tight` in VARmodels, determines the overall weight on all the single-coefficient dummy observations.

The single-coefficient dummy observations are chosen to imply that coefficients on lagged variables are expected to be smaller the longer the lag, via a parameter we'll call θ that in VARmodels is called `decay`.

The general formula for the dummy observation corresponding to the lag coefficient on the j'th variable in the i'th equation at the ℓ'th lag is

$$y_{id} = \tau \ell^{\theta} \sigma_j \bar{B}_{ijl},\ i = 1, \dots, n \tag{3}$$

$$y_{j,d-\ell} = \tau \ell^{\theta} \sigma_j\,. \tag{4}$$

This represents a "prior observation" that when the j'th variable's ℓ'th lag is set to (4), and no other terms on the right-hand-side are non-zero, having the coefficient on that variable at its prior mean of $\bar{B}_{ij\ell}$ produces a perfect fit. The standard deviation of this "noisy observation" is the residual standard deviation of the i'th equation, which we hope is close to σ_i, so for $i = j$ and $\ell = 1$, the implied standard deviation for B_{ii1} itself is approximately $1/\tau$.

III. Persistence dummies

The persistence dummies are four types. One is a single dummy observation in which all current and lagged values of y are set to the same vector $\bar{y}$ and all current and lagged values of x are set to the same vector $\bar{x}$. The second type, of which there are n, sets all current and lagged values of y to zero except those for current and lagged y_i, which are all set to $\bar{y}_i$, and sets all current and lagged values of x to $\bar{x}$. The other two types have the same values for y and set all current and lagged values of x to zero. (Note that in all the dummy observations, lagged values of x are just place-holders; their values do not affect anything.)

Both these types of dummy observations express a belief that if y has been constant at $\bar{y}$ over the past k periods, it is likely to take that same value in the current period. When the sum of own lag coefficient prior means is 1.0, , then conditional on B being at the prior mean, the prior expectation of the coefficients on the x variables, including the constant, are pulled toward zero. This makes sense, because when the sum of own lag coefficients is 1.0 and other coefficients are at their zero means, the model is non-stationary, so the constant becomes the expected linear trend rate of growth, which we usually want to stay similar in magnitude to the residual standard deviation.

When the sum of own lag coefficients is less than one and the only x variable is the constant, both the first two types of persistence dummy imply that the unconditional mean of y should be close to $\bar{y}$. The dummy observation can be rearranged as

$$\bar{y}_i = \frac{C_i}{1 - \sum_\ell B_{ii\ell}}, \tag{5}$$

and the right-hand side of this equation is the model-implied unconditional mean of y_i, when all coefficients except those on own lags in the constant are zero. The difference between the single dummy with all y's set to $\bar{y}$ and the n equation-by-equation dummies is that the latter disfavor persistence arising from cross-variable dependence. The single dummy does not disfavor cointegration, in other words, whereas the n separate dummies do disfavor cointegration.

If there are x's beyond the constant, and these have been given priors via single-coefficient dummy observations (non-null `xsig` in VARmodels), then at the prior means of y coefficients and of the non-constant x coefficients, the constant term is pulled in the same direction as if there were only a constant. However, as the coefficients diverge from their prior means, the prior for the unconditional means for y start to depend on $\bar{x}$ as well as $\bar{y}$ and the constant.

If the sum of coefficients on own lags has prior mean less than one, these persistence dummies (assuming x variables are included) are mainly expressions of belief that unconditional sample means implied by the model should be close to $\bar{y}$. As we have discussed, even when the sum of own lags has prior mean of one, the persistence dummies including the x's disfavor stationary models with $\bar{y}$ far from the model-implied unconditional mean. This is important, because it disfavors estimates that imply stationarity combined with initial conditions that are unlikely to recur. If, as in the VARmodels default for `rfmdd`, $\bar{y}$ and $\bar{x}$ are set as mean values of the initial conditions, this amounts to disfavoring models that imply early-sample behavior is generated by a large initial transient that is unlikely to recur.

The versions of these dummy observations that put no weight on the x coefficients pull the prior toward implying zero unconditional means. This seldom makes sense, but since applied work has often used such dummy observations, the option to use them is present in VARmodels, by making `lambda` and `mu` negative.

IV. Variance dummies

These are observations in which the current value of the i'th variable is set to $w\sigma_i$ and all other variables, including the variable whose coefficient is the constant, are set to zero. This set of observations implies prior belief that the equation residuals have standard deviation $\sigma_i w$ and that they are independent across equations. The observations tend to pull the estimate of Σ toward diagonality. Even if one has no strong prior beliefs about Σ, it may be useful to use these dummy observations, perhaps with small weight w, in order to guarantee a proper prior, especially when n is large. $w = 1$ is the default choice. These dummy observations will have little or not effect on the posterior mean of the B and C coefficients. But if reduced form estimates are mapped into the parameters of a just-identified structural VAR, of course estimates of the SVAR parameters will be affected.

References

THEIL, H. (1961): *Economic Forecasts and Policy*. North-Holland Publising Company.

Department of Economics, Princeton University
Email address: sims@princeton.edu