Empirical Methods for Macroeconomics,

Christiano

BVAR Exercise[1]

1. The following exercise is designed to illustrate some basic properties of BVARs and of the Minnesota priors. The example is much simpler than the type of VAR you would be working with in practice, but the question is not designed for realism. It works only with artificial data that you are asked to generate. Suppose that a 2×1 vector of data, y_t, has a $VAR(1)$ representation:

$$y_t = A_1 y_{t-1} + u_t, \; u_t \sim N(0, \Sigma), \tag{1}$$

where

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

and

$$A_1 = \begin{bmatrix} a & c \\ c & b \end{bmatrix}.$$

Set $\sigma_i = 0.01$, for $i = 1, 2$ and $\rho = 0.9$. Also,with $a = 0.8$, $b = 0.8$ and $c = 0.1$. Generate $T = 100$ observations, $y_1, ..., y_T$, setting $y_0 = 0$. Before generating these data, include the MATLAB command, $rng(1);$. This will ensure that everyone gets the same random numbers and should therefore get the same results. In addition, each time you run your program you will get the same random numbers, which will simplify things for you.

For this question, we adopt the *Naive Minnesota Prior* for Σ. By this is meant that Σ is (i) treated as a known object and (ii) assumed to be equal to $\hat{\Sigma}$, where the elements of $\hat{\Sigma}$ are estimated in a pre-sample. Let the pre-sample be composed of the first 50 of the 100 observations that you generated. In the pre-sample, estimate by OLS a first order scalar autoregression for each variable in y_t. Then, $\hat{\Sigma}$ is the variance covariance matrix of the resulting fitted disturbances. After $\hat{\Sigma}$ is obtained in this way, the Naive Minnesota Prior proceeds by dogmatically[2] assuming $\hat{\Sigma}$ is the true value of Σ. The priors about the

[1] Instructions for reporting results: the basic answer should be written in a wordprocessor. Save the results in the form of a pdf file. Some of the results are produced by MATLAB. The easiest way to integrate the MATLAB results with the material entered produced using the word processor is as follows. Print the MATLAB results to the MATLAB command line and then copy that material into a MATLAB script file in the MATLAB editor. You can type comments into that file explaining the output. Then, create a pdf of the MATLAB script and insert that pdf into the appropriate location of the pdf file generated by the wordprocessor.

[2] The word, dogmatically, is frequently used in the Bayesian literature. It is an adjective for the word 'belief', used to mean that the belief is held without any uncertainty.

elements of A_1 correspond to the usual Minnesota Priors. That is, they are given by:

$$\beta_{ij} \sim N\left(\phi_{ij}, \frac{\hat{\Sigma}_{ii}}{\lambda_1^2 s_j^2}\right), \tag{2}$$

for $i, j = 1, 2$ and $\phi_{ij} = 1$ for $i = j$ and 0 otherwise. The priors are implemented using the usual dummy variables procedure. For this, we require values for the hyperparameters, λ_1, s_j. Initially, specify $\lambda_1 =)0.01)^{1/2}$ (this is smaller than usually specified in practice, but useful for the present pedagogical purposes). The parameters, s_1 and s_2, are equated to the standard deviation, respectively, of the disturbance in the pre-sample of the scalar first order autoregression of the i^{th} variable, $i = 1, 2$ (i.e., $s_i^2 = \hat{\Sigma}_{ii}$, $i = 1, 2$). All analysis should be done on the second set of 50 observations, taking y_{50} as given. The Minnesota priors should be implemented using the dummy strategy discussed in class.

(a) Let $v_1 = (\beta_{11}, \beta_{21})$ and $v_2 = (\beta_{12}, \beta_{22})$. Show that the dummy setup implies that the vector, v_1, is uncorrelated with the vector, v_2. Explain why the prior correlation between β_{11} and β_{21} and between β_{12} and β_{22} depend on $\hat{\Sigma}_{12} = 0$, under the naive Minnesota prior.

(b) Report the dummy matrices, $\bar{X}$ and $\bar{Y}$, for this example. (Careful, the setup here is a little different than in the lecture, because there is no constant term and $A_2 = 0$.)

(c) Verify that the dummy approach to implementing the Minnesota prior 'works'. That is, verify that the prior mean of A_1, $\left(\bar{X}'\bar{X}\right)^{-1}\bar{X}'\bar{Y}$, and the prior variance of, $vec\,(A)$, $\hat{\Sigma} \otimes (\underline{X}'\underline{X})^{-1}$, coincide with the corresponding moments of the Minnesota priors in (2) (careful: note the distinction between A and A_1). Report your results by constructing a table and displaying two columns corresponding to the two ways of computing the prior mean of $vec\,(A)$. Report $\hat{\Sigma}$ and Σ, where the latter denotes the variance-covariance matrix in the data generating mechanism, (1).

(d) Increase the value of λ_1 to $50^{1/2}$. Report the mean and variance of the posterior distribution of the four parameters in A_1 for the two values of λ_1. Also, in a third column report the true values of these parameters (i.e., their values in the data generating mechanism).

(e) Compute the impulse responses to a 'Choleski shock'. That is, compute the lower triangular matrix, C, such that $CC' = \hat{\Sigma}$. Let C_i denote the i^{th} column of C, $i = 1, 2$. The impact effect of shock i is C_i. The period 1 effect is $A_1 C_i$ and the period j effect is $A_1^j C_i$, for $j = 2, 3, .., 10$. Denote the response of variable i to shock j at lag l by

$p(l)_{ij}$, for $i,j = 1,2$ and $l = 0,1,2,...,10$. Graph the posterior mode of $p(l)_{ij}$ in four charts.[3] Also include, in the i,j chart, a 60 percent probability interval for $p(l)_{ij}$ for $l = 0,1,...,10$. To do this, draw A_1 1,000 times from the posterior distribution. This results in 1,000 functions, $p(l)_{ij}$. For each i,j,l sort the $p(l)_{i,j}$'s from the smallest to the largest. The lower and upper boundaries of the 60 percent probability interval are given by the 200^{th} and 800^{th} elements in this sorted list. Do this for $\lambda_1^2 = 0.01$ and $\lambda_1^2 = 500$ and graph the results (the high value of λ_1 is much higher in others parts of this exercise). Also display the true impulse responses, i.e., the ones you get when you use the parameters for the VAR used to generate the data.

- Explain why the impact effects of the shocks are the same for the two values of λ_1. Is this likely to be true when you use the sophisticated Minnesota prior?
- When comparing the posterior modes for the impulse responses with the true impulse responses, there are two dimensions of interest: level and slope. Notice that for the low value of λ_1 (which corresponds closely to OLS), the slopes of $p(l)_{ii}$ with respect to l are lower than the corresponding true slopes. Provide intuition into this (hint: recall the Hurwicz bias).
- Why are the probability intervals around the impulse response functions narrower for the high value of λ_1? Provide intuition.
- In each of the four cases, $p(l)_{i,j}$, $i,j = 1,2$, is the prior pulling you in the right direction relative to the true impulse response function. Explain in intuitive terms.

(f) Compute a mean forecast for y_t for l =1, 5 and 10 periods after the end of the data set. Report the mean forecast under the two circumstances described in the lecture notes: (i) using the stochastic simulation algorithm and (ii) using the non-stochastic simulation algorithm. Do these things for $\lambda_1^2 = 0.01$ and 50. Does it make a difference to the mean forecast whether you do (i) or (ii)? Does the answer depend on λ_1? Provide intuition. (Interesting observation: the amount of shrinkage in the probability interval of the forecast is not the same for different values of l. This is to be expected because the distinct roles played by uncertainty in A_1 versus uncertainty in future ε_t's changes with l.)

(g) Let $f_t(j)$ represent the mode of the forecast of y_{t+j} computed using data available at time t. Estimate the model repeatedly, starting with the sample period, 51 to 60,

[3] An alternative to graphs - which are not easy in this setting - you can report your results in tables generated at the MATLAB command line.

51 to 61, etc., until 51 to 96. This provides a sequence of forecasts,

$$f_{60}(j), f_{61}(j), ..., f_{96}(j).$$

Compute these for $j = 1, 2, 3, 4$. The root mean square error of the forecast, $RMSE$, is defined by:

$$RMSE_j = \left[\frac{1}{96 - 60 + 1} diag\left\{\sum_{t=60}^{96} [f_t(j) - y_{t+j}][f_t(j) - y_{t+j}]'\right\}\right]^{1/2},$$

for $j = 1, 2, 3, 4$. Here, $diag\{Q\}$ means the column vector formed from the diagonal elements of the square matrix, Q.

Compute $RMSE_j$ for $j = 1, 2, 3, 4$ for each of $\lambda_1^2 = 0.01, 1, 5, 10$. Do this again for $a = 0.3$ and $b = 0.3$. Provide an intuitive explanation of the results.

2. Now do the Sophisticated version of the Minnesota Prior, which works with a Wishart prior on Σ (see lecture notes). Redo 2(e). Would your results be different if you redid 2(g)? Assign values to the hyperparameters of the Wishart (i.e., the matrix S^* and the degrees of freedom, ν) as suggested in the lecture notes. Consider $\lambda_1 = 5^{1/2}$ and $50^{1/2}$. Provide intuition for the difference in results.

3. Consider the quarterly dataset "dataVARmedium.mat," which starts in 1959Q1 and ends in 2008:Q8, and includes, in this order, (i) log-real GDP; (ii) log-GDP deflator; (iii) the federal funds rate; (iv) log-real consumption; (v) log-real investment; (vi) log-hours worked; (vii) log-nominal wages. Estimate a 5-lag VAR using **flat priors** and conditioning on the initial 5 observations (note that no prior hyperparameters need to be set in this case). Start with the estimation sample that ranges from 1959Q1 to 1974Q4, and then iterate the same procedure updating the estimation sample, one quarter at a time, until the end of the sample, i.e. 2008Q4. For each estimation sample, set the VAR coefficients at their posterior mode and generate the 1–quarter and 4-quarter ahead forecasts of log-real GDP and log-GDP deflator (i.e., non-stochastic method (ii) in question 2(f)). Denote the out-of-sample forecasts by $f_{i,t}(h)$, where $h = 1$ and 4, for $i = 1$ or 2 (i.e., GDP and deflator). Compute the root mean square forecast error, RMSE, for $i = 1, 2$.

4. Repeat the exercise in question 4 by estimating the VAR with the four sets of dummy priors described in the lecture notes, and using the inverted Wishart prior for the variance-

covariance matrix of the disturbances in the VAR. Set

$$\lambda_2 = 1, \ \lambda_3 = 2, \ \lambda_4 = 5.$$

These settings are taken from Lubik and Schorfheide's 2005 NBER Macro Annual paper. The other hyper parameters are set as in the lecture notes, based on the standard deviation of univariate, lag one autoregressions fit to the 7 variables in the dataset. What value of λ_1 (limit yourself to integer values of λ_1) makes the one-quarter-ahead RMSE on GDP smallest?