MIDTERM EXAM

The exam lasts 90 minutes. Answer all three questions.

(1) Consider this model:

$$x_t = 1.2x_{t-1} - .62x_{t-2} + .2y_{t-1} - .23y_{t-2} + \varepsilon_t - .7\varepsilon_{t-1} - .9\nu_{t-1} \quad (1)$$

$$y_t = 1.4y_{t-1} - .62y_{t-2} + .1x_{t-1} - .08x_{t-2} + \nu_t - .6\nu_{t-1} + .2\varepsilon_{t-1} \quad (2)$$

$$\begin{bmatrix}\varepsilon_t \\ \nu_t\end{bmatrix} \bigg| \{x_s, y_s, s < t\} \sim N(0, I) \,. \quad (3)$$

(a) Show that the model above is exactly equivalent in its implications for the distribution of the x, y stochastic process to the one below:

$$x_t = .5x_{t-1} - .7y_{t-1} + \varepsilon_t \quad (4)$$

$$y_t = .8y_{t-1} + .3x_{t-1} + \nu_t \,, \quad (5)$$

where the assumptions on ε_t and ν_t are as before.

In matrix notation, the first version of the model is

$$(I - AL - BL^2)z_t = (I - CL)\zeta_t \,, \text{ where}$$

$$z_t = \begin{bmatrix}x_t \\ y_t\end{bmatrix} \,, \quad \zeta_t = \begin{bmatrix}\varepsilon_t \\ \nu_t\end{bmatrix} \,,$$

$$A = \begin{bmatrix}1.2 & .2 \\ .1 & 1.4\end{bmatrix} \,, \quad B = \begin{bmatrix}-.62 & -.23 \\ -.08 & -.62\end{bmatrix} \,, \quad C = \begin{bmatrix}.7 & .9 \\ -.2 & .6\end{bmatrix} \,.$$

The second, smaller version of the system is in matrix notation

$$(I - DL)z_t = \zeta_t \,, \text{ where}$$

$$D = \begin{bmatrix}.5 & -.7 \\ .3 & .8\end{bmatrix} \,.$$

The way two systems like this can be equivalent is through root cancellation, that is through the polynomials on either side of the larger system containing a common factor. That is, we should be able to multiply the smaller system by some first-order matrix polynomial in the lag operator that would give us the larger system. But there is

Date: January 12, 2011.

only one first-order polynomial that will convert ζ_t in the smaller system to $(I - CL)\zeta_t$ in the larger system, namely $I - CL$. So we need to check whether

$$(I - CL)(I - DL) = I - AL - BL^2 .$$

The rest is just arithmetic.

Another way to approach this problem is to observe that the smaller pair of equations can be thought of as expressing ε_t and ν_t as functions of current and lagged x_t and y_t. One can then plug in these expressions in place of ε_{t-1} and ν_{t-1} in the larger system, to obtain a new system — which exactly matches the smaller system. This approach leads you through the same arithmetic as you would carry out in the matrix polynomial approach.

(b) Is this model consistent with x_t, y_t being jointly stationary? Explain your answer.

We can work with the first-order version of the model. A first order VAR $y_t = Ay_{t-1} + \zeta_t$ is consistent with stationarity if and only if all the eigenvalues of A are less than one in absolute value. The characteristic equation of A here is $(.5 - \lambda)(.8 - \lambda) + .21$, which has roots

$$\frac{1.3 \pm \sqrt{1.69 - 2.44}}{2} ,$$

i.e. two complex roots whose squared absolute value is $2.44/4 = .61 < 1$, so the model is consistent with stationarity.

(c) Suppose $x_3 = 2$, $x_2 = 1$, $y_3 = -2$, $y_2 = 1$. What is the conditional expectation of x_5 based on all the data for y and x dated 3 and earlier?

$$E_3 \begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = D^2 \begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1.9 \\ -.08 \end{bmatrix} ,$$

so the answer is $E_3[x_5] = 1.9$.

(2) Suppose you have data on $y_t, t = 1, \ldots, T$ and you wish to consider two possible models for its behavior:

$$y_t \mid \{y_{t-s}, s \geq 1\} \sim N(\gamma + \rho y_{t-1}, \sigma^2) \tag{6}$$

$$\log y_t \mid \{y_{t-s}, s \geq 1\} \sim N(\delta + \theta \log y_{t-1}, \nu^2) . \tag{7}$$

Assuming your priors on ρ and θ are the same, make the two parameters a priori independent, and are both flat relative to the likelihood, and that the same holds true for the priors on σ^2 and ν^2 and γ and δ.

How would you go about constructing a posterior odds ratio for the two models based on the observed data? Be as explicit as you would have to be in instructing an undergraduate RA who knows matlab but does not know statistics.

We would like to integrate the posterior for each model to obtain a posterior odds ratio. The first complication to account for is that the second equation defines a conditional distribution for $\log y_t$, not y_t, so to make the models apply to the same observed data, we need to recognize the Jacobian from the transformation from $\log y$ to y. Second, the problem statement does not tell us how to handle initial conditions. So as in a problem with real data, one has to make decisions about whether to assume stationarity and use the model's implied unconditional distribution for the initial conditions. One reasonable answer would be to use the stationarity assumption, in which case the likelihoods for the two models are

$$\exp\left(-\frac{(y_1-\gamma/(1-\rho))^2(1-\rho^2)}{2\sigma^2}\right)(2\pi\sigma^2)^{-T/2}\prod_{t=2}^{T}\exp\left(-\frac{(y_t-\rho y_{t-1}-\gamma)^2}{2\sigma^2}\right) \quad \text{and}$$

$$\exp\left(-\frac{(\log y_1-\delta/(1-\theta))^2(1-\theta^2)}{2\nu^2}\right)(2\pi\nu^2)^{-T/2}\prod_{t=2}^{T}\exp\left(-\frac{(\log y_t-\theta y_{t-1}-\delta)^2}{2\nu^2}\right)/y_t\,.$$

In general,model comparisons via marginal likelihood require proper priors, However, in a case like this one, where the parameters in the two models can reasonably be taken to have the same priors, assuming they are both "flat" relative to the likelihood is OK. The level of the constant pdf one uses to represent a flat prior is arbitrary, yet affects the level of the integrated posterior. But when the parameters in the two models have similar interpretations, assuming that their flat priors have the same height (and thus cancel out) is reasonable.

Following that course, we have to integrate the two expressions above over ρ, γ, σ^2 and θ, δ, ν^2, respectively. If it were not for the initial conditions component of the likelihood, this could be done analytically, but with that component, some kind of MCMC sampling of the posterior is needed. The brute force approach would be just to start up a Markov chain at the OLS estimates and their associated variance estimates, using random walk Metropolis with normally distributed steps. One could also use the normal-inverse-gamma distribution implied by the likelihood without initial conditions as a proposal distribution in an independence Metropolis-Hastings sampling scheme. Note that the denominator y_t's have no effect on the shape of the posterior as a function of the parameters, so the likelihood conditional on initial conditions for that model is still normal-inverse-gamma. The Jacobian terms weights do, though, affect the level of the integrated posterior, and hence are important for the model comparison. I won't go into more detail here about how to do the MCMC, since there are multiple possible approaches, but an A+ answer would have given the kind of detail necessary to instruct the hypothetical undergraduate RA in the question statement.

Of course to use the MCMC sample to get an estimate of the integrated posterior, one has to take an average of an appropriate function of the sampled likelihood. Most commonly, people would use a "modified harmonic mean" approach, i.e. take averages of $g(\theta_j)/\ell(\theta_j)$, where $g(\cdot)$ is a probability density, $\ell(\cdot)$ is the unnormalized posterior density, and θ_j is an MCMC draw of the parameters. These averages will converge to the inverse of the integrated posterior. Usually people use as g a normal density corresponding to the second order expansion of the log likelihood about its maximum, truncated at some χ^2 value (e.g. $2n$, where n is the dimension of θ) and rescaled to integrate to one with the truncation.

If you answered by treating the initial conditions as uninformative about the parameters, you should have seen that the integrated posteriors can be calculated analytically and explained how to do that integration, rather than suggesting MCMC. Because the likelihood is normal-inverse-gamma, the integral involves integrating the normal in γ, ρ or in θ, δ, followed by integration of the resulting gamma marginal distribution for the variance parameter. Since the likelihood is only proportional to these distributions, you have to keep track of the constants of integration.

For the first stage, integrating over ρ and γ, e.g., we need to recognize that the likelihood can be rewritten as

$$(2\pi\sigma^2)^{-(T-1)/2} \exp\left(-\frac{-\sum \hat{u}_t^2 - (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})}{2\sigma^2}\right) ,$$

where $\hat{\beta}$ is the OLS estimator of β, which is the vector ρ, γ. X is the $(T-1) \times 2$ matrix consisting of y_1 to y_{T-1} in the first column and ones in the second. Then integrating out the normal density in β leaves us with

$$(2\pi\sigma^2)^{-(T-3)/2} \left|X'X\right|^{-\frac{1}{2}} \exp\left(-\frac{\sum \hat{u}_t^2}{2\sigma^2}\right) .$$

If we treat the prior as flat in $1/\sigma^2$, this is proportional to a $\text{Gamma}((T-1)/2, \sum \hat{u}_t^2)/2$ distribution, and when we integrate it we get

$$\left(\frac{\sum \hat{u}_t^2}{2}\right)^{-(T-1)/2} 2^{-1} \left|X'X\right|^{-\frac{1}{2}} .$$

For the second model, the calculations are the same, with $\log y_t$ replacing y_t in the definitions of $\hat{u}_t$ and X and θ, δ becoming the β vector. And of course for this model there is also the Jacobian term factor $1/\prod_1^{T-1} y_t$.

(3) Suppose we consider a linearized DSGE model of Canadian and US data, where for each country the observables are a short interest rate, real GDP, the GDP deflator, and the wage rate. There are a small number of structural parameters for the US, and a similar small number of structural parameters for Canada, and we treat them as independent in our prior. The linearized model implies a restricted VAR form, with coefficients nonlinear functions of the structural parameters, and has US GDP and GDP deflator entering the equations for the Canadian part of the model, but not the US interest rate or the US wage rate. No Canadian variables appear in the US equations. The residuals in all equations are i.i.d. across equations and zero mean and they span the same space as the variable innovations.

(a) Does the linearized model imply a Granger causal ordering, with the US first in the ordering?

Yes. The criterion for a Granger ordering is that the system can be arranged so that the coefficients on lagged variables are block triangular. That is true here, if we treat the US as the Granger-causally prior block.

(b) How would you test the Granger causal ordering restriction? Can it be done by considering the US equations alone?

As we discussed in class, tests for Granger orderings can be carried out by looking at the restricted subsystem (that for the variables first in the ordering) alone, under some conditions. The restrictions implied by the ordering are only on the US equations, which is one of the conditions. But if residuals are correlated across equations, an additional condition is that the other equations of the system are unrestricted, so that if we were to take linear combinations of the equations to make the errors in the two blocks uncorrelated, the transformed version of the Canadian block would be neither more nor less restricted than the original Canadian block. Here, the Canadian block originally did not contain all the US variables, so an orthogonalizing transformation, by introducing all the lagged US variables, would change it. Therefore testing the Granger ordering requires estimating the full system. Of course, one could relax the restrictions on the US variables in the Canadian equation and test the Granger ordering restriction in that looser model. However the test would give less sharp results if done that way.

The problem statement included the assertion that "residuals in all equations are i.i.d. across equations". which is unclear at best. What I meant was "i.i.d. over time and independent across equations". If interpreted that way, it implies that the linearized system has the form $A(L;\theta)y_t = \varepsilon_t$, with $\mathrm{Var}(\varepsilon_t)$ diagonal. If we treat the US as the second of two blocks, the restriction is that the second block can be written as $A_{22}(L)y_{2t} = \varepsilon_{2t}$, and it might seem that, since the residuals of this block are assumed independent of those in the first, we could carry out the test using the second block alone. But equations are not directly likelihoods. To carry out the test we need to consider what the test implies for the reduced form of the model, after we have multiplied through by A_0^{-1}. With the restriction imposed, the y_2 block involves US-equation parameters alone. But in the unrestricted model, if there are any Canadian variables allowed contemporaneously in the US block, there will be parameters from the first block in the reduced form equation for y_2 and vice versa. Thus in general the fact that the residuals in the two blocks are uncorrelated does not justify testing the restriction by looking at residuals from the second block alone. The special case where this would be justified is where A_0 is block diagonal, so that multiplying through by A_0^{-1} does not mix up the equations.