Econ. 556b Spring 1999 Chris Sims

Take-Home exam answers

1. (a) I was surprised that 3 out of 4 of those taking the exam missed the point of this question. As we discussed in class and notes, S_R/T formed from the OLS residuals is *not* the MLE of the covariance matrix of residuals for the restricted model. One must either base test statistics on the lower right $(m-q)\times(m-q)$ submatrix of S and S_R, or else use the full system covariance matrix of residuals from a triangularized (or otherwise orthogonalized) system. Since this point was discussed in detail in section 3 of "Notes on Granger Causal Priority and Testing It", posted April 4th, it was surprising to me that only one person got it.

   (b) The flat-prior posterior makes the joint distribution of the VAR parameters approximately Gaussian, around OLS estimates as the means and with $(S/T)\otimes(X'X)^{-1}$ as posterior covariance matrix of residuals. (They are not exactly Gaussian, except conditional on the true covariance matrix. But asymptotically the randomness of Σ about S/T becomes negligible.) This means that the $(m-q)qk$ parameters that are set to zero in the restricted model have an approximately Gaussian posterior with a covariance matrix that can be extracted as a submatrix of $S/T\otimes(X'X)^{-1}$. If we use γ to denote the parameters set to zero in the restricted model and $\hat{\gamma}$ to denote their OLS estimates in the unrestricted model, then the classical LR test statistic has the form

$$\hat{\gamma}'\Sigma_\gamma^{-1}\hat{\gamma}\,, \tag{A1}$$

   where Σ_γ is the usual estimate of the asymptotic covariance matrix of $\hat{\gamma}$ conditional on true values, and also the usual approximation to the posterior covariance matrix of γ. This is a number, not (once we have seen the data) a random variable. The related random variable, which has an approximate $\chi^2((m-1)qk)$ distribution quantity conditional on the data under a flat prior, is

$$(\gamma-\hat{\gamma}')\Sigma_\gamma^{-1}(\gamma-\hat{\gamma})\,. \tag{A2}$$

   Thus when we are told the value of the LR statistic, we can look it up in the χ^2 table to find out the probability of one particular set. Using θ to denote the whole vector of elements of $\{B_{ij}(s)\}$ and $\gamma(\theta)$ to denote the restricted subvector extracted from θ, the set whose probability we obtain from the table is

$$\left\{\theta \mid (\hat{\gamma}-\gamma(\theta))'\Sigma_\gamma^{-1}(\hat{\gamma}-\gamma(\theta)) < \hat{\gamma}'\Sigma_\gamma^{-1}\hat{\gamma}\right\}\,. \tag{A3}$$

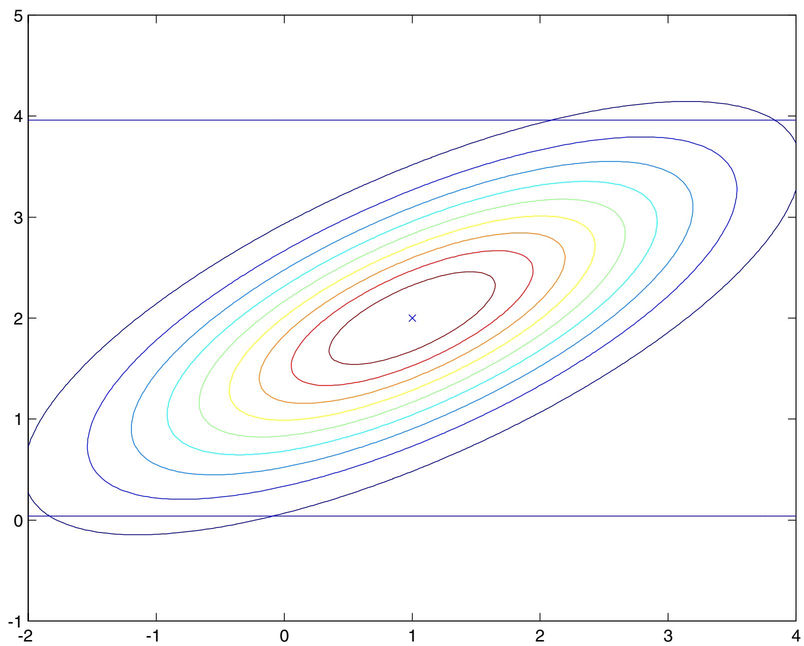
   The probability of this set will generally not be exactly .05, since the statistic will not be exactly at the .05 significance level. To find a set of .05 probability,

we would extend or shrink the set by replacing the right-hand side of the inequality in the set definition (A3) with the exact .05 level of the test.
To understand what this means (and what its limitations might be), consider the simpler case of a regression model with a two-dimensional parameter vector β, in which our estimates imply

$$\beta \sim N\left(\begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\right) . \tag{A4}$$

If we consider a constraint on β_2, we construct the region from a $\chi^2(1)$ statistic, and the region, centered on $\hat{\beta}$, that has probability .05 is that above and below the horizontal lines in the figure below.

Likelihood contours and .05 prob. region from distribution of β_1



(c) Here we are using the theory laid out in the handout on testing linear restrictions. The relevant formula, giving the log of the posterior odds against the restricted space based on the Gaussian approximation, is (10) in those notes:

$$\log\left(\frac{p(X|\hat{\theta})}{q(X|\hat{\phi})}\right) + \frac{m-n}{2}\log(2\pi) - \frac{1}{2}\log\left(\frac{|\Sigma_\theta|}{|\Sigma_\phi|}\right) . \tag{A5}$$

The first term in this expression, the log of the likelihood ratio, is $\frac{1}{2}$ times the χ^2 statistic that we have been discussing.

(d) The Bayesian asymptotics do not depend at all on whether unit roots are present. The classical asymptotics may change drastically if unit roots are present, but do not necessarily change. Since we did not discuss classical unit root theory in the course, you were not required to be able to give a more detailed answer. Generally, if there are two blocks of variables Y and X, and if we test (as here) a restriction that $X \,\mathrm{GCP}\, Y$, special unit-root distribution theory does not come into play if X and Y are cointegrated, so that there is a matrix M of cointegrating vectors such that $Y - MX$ is stationary.

(e) The accuracy of the Bayesian asymptotics can be checked by exploring the shape of the actual likelihood to see if it matches the approximation. For example, the height, relative to the height at the maximum, of the likelihood along the ellipsoid in the parameter space that forms the boundary of the flat-prior 95% probability region under the Gaussian approximation can be checked to see it matches what is predicted by the approximation. It is not even a meaningful question whether classical asymptotics are accurate in a particular sample; they have nothing to say about particular samples.

2. The model specified in this problem, I am sad to say, was internally inconsistent. Anyone who figured this out would have received full credit, but nobody did. Though I could detect some differences in levels of understanding in the answers to this question, I did not award points for it in the grading. I should have left $y(t)$ out of the model's second equation ((5) on the exam), which would have created the kind of simple likelihood I meant for you to consider. Below I first describe the inconsistency, then describe a small change in the model's specification that removes the inconsistency, then present an answer to the question for an altered model.

The inconsistency: substituting (3), the equation for y, into (5), the equation for S, produces

$$S(t) = -\gamma(S(t)) - \varepsilon(t) + \rho S(t-1) + \nu(t)\,, \tag{A6}$$

where $\gamma(S(t))$ is defined by

$$\gamma(S(t)) = \begin{cases} \gamma_0 & \text{for } S(t) < 0 \\ \gamma_1 & \text{for } S(t) \geq 0 \end{cases}. \tag{A7}$$

If we use the notation $F(S(t)) = S(t) + \gamma(S(t))$, then (A6) asserts that, conditional on $S(t-1)$,

$$F(S(t)) \sim N(\rho S(t-1), \sigma_\varepsilon^2 + \sigma_\nu^2)\,. \tag{A8}$$

But F is a function that maps the real line $\mathbb{R}$ into $\mathbb{R} - [\gamma_0, \gamma_1)$. That is, $F(S(t))$ can never take on any value in the half-open interval $[\gamma_0, \gamma_1)$. When $S(t) < 0$, $F(S(t)) < \gamma_0$, while when $S(t) \geq 0$, $F(S(t)) \geq \gamma_1$.[1] This means that it is in fact impossible for $F(S(t))$ to be normally distributed, contradicting (A8).

Fixups: To make the model coherent while retaining y in (5), we have to recognize that it only works if $S(t) = 0$ has positive probability and if in this case

[1]This discussion assumes that $\gamma_1 > \gamma_0$, as fits the interpretation of recession and normal rates of growth. if instead $\gamma_1 < \gamma_0$, the model, instead of being internally inconsistent, fails to define a determinate distribution for the observed data.

equation (5) on the exam determines $y(t)$, replacing for this case the original y equation, (3) from the exam. This results in a well-defined model, for which MCC sampling is possible. It is a rather complicated model, though, and since no one who wrote the exam actually used this model, I won't go through here how to do MCC sampling for it. To show the kind of setup I meant to be confronting you with, consider a model in which $y(t)$ is simply deleted from equation (5) on the exam. This results in an easy-to-handle model.

Sampling: For the easy-to-handle model with y gone from (5), sampling from the likelihood can proceed as follows. With the S sequence fixed, (3) and (5) are simply a pair of regression equations with independent disturbance terms. We can apply OLS to them separately. To make this work we must suppose that the distribution of the initial condition $S(0)$ does not depend on any of the parameters of the model. Separate OLS on equation (3) amounts to estimating γ_0 as the mean of y for recession ($S(t) < 0$) periods and γ_1 as the mean of y over the rest of the sample. The joint posterior on the γ's, ρ, and the variances of ε and ν is then in the usual Normal-Inverse-Gamma form and is easy to sample from. So the first phase of the sampling algorithm is to draw from this simple conditional distribution of all the parameters given the data and the S sequence. The second phase is to draw from the conditional distribution of the S sequence given the parameters and the data. This conditional distribution is non-standard, so drawing from it requires its own Metropolis-Hastings component. The joint pdf for the y's and S's itself is easy enough to write down:

$$\begin{aligned}\log p\left(\{y(t)\}_{t=1}^{T}, \{S(t)\}_{t=0}^{T} \,\middle|\, \gamma_0, \gamma_1, \sigma_\varepsilon^2, \sigma_\nu^2\right) \\ = q(S(0)) - T\log\sigma_\varepsilon - \frac{1}{2}\sum_{t=1}^{T}\frac{(y(t)-\gamma(S(t)))^2}{\sigma_\varepsilon^2} \\ - T\log\sigma_\nu - \frac{1}{2}\sum_{t=1}^{T}\frac{(S(t)-\rho S(t-1))^2}{\sigma_\nu^2}\,, \qquad \text{(A9)}\end{aligned}$$

where $\gamma(S(t))$ is the function defined in (A7). What makes the distribution nonstandard is the appearance of $S(t)$ as the argument of the γ function in the first term. A reasonable way to implement Metropolis sampling of S would be to proceed sequentially through $t = 0, \ldots, T$. For $t = 0$ the draw could be from q, the marginal pdf for $S(0)$, and for subsequent $S(t)$'s it could be either from $N(\rho S(t-1), \sigma_\nu^2)$ (the conditional pdf for $S(t)|\,S(t-1)$, ignoring the information in y) or from $N(\rho\cdot(S(t+1)+S(t))/(1+\rho^2), \sigma_\nu^2/(1+\rho^2))$, the conditional distribution of $S(t)|\,S(t+1), S(t-1)$. The acceptance algorithm then follows the usual Metropolis-Hastings rule, using the actual likelihood (A9) to form the weights.

The overall algorithm, then, is Gibbs sampling, alternating between drawing from the distribution of the parameters given S and y and drawing from the distribution of S given the parameters and y. The latter part of the algorithm requires using a Metropolis-Hastings accept/reject rule.