Midterm Exam*

(1) Is it true that in large enough samples, with i.i.d. observations, the Schwarz criterion eventually becomes a good approximation to the log of the posterior probability weights on models being compared? If so, sketch the argument for why this is true, mentioning the main regularity conditions that have to be assumed. If not, discuss whether there is any Bayesian argument that the Schwarz criterion is "asymptotically valid", and explain why or why not.

   The only sense in which this is true is that, with i.i.d. observations plus other regularity conditions, the Schwarz criterion converges to $\pm\infty$ along with the odds ratio. It is not true that the Schwarz criterion gets close to the log of the odds ratio in any ordinary sense. Only one or two people made this point correctly.

   The regularity conditions, besides the i.i.d. assumption, are that the prior pdf should be continuous in a neighborhood of the true value of the parameter and that the likelihood is regular enough to make the Taylor expansion of its log asymptotically accurate. Among other things, we certainly require that

$$E\left[\frac{\partial^2 \log p(X_t \mid \beta)}{\partial\beta\partial\beta'}\right]$$

   exist and be finite, so that the i.i.d. assumption allows us to apply the law of large numbers.

(2) Suppose $\varepsilon(t)$ is an i.i.d. sequence of two-dimensional $N(0,\Sigma)$ random variables and that, for all integer t,

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \varepsilon_1(t) + .3\varepsilon_1(t-1) + .4\varepsilon_2(t-1) \\ \varepsilon_2(t) - .8\varepsilon_1(t-1) + 1.5\varepsilon_2(t-1) \end{bmatrix}.$$

Is x covariance stationary? Is $\varepsilon(t)$ the innovation in $x(t)$? Justify your answers.

   Any process that is represented as a finite order moving average of i.i.d. variables with finite second moments is covariance-stationary. So this process is stationary. Whether it is fundamental depends on whether we can be sure that in the representation

$$x(t) = A(L)\varepsilon(t)$$

The roots of the polynomial $|A(z)|$ all lie outside the unit circle, which is here equivalent to the eigenvalues of

$$A = \begin{bmatrix} .3 & .4 \\ -.8 & 1.5 \end{bmatrix}$$

all lying inside the unit circle. The quick check of finding the determinant of A did not give a definite answer here, since the determinant is less than one. The eigenvalues

themselves are 1.1 and .7, which of course means this is not the fundamental representation and ε is not the innovation vector.

(3) Suppose our model is

$$y(t) = \gamma + \rho_1 y(t-1) + \rho_2 y(t-2) + \varepsilon(t),$$

with $\varepsilon(t)$ as usual assumed independent of all past y's and $\varepsilon(t) \sim N(0, \sigma^2)$.

(a) Explain how to construct one or more dummy observations that express a prior belief that $\rho_1 + \rho_2$ is approximately one, with a standard error of $.5\sigma$ on this prior mean for the sum of coefficients.

Rewriting the regression equation as

$$y(t) = X(t)\beta + \varepsilon(t) \quad \text{with}$$
$$X(t) = \begin{bmatrix} 1 & y(t-1) & y(t-2) \end{bmatrix},$$

we can construct the dummy observation at "t^*" as

$$y(t^*) = 2$$
$$X(t^*) = \begin{bmatrix} [0 & 2 & 2] . \end{bmatrix}$$

This dummy observation reads

$$2 = 2\rho_1 + 2\rho_2 + \varepsilon(t^*),$$

which can be rearranged as

$$\rho_1 + \rho_2 = 1 + \frac{\varepsilon(t^*)}{2},$$

which makes it clear that we the observation corresponds to a prior mean of 1 and a prior standard error of $\sigma/2$.

(b) It is common practice to use dummy observations in which the scale of the dummy observations is determined by taking an average of y values over the initial conditions. Is this different from what you have proposed in part (3a)? If not, explain; if so, what might be an argument for proceeding this way rather than as you suggest in part (3a)?

Here, instead of weighting by 2 as in the previous part, we weight by $\bar{y}$, an average of initial observations, making the rearranged dummy observation read

$$\rho_1 + \rho_2 = 1 + \frac{\varepsilon(t^*)}{\bar{y}}.$$

This makes the standard deviation of the prior $\sigma/\bar{y}$. The argument for proceeding this way is that usually we have no clear idea of how directly to choose the precision for our prior beliefs about a sum of coefficients. Weighting the observation by $\bar{y}$ makes the prior assert that the our uncertainty about the ρ values contributes about as much to forecast uncertainty, when lagged y's are at $\bar{y}$, as does equation error itself. This seems like a reasonable order of magnitude.

(4) Is it true that if in an autoregressive time series model y does not Granger-cause x, then there is a regression equation relating y to current and past x, possibly including other variables but not including lagged y, in which x is strictly exogenous? Does the answer depend on whether the time series model consists of y and x alone or instead includes other variables as well? Explain your answer.

What is true is that if x is part of a block of variables that is Granger-causally-prior to y, ($x\ GCP\ y$), then there is such a regression equation. However, in a multivariate model $\sim y\ GC\ x$ does not imply $x\ GCP\ y$, because of the intransitivity of the GC relation. In a bivariate model the implication does hold, however. Stating these relationships clearly earned substantial credit. An excellent answer added some detail, defining GCP in the context of an AR model and showing the difference between it and $\sim GC$. A few people were confused by the fact that our in-class discussion only went through a derivation of the fact that $x\ GCP\ y$ implies that there is a regression of y on current and lagged x and other variables z in its GCP block *and lagged y's*, in which x is strictly exogenous. However, in such an equation the polynomial on the lag operator applying to y can be inverted to yield the required equation without lagged y's on the right. This is asserted, but not proved, in the notes on Granger Causal Priority on the web page for last year's version of the course (they are labeled "10/26 notes").

(5) Consider the model

$$y_1(t) = y_1(t-1)^\theta + \gamma y_2(t-1) + \varepsilon_1(t)$$
$$y_2(t) = \delta y_1(t-1) + \phi y_2(t-1) + \varepsilon_2(t),$$

where the $\varepsilon(t)$ vector is i.i.d. $N(0,\Sigma)$ conditional on all past $y's$. Are we justified in using an ordinary t-statistic on the OLS estimate of δ to test whether $\delta=0$, or does this fail to be equivalent to a likelihood ratio test? Explain your answer.

You got some credit for noticing that this is a question of whether here, as in the case of testing for GCP in a purely linear AR, the likelihood factors appropriately to allow single-equation estimation to correspond to ML. The factorization does not work here, unless we know a priori that $\theta=1$, so the system is after all linear. The argument in the linear case is that we can find the coefficient λ in the projection of $\varepsilon_1(t)$ on $\varepsilon_2(t)$ and that this allows us t form a revised first equation:

$$y_1(t) = \lambda y_2(t) + y_1(t-1)^\theta - \lambda\delta y_1(t-1) + (\gamma-\lambda\phi)y_2(t-1) + \varepsilon_1(t) - \delta\varepsilon_2(t).$$

In the linear case, the terms in $y(t-1)$ merge, so that the range of right-hand-side functions of $y_1(t-1)$ and $y_2(t-1)$ achievable by varying the parameters in the modified first equation is unaffected by restrictions on δ or ϕ. But in this linear case, the right-hand side has a linear component that disappears if we force $\delta=0$. Thus the restrictions on δ affect the likelihood both through the "OLS" piece of the likelihood we get from the second equation alone and from the component related to the revised first equation. There is no way to get a likelihood-ratio test by using the second equation in isolation.