Econ. 513, Time Series Econometrics Fall 1999 Chris Sims

Take-Home Final Exam*

(I) Suppose

$$y(t) = \begin{bmatrix} 0.3 & -0.2 \\ -0.3 & 0.9 \end{bmatrix} y(t-1) + \varepsilon(t)$$

and $\varepsilon(t) \mid \{y(s), s < t\}] \sim N(0, \Omega)$, for all t.

(a) Prove that y is stationary.

(b) Prove that, if $\{A_s, s = 0, \dots, \infty\}$ are the coefficients in the fundamental MAR representation of y in terms of its innovations, the following sequence $\{A_s^*\}$ of three matrices are *not* the first three elements of the A_s sequence:

$$A_0^* = I \quad A_1^* = \begin{bmatrix} 0.3 & -0.2 \\ -0.3 & 0.9 \end{bmatrix} \quad A_2^* = \begin{bmatrix} 0.15 & -0.24 \\ -0.30 & 0.87 \end{bmatrix}$$

(A:a) This just asks you to apply Proposition 4.ii from the "Innovations" notes. The roots of $|I - Az|$ are all outside the unit circle iff the eigenvalues of A are all inside the unit circle. The eigenvalues of A are .2127 and .9873, so y is stationary. (Actually an implicit assumption here and in the proposition is that y and ε are thought of as defined for all t, $-\infty < t < \infty$. Otherwise the distribution of y can depend on initial conditions, even if the eigenvalues of A are all less than one in absolute value.)

(A:b) Since we have stationarity and an AR representation, the MA coefficients must be the inverse in positive powers of L of $I - AL$. So if the MA polynomial is $B(L)$, we must have $(I - AL)B(L) = I$. But if we calculate we find

$$\begin{array}{ccccccccc} I & + & A_1^* L & + & A_2^* L^2 & + & \dots \\ -(& & AL & + & AA_1^* L^2 & + & \dots \\ \hline I & + & 0L & + & \begin{bmatrix} 0 & 0 \\ -.06 & 0 \end{bmatrix} L^2 & + & \dots . \end{array}$$

Since this isn't I, the A^* sequence is not the MAR.

(II) Consider the model

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

in which $\varepsilon(t) \mid \{y(t-s),\ s > 0\} \sim N(0, \sigma^2)$. Applying least squares (i.e. maximum likelihood conditional on initial observations) to the model, we obtain the estimates $\hat{\rho}_1 = 1.9$, $\hat{\rho}_2 = -.9024$, $\hat{c} = 3$, $\hat{\sigma}^2 = .2$. It can be shown that if these estimates were the parameters generating the y process, we would have, unconditionally, $\mathrm{Var}(y(t)) = 427.1820$, $\mathrm{Cov}(y(t), y(t-1)) = 426.6431$.

(a) Suppose the initial conditions that we have treated as fixed in the OLS estimation are $y(-1) = 1.7$, $y(0) = 1.1$. Show that, according to the estimated model, these initial conditions are implausible if they are thought to have been generated from the same model as the later observations.

(b) Construct a dummy observation that implies that if lagged y's are set at the sample mean of the initial observations, the model should predict that current y is also near the mean of the initial observations, with a standard deviation on this dummy observation of $.2\sigma^2$, where this σ^2 is the variance of the equation disturbances as defined above. You are expected to give numerical values for all elements of the dummy observation.

(c) Explain why adding this dummy observation to the sample is likely to reduce the extent to which the estimated parameters imply the initial conditions are implausible.

(A:a) The model implies an unconditional mean for y of $\bar{y} = c/(1 - \rho_1 - \rho_2) = 1250$. According to the covariance matrix the problem gives you, the standard deviation of y should be $\sqrt{427.1820} = 20.6684$. Thus both initial values of y are more than 60, standard deviations away from their means, which is extremely unlikely if they were indeed drawn from the model's implied unconditional distribution. We could make the same point with the two initial values jointly by forming a $\chi^2(2)$ statistic as $(y - \bar{y})'\Sigma^{-1}(y - \bar{y}) = 3652$, which is far out in the tail of the $\chi^2(2)$.

(A:b) If the "$.2\sigma^2$" above is corrected to $.2\sigma$, The appropriate dummy observation is

$$\check{y} = c + \rho_1 \check{y} + \rho_2 \check{y} + .2\check{\varepsilon} \ ,$$

or as a row of data for the data matrix, dividing through by .2 to match the size of the error in the dummy observation to that in the real data, we get

$$\begin{bmatrix} 5\check{y} & 5\check{y} & 5\check{y} & 5 \end{bmatrix} \ ,$$

where we have taken the t'th row of the data matrix to be

$$\begin{bmatrix} y(t) & y(t-1) & y(t-2) & 1 \end{bmatrix}$$

and $\check{y} = (1.7 + 1.1)/2$ is the average of the initial conditions. If you did not get the (late) correction of the typo in this question and tried to use $.2\sigma^2$ as the standard deviation of the dummy observation, you would have the problem that the appropriate weight on the dummy observation depends on the unknown σ^2. An approximate solution to this problem is to use $\hat{\sigma}^2$ as if it were the true value, in which case the appropriate weight is $\hat{\sigma}/(.2\hat{\sigma}^2) = 1/(.2\sqrt{.2}) = 11.18$.

(A:c) The dummy observation has a residual of -2.966, or about 6.6 standard deviations, if evaluated at the OLS estimates. Incorporating the dummy observation will therefore have a substantial effect on the estimates. To make the residual zero would require that the parameters satisfy

$$\check{y} = c/(1 - \rho_1 - \rho_2)\,, \tag{1}$$

i.e. that the parameters imply the unconditional mean matches $\check{y}$. However, since the level of "implausibility" will depend on the implied unconditional variance as well as on (1) being close to true, it is not necessarily true that the initial conditions look more plausible once the dummy observation is added; this is only a tendency.

(III) Consider the model

$$\underset{2\times 1}{y(t)} = Ay(t-1) + \varepsilon(t)\,,$$

in which as usual we assume $\varepsilon(t)$ to be the innovation in y. We assume to keep the problem simple $\varepsilon(t) \,|\, \{y(s),\, s < t\} \sim N(0, I)$, with no uncertainty about this covariance matrix. We have OLS estimates conditional on initial observations in which we emerge with

$$\hat{A} = \begin{bmatrix} .7 & .2 \\ .2 & .7 \end{bmatrix}\,, \quad \frac{\hat{u}'\hat{u}}{T} = I\,,$$

where the $T \times 2$ $\hat{u}$ matrix is the OLS residuals. The sample second-moment matrix of the right-hand side variables (what we usually label "$X'X$") is

$$\begin{bmatrix} 30 & 20 \\ 20 & 30 \end{bmatrix}\,.$$

When the second equation alone is estimated with a_{21} constrained to zero, it produces $\hat{a}_{22} = .9$ and $\hat{u}_2'\hat{u}_2/T = 1.2$. We have a prior pdf over A that is Gaussian with mean $\bar{A} = I$ and covariance matrix

$$I \otimes \begin{bmatrix} .5 & -.4 \\ -.4 & .5 \end{bmatrix}$$

conditionally on the parameter space with a_{22} unconstrained. Conditionally on the hypothesis that $a_{21} = 0$, our prior mean is the same and our prior covariance matrix is

$$\begin{bmatrix} .5 & -.4 & 0 \\ -.4 & .5 & 0 \\ 0 & 0 & .25 \end{bmatrix} ,$$

where the upper left 2×2 block corresponds to the coefficients of the first equation and the lower right to the single unconstrained coefficient on the second equation. Our prior puts equal probability on the constrained and unconstrained versions of the model. What is the posterior probability that the constraint is true? [Note: This problem involves a lot of algebra and numerical calculation. Be sure you have explained how you are setting up the calculation and why, so that if you run out of time or make algebra or arithmetic mistakes you can get partial credit.]

For the unconstrained parameter space, the log of likelihood times prior is

$$-(T+2)\log(2\pi) - \log|\Omega| \\ -\frac{1}{2}\left(\sum_{i=1}^{2}\{\hat{u}_i'\hat{u}_i + (\beta_i - \hat{\beta}_i)'X'X(\beta_i - \hat{\beta}_i) + (\beta_i - \bar{\beta}_i)'\Omega^{-1}(\beta_i - \bar{\beta}_i)\}\right) , \quad (2)$$

where β_i is the ith row of A, arranged as a column, $\hat{\beta}_i$ is the OLS estimate of β_i, $\bar{\beta}_i$ is the prior mean of β_i, and Ω is the prior covariance matrix of β_i (which happens to be the same for both values of i.) To evaluate the posterior probability of the whole sample space, we have to integrate β_1 and β_2 out of this pdf. The pdf is Gaussian as a function of the β's, since its log is quadratic in them. So the hard part of the integration task is an exercise in completing the square. Let

$$\beta_i^* = \left(X'X + \Omega^{-1}\right)^{-1}\left(X'X\hat{\beta}_i + \Omega^{-1}\bar{\beta}_i\right)$$
$$\Omega^* = \left(X'X + \Omega^{-1}\right)^{-1} .$$

Then we can rewrite (2) as

$$-(T+2)\log(2\pi) - \log|\Omega| - \frac{1}{2}\left(\sum_{i=1}^{2}\{\hat{u}_i'\hat{u}_i + (\beta_i - \beta_i^*)'(\Omega^*)^{-1}(\beta_i - \beta_i^*) \right. \\ \left. + \bar{\beta}_i'\Omega^{-1}\bar{\beta}_i + \hat{\beta}_i'X'X\hat{\beta}_i - \beta_i^{*\prime}(\Omega^*)^{-1}\beta_i^*\}\right) , \quad (3)$$

To integrate out the β_i's is now easy, using our knowledge of the multivariate normal pdf. The result is

$$-T\log(2\pi)+\log|\Omega^*|-\log|\Omega| \\ -\frac{1}{2}\left(\sum_{i=1}^{2}\{\hat{u}_i'\hat{u}_i+\bar{\beta}_i'\Omega^{-1}\bar{\beta}_i+\hat{\beta}_i'X'X\hat{\beta}_i-\beta_i^{*\prime}(\Omega^*)^{-1}\beta_i^*\}\right) . \quad (4)$$

For the restricted parameter space, the calculations are the same in form. Because it is assumed in the problem statement that we know a priori that the disturbances of the two equations are independent, and because the priors on the parameters of the two equations are assumed independent, we end up with no cross terms in the likelihood and the MLE for the first equation's parameters is unchanged by the restriction on the second equation. The difference in the logs of the integrated pdf's for the two parameter spaces therefore emerges as

$$\frac{1}{2}\left(2\log|\Omega^*|-\log|\Omega^{**}|+\log\left(\frac{\omega_2}{\omega_2^{**}}\right)-\log|\Omega|\right) \\ -\frac{1}{2}\left(\hat{u}_2'\hat{u}_2-\hat{\hat{u}}_2'\hat{\hat{u}}_2+\bar{\beta}_2'\Omega^{-1}\bar{\beta}_2-\frac{\bar{a}_{22}^2}{\omega_2}+\hat{\beta}_2'X'X\hat{\beta}_2 \right. \\ \left. -30\hat{a}_{22}^2-\beta_2^{*\prime}(\Omega^*)^{-1}\beta_2^*+\frac{a_{22}^{**}}{\omega_2^*}\right) , \quad (5)$$

where $\hat{\hat{u}}_2$ is the residual vector from an OLS estimate of the constrained equation 2, $\omega_2 = .25$ is prior variance on a_{22} in the constrained space, 30 is the lower right element of $X'X$, and a_{22}^{**} is the posterior mean for a_{22} in the constrained space. All the elements of this expression were given to you numerically in the problem statement except, sad to say, T, which is needed to convert $\hat{u}'\hat{u}/T$ to $\hat{u}'\hat{u}$. So getting the answer into the form above was enough to get full credit. You could also have saved some writing out of algebra by assuming a value for T, say $T = 100$.

(IV) (a) If $y(t) = 10y(t-1)+\varepsilon(t)$, with ε i.i.d. across time and distributed as $N(0,1)$, and if $y(t)$ is stationary, find an expression for the innovation in y as a function of current and past y and find the variance of the innovation.

(b) If the estimated time path of coefficients from the Kalman smoother is in fact much less smooth then the estimated time path of coefficients, using the same data, based on the Kalman filter, does this mean you have made a mistake in calculation? Does it mean the model is wrong?

(c) If the spectral density of y shows a single very sharp peak at frequency π, then y has strong seasonal fluctuations, and we do not have to know whether the time unit is a month, a quarter, a day, etc. in order to assert this. Is this true or false? Why?

(A:a) The equation can be rearranged as $y(t) = .1y(t+1) - .1\varepsilon(t+1)$ and then "solved forward" as $y(t) = -.1L^{-1} \cdot (1 - .1L^{-1})^{-1}\varepsilon(t)$, which implies an autocovariance function equal to the coefficients of the polynomial

$$R_y(L) = \frac{.01}{(1 - .1L)(1 - .1L^{-1})} .$$

This is easily recognized as the ACF of a simple first-order AR with representation

$$y(t) = .1y(t-1) + \eta(t)$$

in which $\operatorname{Var}\eta(t) = .01 \cdot \operatorname{Var}\varepsilon(t) = .01$.

(A:b) It is true that the Kalman filter estimate $\hat{\beta}_t = E_t\beta_t = E_t\big[[E_T[\beta_t]\big] = E_t[\hat{\hat{\beta}}_t]$, where we are using $\hat{\beta}_t$ for the Kalman filter estimate of β_t and $\hat{\hat{\beta}}_t$ for the Kalman smoother estimate. So the Kalman smoother estimate can be represented as the Kalman filter estimate plus an orthogonal error, and thus has *larger* variance than the Kalman filter estimate. But this larger variance would be observed as a sampling variance only if we made repeated draws from our prior on β_0 and then on the disturbances in the plant and observation equations. In a single sample, the time path of the smoother is often, but need not be "smoother" than the time path of the filter. Of course there is a limiting case, that in which the plant equation implies constant β, in which the smoother certainly always yields a "smoother" time path — in fact a constant time path. But one can construct examples where the opposite is likely. For example, if the plant equation is $\beta_t = -\beta_{t-1}$, i.e. the β's oscillate in sign, non-randomly, and if the prior mean on β_0 is zero, then it is very likely that the Kalman filter will show only small oscillations at first, until the data pull the estimated absolute value of β away from zero. The smoothed estimates will always be a perfect sawtooth pattern, constant in absolute value. Another special case, which some who took the exam had in mind, is that where the plant equation is $\beta_{t+1} = \beta_t + \nu_{t+1}$, so β is a martingale. [The idea for the argument that follows comes in good part from one of the exam answers.] In this case

$$\hat{\hat{\beta}}_{t+1} - \hat{\beta}_t = (I + \Omega\Sigma_t^{-1})(\hat{\hat{\beta}}_t - \hat{\beta}_t) ,$$

where Ω is the variance of $\nu(t)$ and Σ_t is the variance of $\hat{\beta}_t$. (The expression holds only if $|\Sigma_t| \neq 0$, obviously.) This in turn implies

$$\hat{\hat{\beta}}_{t+1} - \hat{\hat{\beta}}_t = \Omega \Sigma_t^{-1} (\hat{\hat{\beta}}_t - \hat{\beta}_t) \, .$$

This means that, for the scalar case, the filtered estimate lies below the smoothed estimate whenever the smoothed estimate is rising, and above it whenever the smoothed estimate is falling. Therefore the absolute change in the filtered estimates from the beginning to the end of the sample — $|\hat{\beta}_T - \hat{\beta}_0|$ — must exceed the absolute change in the smoothed estimates. However, it is nonetheless possible for the changes in the filtered estimates, period by period, to be smaller than the changes in the smoothed estimates over long stretches of time.

(A:c) The seasonal frequencies are $2\pi j/S$, $j = 1, \ldots, S$, where S is is the number of time units in a year. Any process whose spectral density shows sharp peaks at any or all of these frequencies will show seasonal oscillations. So long as S is an even number, as it is for monthly, quarterly, and weekly data, π is the seasonal frequency corresponding to $j = S/2$, and reflects the presence of oscillations with period 2. With true daily data, where weekends are treated like other days, $S = 365$ and π is not a seasonal frequency. With workday data, the convention is that there are 52 weeks and therefore 260 workdays, which is even. (However, once we go below the monthly time unit, the need to account for leap year and for the fact that there is not an integer number of weeks in a year raise important difficulties for interpreting and modeling seasonals).