FINAL EXAM

The exam consists of five questions containing a total of 140 points. You are to answer all 5 questions. The points for each question are shown in parenthesis after the question number. Allocate your 180 minutes accordingly.

(1) (30) Here's a VAR system, with the usual assumptions making ε_t the innovation in y_t.

$$y_t = \begin{bmatrix} 1.67 & .14 \\ -.24 & 1.62 \end{bmatrix} y_{t-1} - \begin{bmatrix} -.7 & -.2 \\ .2 & -.7 \end{bmatrix} y_{t-2} + \varepsilon_t \,. \tag{1}$$

(a) Show that this system is cointegrated, with a cointegrating vector [1, 2].

The first step is to show that there is a unit root. The most straightforward way, if we think of the system as $B(L)y_t = \varepsilon_t$, is to replace L with 1 and see if we get that $B(1)$ is singular:

$$|B(1)| = \left| I - \begin{bmatrix} 1.67 & .14 \\ -.24 & 1.62 \end{bmatrix} + \begin{bmatrix} -.7 & -.2 \\ .2 & -.7 \end{bmatrix} \right| = \begin{vmatrix} .03 & .06 \\ .04 & .08 \end{vmatrix} = 0 \,.$$

So there is in fact a unit root. Also, since $B(1)$ is rank 1 not rank 0, there is a single unit root and the system must show cointegration. It is easy to see that the right eigenvector of $B(1)$ corresponding to the unit root (the eigenvector associated with $B(1)$'s 0 eigenvalue) is proportional to (2,-1). All vectors orthogonal to this one will be cointegrating vectors. (This doesn't follow immediately from the notes I gave out, which worked entirely in terms of the system matrix for the stacked system. That is why I just gave you the cointegrating vector, (1,2).) If we let $z_t = [1 \quad 2]y_t$, we can multiply the system on the left by $[1 \quad 2]$ to obtain

$$z_t = [1.19 \quad 3.28]y_{t-1} - [.3 \quad 1.6]y_{t-2} + \xi_t \,,$$

where $\xi_t = [1 \quad 2]\varepsilon_t$. To show that z_t is stationary, we want to express the right-hand side in terms of lagged z and differences of y (which are stationary if there are no roots larger than one). This is simply done by writing

$$z_t = [1.19 \quad 2.38]y_{t-1} + y_{t-1} - [.3 \quad .6]y_{t-2} - y_{t-2} + \xi_t = 1.19z_{t-1} - .3z_{t-2} + \Delta y_t + \xi_t \,.$$

The lag operator on z in this equation, $1 - 1.19L + .3L^2$, is invertible. (Its roots are 1/.83 and 1/.36.) So the equation allows us to express z_t as a convergent linear combination of current and past values of Δy_t and ξ_t, which is stationary so long as Δy_t is stationary. This much would be a full-credit answer. A more complete answer would also verify that there are no roots larger than one to the system. This is straightforward with a computer, but more computation than is reasonable for an in-class exam without computer.

Date: January 27, 2009.

(b) Rewrite the system in VECM format.

This is just re-expressing $B(L)y_t = \varepsilon_t$ as $C(L)\Delta y_t = \alpha\gamma y_{t-1} + \varepsilon_t$. What emerges is

$$\Delta y_t = \begin{bmatrix} .7 & .2 \\ -.2 & .7 \end{bmatrix} \Delta y_{t-1} + \begin{bmatrix} -.03 & -.06 \\ -.04 & -.08 \end{bmatrix} y_{t-1} + \varepsilon_t \,.$$

The coefficient on the level y_{t-1} is obviously a singular matrix and can be written as $\alpha\gamma$ with (e.g.) $\alpha = [-.03 \quad -.04]'$ and $\gamma = [1 \quad 2]$.

(2) (20) Dividends δ_t depend on technology shocks τ_t and demand shocks ε_t, which are zero mean, independent of each other, and independent across time. The stock price Q_t (this is a naive model, but accept it for exam-taking purposes) satisfies

$$Q_t = \beta E_t \left[Q_{t+1} + \delta_{t+1} \right] \,, \tag{2}$$

where $\beta \in (0, 1)$. The dividend process is

$$\delta_t = \tau_t + b\tau_{t-1} + \varepsilon_t + d\varepsilon_{t-1} \,. \tag{3}$$

(a) Display the Q_t, δ_t process as a moving average of the τ_t, ε_t process. Assume that explosive growth of Q_t can be ruled out.

(2) can be solved forward (assuming that explosive growth of Q_t can be ruled out) as

$$Q_t = E_t \left[\sum_{s=1}^{\infty} \beta^s \delta_{t+s} \right] \,.$$

Assuming (as should have been made explicit in the problem) that the information being conditioned on at time t is just $\{\delta_s, Q_s, s \le t\}$, $E_t\delta_{t+1} = b\tau_t + d\varepsilon_t$, and $E_t\delta_{t+s} = 0$, all $s > 1$. So we can conclude that

$$Q_t = \beta b \tau_t + \beta d \varepsilon_t \,.$$

But then combining this equation with (3) gives us a MAR for the joint process:

$$\begin{bmatrix} \delta_t \\ Q_t \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 \\ \beta b & \beta d \end{bmatrix} + \begin{bmatrix} b & d \\ 0 & 0 \end{bmatrix} L \right) \begin{bmatrix} \varepsilon_t \\ \tau_t \end{bmatrix}$$

(b) Under what conditions on b, d and β could an econometrician, observing the history of Q and δ up to time t and knowing the values of b and d, recover the technology and demand shocks at t?

First we need to require that the matrix of coefficients on the current-period shocks is non-singular. Otherwise there is some direction of variation in the current shock vector that has no effect on observables, and hence must itself not be recoverable from current and past data. The condition fo non-singularity of the contemporaneous coefficient matrix is $b \neq d$. But assuming that, there remains the question of whether this MAR operator is "invertible" — that is, whether we can write down a convergent polynomial in positive powers of L that expresses τ_t, ε_t as a linear combination of past Q and δ values. The most straightforward way to do this is, denoting by A_0, A_1 the coefficient matrices in (2a), to form $A_0^{-1}A_1$ and check whether its eigenvalues are all less than one in absolute value. One could also check the generalized eigenvalues of the A_0, A_1 pair (which you would have had to learn in some other course) by solving $|\lambda A_0 - A_1| = 0$. Either of these methods produces the result that there are two, repeating roots of 0. And it is indeed true that $(A_0^{-1}A_1)^2 = 0$, verifying that all its eigenvalues must be zero. Zero is certainly less than one, so the MAR operator is indeed invertible, so long as $b \neq d$.

(3) (60) Suppose we believe that $y_t \sim N(\mu_j, \sigma_j^2)$ for $t \in (b_j, b_{j+1})$, where $j = 1, \dots, k$ are k "regimes". Assume that we do not know the μ_j's or the σ_j^2 or the b_j's, though k is known.

(a) Assuming we have a sample $y_1, \dots, y_T$, display the likelihood function.
The log likelihood is

$$-\frac{T}{2}\log(2\pi) - \sum_{j=1}^{k}\Big((b_{j+1} - b_j)\log\sigma_j - \frac{1}{2}\sum_{t\in(b_j,b_{j+1}]}\frac{(y_t-\mu_j)^2}{\sigma_j^2}\Big)\,. \qquad (\dagger)$$

(b) Show that the likelihood function is unbounded and has multiple infinite peaks.
Pick $b_j = b_{j+1}$ for some j. Set $\mu_j = y_{b_j}$. Then as $\sigma_j^2 \to 0$, the j'th term in the log likelihood sum goes to plus infinity.

(c) Now suppose that we have a prior for $\{\mu_j\}$ and $\left\{\sigma_j^2\right\}$ that makes them independent draws from a normal-inverse-gamma prior of the form

$$K\sigma_j^{-m-1}\exp\big(-\frac{1}{2\sigma_j^2}\big(m(\mu_j - \bar{\mu})^2\big)\,, \qquad (*)$$

where K is a constant of integration and $m > 1$ is a known value. The break dates have a flat prior subject to the condition that $b_j - b_{j-1} \geq 3$ for all j — meaning that every collection of k regime intervals satisfying this restriction has the same prior probability. Find the posterior pdf.
This is more or less trivial, so long as one wants just the kernel of the posterior. To find the log posterior one just adds the log of the prior pdf $(*)$ to the log likelihood displayed in $(\dagger)$. This gives the log of the kernel of the posterior, which is all that is needed for posterior simulation. Of course one also has to note that the resulting expression only applies when $b_{j+1} \geq b_j$, all j, with the posterior density zero for all other $\{b_j\}$ sequences. To go from the kernel to the normalized posterior, that integrates to one, requires integrating the posterior, which probably can be done analytically for each given $\{b_j\}$, but is beyond what was expected on the exam. (So I should have stated that only the kernel was required.) A fully normalized posterior would also have to calculate the sum of the integrated posteriors across all the possible $\{b_j\}$ sequences, again more than expected on the exam.

(d) With this prior and posterior, formulate a MCMC scheme to sample from the posterior. The more specific you can be, and the better you exploit the special features of this model, the better your score will be.

The basic idea here is that, conditional on $\{b_j\}$ and $\left\{\sigma_j^2\right\}$, the log posterior is quadratic in $\bar{\mu}$ and $\{\mu_j\}$ jointly, so that one could draw directly from this conditional distribution. Then, conditional on $\{b_j\}$ and on $\bar{\mu}$ and on $\{\mu_j\}$, it has the form of k independent inverse-gamma distributions on $\left\{\sigma_j^2\right\}$. So from this conditional density we can also draw directly. The tricky part is drawing from the posterior on $\{b_j\}$ conditional on $\bar{\mu}$, $\{\mu_j\}$, and $\left\{\sigma_j^2\right\}$. One way to proceed is with an independence Metropolis-Hastings step, but for this one needs to formulate a proposal density. One possibility would be to draw $\{b_j, j = 2, \dots, k\}$ (note that $b_1 \equiv 1$ and $b_{k+1} \equiv T$) as k independent draws from a uniform distribution over $1, \dots, T$. This makes the proposal density easy to calculate (i.e., $1/T$). One then would accept or reject the draw based on the M-H criterion. Draws in which $\min_j \{b_{j+1} - b_j\} < 3$ of course have zero density in the true distribution and are rejected. If k is relatively small and T large, this might work. If k is on the order of, say, $T/4$, a large fraction of draws will fail to have the minimal separation of break points, so the sampling scheme might be very inefficient.

Another possibility would be make the $\{b_j\}$ draw for the $n+1$'st draw consist of independent draws of changes in b_j, $j = 2, \dots, k-1$ relative to the n'th draw. The changes might be taken to be distributed uniformly over $(-1, 0, 1)$, for example. This also has an easily calculated density: 3^-k. Draws might still violate the separation restriction, especially if k is large relative to T, but it seems likely that the violations would be less frequent than with i.i.d. uniform draws. Furthermore, this scheme would tend to stay close to the most likely $\{b_j\}$ sequence if the data sharply prefer breaks at certain dates, whereas the independence scheme would produce many rejected draws if the data sharply favor a certain pattern of break points.

I may give more detail on how to make the draws for $\{\mu_j\}$, $\bar{\mu}$ and $\{\sigma_j\}$, depending on what I see when I grade the exams.

(4) (15) "Lindley's paradox" is that when comparing models using Bayesian posterior odds, we often find the odds ratios unreasonably decisive.

(a) Explain why we tend to get extreme posterior probabilities, and what can be done about it.

Commonly the collection of models we consider is not really an exhaustive list of all the models we consider possible. For example when one model gets posterior probability .999, but we don't believe this result, it is likely that we believe that by relaxing some restrictions (e.g. on lag length or functional form) in the low-probability model we could get it to fit better. Some authors (like Gelman, Carlin, Stern and Rubin) argue that it is almost always possible to embed what we initially treat as distinct models in a larger model in which each of the original models arises when some continuously varying parameters are restricted. This may not always be practical, but even when it is not, the right response to an implausibly sharp posterior on a collection of models is to introduce other models, if only as discrete additional models, "in between" the original collection of models.

(b) Odds ratios can be used to choose lag length, treating each lag length as a different model. Do Lindley paradox effects show up in this application of odds ratios? Why or why not?

Lindley paradox effects do not usually arise in choosing lag lengths, because we usually consider a fairly rich collection of lag lengths and we do not (in a discrete time model) think that there can be "in between" lag lengths. Occasionally researchers will consider a disconnected set of lag lengths, say one, two, six or 12 months. In such a case a Lindley paradox effect could arise. If we found probability .999 on a lag length of 12 in such a situation, we would probably want to bring into consideration models 8-24, for example, because the high posterior probability on 12 suggests we were too confident that the lag lengths were short.

(5) (15) The Schwarz criterion (or BIC) chooses among models without making any reference to a prior for any of the models. We know that posterior odds calculations require a proper prior and are in general meaningless if improper priors are involved. The Schwarz criterion gives asymptotically the same results as posterior odds ratios.

The foregoing statements seem contradictory. Explain how they can all be true at once.

The BIC is derived by assuming the asymptotically Gaussian shape of the posterior is a good approximation and by concentrating attention in the log odds ratio formula on those components that increase in magnitude with T. The component of the posterior contributed by the prior in this asymptotically Gaussian case does not change with T, so it drops out of the BIC formula. Assuming one and only one of the models is true, the log posterior odds will converge to $\pm\infty$, and which it converges to is determined of course entirely by the terms that increas in magnitude with T. Therefore in large enough samples, BIC and log posterior odds will agree in sign.

The priors contribute a term to the posterior odds that does not increase in magnitude with T, but does not decrease either. In a given sample, if we double the prior odds ratio between models A and B, we will automatically double the posterior odds on the two models as well. In this sense posterior odds are always sensitive to the prior.

There is no contradiction here, because BIC and the posterior odds ratio agree in large samples only in the sense that both have the same sign in large samples. The BIC does not asymptotically become close in numerical value to the posterior odds ratio.

With sample size fixed, the prior odds ratio has a direct and proportional effect on the log odds ratio, but with the prior odds ratio fixed and sample size increasing, the odds ratio converges to zero or one, and the prior cannot influence which it converges to. This is how the statements can be reconciled.