FINAL EXAM

Answer all questions. You can take up to three hours to complete the exam.

(1) Consider the model

$$y_t = a + bx_t + cx_t\varepsilon_t \qquad t = 1, \dots, T\,, \tag{1}$$

where a, b and c are unknown parameters, the data are i.i.d. over t, and $\varepsilon_t \mid x_t \sim N(0,1)$.

(a) Display the likelihood function for this model and sample.

The data are i.i.d., and ε_t is independent of x_t, so $y_t \mid x_t$ is distributed as $N(a + bx_t, c^2x_t^2)$. This makes the likelihood

$$(2\pi)^{-T/2}c^{-T}\prod_1^T |x_t|^{-1} \exp\left(-\sum_1^T \frac{(y_t - a - bx_t)^2}{c^2x_t}\right)\,.$$

Quite a few people missed the c^{-T} term in the likelihood, and this made it impossible to derive the inverse-gamma form for the marginal distribution of c^2.

(b) Explain how to make draws from the posterior joint distribution for a, b, and c. It may be useful for you to know that if w is inverse-gamma distributed with parameters n (for shape) and α (for scale), it has pdf

$$w^{-n-1}\alpha^n \exp(-\alpha/w)\,dw \tag{2}$$

and expectation $\alpha/(n-1)$. You can assume the availability of functions that generate draws from normal and of inverse-gamma pdf's.

The marginal posterior density of c^2 under a flat prior is found by integrating out a and b. For this it was helpful to realize that this is just a standard GLS model with known form for the dependence of the residual variance on x, so the MLE is weighted least squares. That is, if we write $\tilde{y}_t = y_t/x_t$, $\tilde{x}_t = 1/x_t$, then the model is

$$\tilde{y}_t = b + c\tilde{x}_t + \eta_t\,,$$

where $\eta_t \sim N(0, c^2)$, which is exactly the standard normal linear model. So following the standard derivation for that model (or just quoting the result), integrating out a and b gives us the marginal posterior density for c^2 proportional to

$$c^{-T+2}\exp\left(-\frac{1}{2c^2}\sum_1^T \hat{u}_t^2\right)\,,$$

where $\hat{u}_t$ is the residual from the least-squares fit of $\tilde{y}_t$ to $\tilde{x}_t$ and a constant. This has the form of an $\mathrm{IG}((T-2)/2, \frac{1}{2}\sum \hat{u}_t^2)$ distribution. If the prior were flat on $\log c$ (i.e. dc/c) rather than on c^2, the degrees of freedom would be $(T-1)/2$. Best answers showed all this. Next best noticed equivalence to the SNLM after weighting, quoted

Date: June 5, 2018.

the standard results for that model. Next best just said c^2 distributed as IG, without showing degrees of freedom or scale factor. A final variant noted that the *conditional* density for c given a, b is IG, though with a different degrees of freedom and scale factor.

Because we have an exact standard marginal distribution for c^2 and an exact (normal) conditional distribution for a, b given c^2, we can draw c^2 from its marginal, then a and b from their conditional. Each time we do this we get an independent draw from the posterior, so no MCMC is necessary, not even Gibbs sampling.

(c) Suggest a way to construct a consistent estimate of the (frequentist) covariance matrix of the maximum likelihood estimates of a and b for this model.

As we've already noted, the MLE is the GLS estimator with the MLE for c^2 plugged in for its true value. c^2, the residual variance of the weighted regression, and has the usual SNLM normal distribution. c^2 can be estimated as the sample variance of the residuals. There is one important qualification here, though, that no one noticed on the actual exam. If x_t has a non-zero, continuous density at $x_t = 0$ (as it would if it were normally distributed, for example), then $1/x_t$ has infinite variance. This makes the usual argument for asymptotic normality of GLS break down, so there may be no asymptotic covariance matrix. This despite the fact that the Bayesian posterior remains nicely behaved without the need for a finite variance assumption on $1/x_t$.

(d) Would estimation of a and b by an OLS regression of y on x and a constant be consistent? How could you construct a (frequentist) covariance matrix for such estimates?

OLS is consistent. The residual in the unweighted regression is $cx_t\varepsilon_t$, and because you were given that $E[\varepsilon_t \mid x_t] = 0$,

$$E[c\varepsilon_t x_t] = E[cx_t E[\varepsilon_t \mid x_t]] = 0\,.$$

Since the data are i.i.d., the only additional assumption needed is that x_t is not constant. Even if $1/x_t$ has infinite variance, which it will if its density is continuous and positive at $x_t = 0$, OLS will be consistent. The usual formula for the covariance matrix of OLS with $\mathrm{Var}(\varepsilon) = \Omega \neq \sigma^2 I$ applies:

$$(E[X'X])^{-1}E[X'\eta\eta'X]E[(X'X)^{-1}]\,.$$

The middle term in this expression can be estimated by the usual "robust" formula, though here, since we know the form of dependence of the variance of η on x, we can do better by estimating the middle term as

$$\sum \hat{c}^2 x_t^4\,.$$

Of course this requires an additional assumption, that $E[x_t^4]$ is finite; otherwise asymptotic normality may not apply.

(e) Suppose now that the last term on the right-hand side of (1) is, instead of $cx_t\varepsilon_t$, $c\,|x_t|^{\theta}\,\varepsilon_t$. Sampling from the posterior over a, b, c and θ requires MCMC iteration. Explain how you could use your results from part 1b to put the iterations into "Metropolis within Gibbs" form.

We break the parameters into two groups: (a, b, c) and θ. Conditional on θ the model is just as before, except with weighted least squares using $x_t^{-\theta}$ rather than x_t^{-1} as weights. So we can draw directly from the posterior on $(a, b, c) \mid \theta$ by the methods described in the answer to 1b. This is a standard Gibbs step. Then conditional on the draw of (a, b, c), draw θ with a Metropolis step, because θ's conditional posterior is not standard. Alternating these two types of step should give a valid MCMC algorithm.

(2) Alice wishes to estimate a regression of income, for a sample of many thousands of individuals, on a constant, the individual's age, average GDP in the person's state of residence, and average unemployment in the person's state of residence.

(a) She realizes that because of the nature of her variables, she cannot use a fixed-effects regression with state fixed effects. Why?

All the individuals in a state will have the same state GDP and state unemployment rate value, so within-state regressions will not be able to separate state fixed effects from these other two variables that are perfectly collinear with it.

(b) She also knows that a standard random effects model should allow her to estimate her model. However, she knows that a random effects regression estimate can supposedly be characterized as a weighted average of a "between" and "within" regression, where the "within" regression is the fixed effects estimator. Since in her case the fixed effects estimator provides no estimates of coefficients on some of the variables in her regression, she wonders how to take advantage of this weighted average characterization in her case. Is there a way to do it? Explain how to do it, or why the weighted average characterization breaks down in this case.

The "between-within" formula expresses the fixed-effects estimator as

$$(\alpha_1 \tilde{X}'\tilde{X} + \alpha_2 \bar{X}'\bar{X})^{-1}(\alpha_1 \tilde{X}'\tilde{X}\tilde{\beta} + \alpha_2 \bar{X}'\bar{X}\bar{\beta}),$$

where $\tilde{}$ indicates "within" and $\bar{}$ indicates "between". So the $\tilde{}$ variables are deviations from state means and and the $\bar{}$ variables are state means. Even if there are no variables constant within states, like GDP and unemployment here, the $\tilde{X}'\tilde{X}$ matrix has zeros in rows and columns corresponding to the constant term and thus is singular. But the between-within formula does not require inverting this matrix. The components of $\tilde{\beta}$ corresponding to variables constant within states are indeterminate, but they are all multiplied by zeros so they do not affect the weighted average.

(3) Consider the model

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = y_t = c + Ay_{t-1} + \varepsilon_t\,, \tag{3}$$

where $\varepsilon_t \sim N(0, \Sigma)$ and is independent of all y_s values with $s < t$. The parameters of the model take the numerical values

$$c = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \qquad A = \begin{bmatrix} 0.53 & 0.07 & 0.50 \\ -0.12 & 0.87 & 0.05 \\ 0.23 & 0.07 & 0.80 \end{bmatrix} \tag{4}$$

The Jordan decomposition of A is given by

$$\begin{bmatrix} 0.33 & -0.33 & -0.67 \\ -0.17 & 0.67 & -0.17 \\ 0.33 & -0.33 & 0.33 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & .9 & 0 \\ 0 & 0 & .3 \end{bmatrix} \begin{bmatrix} 1.00 & 2.00 & 3.00 \\ 0.00 & 2.00 & 1.00 \\ -1.00 & 0.00 & 1.00 \end{bmatrix}. \tag{5}$$

(a) Which, if any, of the individual elements of the y_t vector are stationary?
None are stationary. Jordan decomposition $A = P\Lambda P^{-1}$ lets us rewrite the system in terms of $z = P^{-1}y$, where it is clear that the first component of z is non-stationary and the second two are stationary. Since the first column of P is non-zero in all its elements and $y = Pz$, each element of y loads on the non-stationary z_1 and is thus itself non-stationary.

(b) Show that the system exhibits cointegration and display the cointegrating vectors (i.e. the linear combinations of y that are stationary).
Since the system has three variables but just one unit eigenvalue, it displays cointegration. The cointegrating vectors are the second two rows of P^{-1}, i.e. $(0, 2, 1)$ and $(-1, 0, 1)$. (Any linear combination of these two is also a cointegrating vector.) Some people didn't see this and instead went directly to finding eigenvalues and eigenvectors of the given A or of $A - I$, not recognizing that the eigenvalues and eigenvectors were already given by the Jordan decomposition. With only two significant figures of accuracy in the given A and P, this led to some messy calculations.

(c) What are the unconditional means of the stationary components of y that you have identified?

$$\begin{aligned} z_{2t} &= [0, 2, 1][1, 0, -1]' + .9z_{2,t-1} + \eta_{2t} = -1 + .9z_{2,t-1} + \eta_{2t} \\ z_{3t} &= [-1, 0, 1][1, 0, -1]' + .3z_{3,t-1} + \eta_{3t} = -2 + .3z_{3,t-1} + \eta_{3t}\,. \end{aligned}$$

From these equations, substituting in $\bar{z}$ for every occurrence of z, we get $E[z_{2t}] = -10$, $E[z_{3t} = -2/.7]$.

(d) How could you calculate the unconditional covariance matrix of the stationary components of the model?
The covariance matrix of η_t, the residual vector in the transformed system, is $P^{-1}\Sigma(P')^{-1}$, where Σ is the covariance matrix of ε_t. Knowing that, we can use the usual formula

for the variance of a first-order scalar AR to get

$$\mathrm{Var}(z_{2t}) = \frac{\mathrm{Var}(\eta_2 t)}{1 - .81}, \qquad \mathrm{Var}(z_{3t}) = \frac{\mathrm{Var}(\eta_{3t})}{1 - .09}.$$

You were asked for the whole covariance matrix, which requires calculating $\mathrm{Cov}(z_{2t}, z_{3t})$. One way to do this, which we did not cover in class, is to calculate

$$\frac{\mathrm{Cov}(\eta_{2t}, \eta_{3t})}{1 - .9 \cdot .3}.$$

That this works could have been derived from the general matrix solution, which we did discuss in class and most people who answered correctly used. It uses the equation (where $w_t = (z_{2t}, z_{3t})$)

$$R_w(0) = \sum_{s=0}^{\infty} \Lambda^s P^{-1} \Sigma (P')^{-1} \Lambda^s,$$

(e) Write the system in VECM form. (Note that in this pure first-order case, the VECM form has no first-differenced terms on the right-hand side.)
Full credit for just observing that the system can be written

$$\Delta y_t = c + (A - I) y_{t-1} + \varepsilon_t,$$

and that this is the VECM form. $A - I$ is singular and has the form $\alpha\beta$, where

$$\beta = \begin{bmatrix} 0 & 2 & 1 \\ -1 & 0 & 1 \end{bmatrix}.$$

Figuring out numerically what α is was not required, though many tried and some succeeded. The easiest way, which most people did not see, is to observe that

$$A - I = P\Lambda P^{-1} - PIP^{-1} = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & .9 & 0 \\ 0 & 0 & .3 \end{bmatrix} P^{-1}.$$

If we use Q to denote the last two columns of P, β for the last two rows of P^{-1}, and M to denote the lower 2 by 2 diagonal of Λ, the expression above is equivalent to

$$A - I = QM\beta,$$

So we can set $\alpha = QM$. Numerically, using the given Jordan decomposition, this is

$$\alpha = \begin{bmatrix} -0.30 & -0.20 \\ 0.60 & -0.05 \\ -0.30 & 0.10 \end{bmatrix},$$

give or take some rounding error. Of course if you arrived at some other linear combination of the second and third right eigenvectors as your cointegrating vectors, you would have arrived at a different α.