Econ. 513, Time Series Econometrics Fall 1999 Chris Sims

Answers to Final Review Questions*

Disclaimer: These are questions to help you review, not sample take-home exam questions. They should remind you of topics we have studied and exercise your understanding of them. They haven't been written with a view to making them easily gradeable or definitely doable in a finite time.

(1) Suppose a particular sample and model has produced the likelihood function, in terms of a single unknown parameter μ, $L(\mu \mid \text{Data}) = e^{-10|.3-\mu|}$, and that we began with the prior pdf on μ, $\pi(\mu) = .5e^{-|\mu|}$.

(a) From the information given, can you determine whether the interval $(-3, 3.1)$ is close to a 95% posterior probability region for μ? If not, why not? If so, compute the interval's posterior probability.

This is enough information to compute the interval's probability. The posterior is proportional to $e^{-|\mu|-10|.3-\mu|}$. This can be integrated analytically by treating it in pieces:

$$\int_{-\infty}^{0} e^{\mu-3+10\mu}\,d\mu + \int_{0}^{.3} e^{-\mu-3+10\mu}\,d\mu + \int_{.3}^{\infty} e^{-\mu+3-10\mu}\,d\mu = \frac{e^{-3}}{11} + e^{-3}\frac{e^{2.7}-1}{9} + e^{3}\frac{e^{-3.3}}{11}. \quad \text{(A1)}$$

The integral over the interval $(-3, 3.1)$ can similarly be computed in pieces as

$$\int_{-3}^{0} e^{\mu-3+10\mu}\,d\mu + \int_{0}^{.3} e^{-\mu-3+10\mu}\,d\mu + \int_{.3}^{3.1} e^{-\mu+3-10\mu}\,d\mu = \frac{e^{-3}-e^{-33}}{11} + e^{-3}\frac{e^{2.7}-1}{9} + e^{3}\frac{e^{-3.3}-e^{-34.1}}{11}. \quad \text{(A2)}$$

The posterior probability of the interval is just this integral over $(-3, 3.1)$, divided by the integral over the whole real line. It is easy to see, even without carrying the calculations out, that (A1) and (A2) differ only by the terms $e^{-33}/11$ and $e^{-34.1}/11$, which are extremely close to zero. So the quoted interval is very close to a 100% probability interval, not a 95% interval.

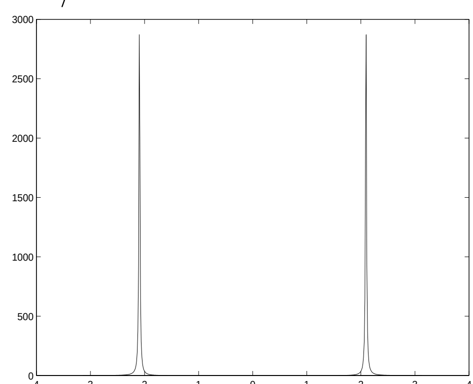
(b) From the information given, can you determine whether the interval $(-3, 3.1)$ is close to a 95% classical confidence interval for μ? If not, why not? If so, compute the interval's confidence level.

Not enough information has been given. We are given only the likelihood of this sample, which does not allow us to recover the distribution of an estimator of μ, which is what we would need to form a confidence interval.

(2) One or more of the models below imply that y will display fairly persistent seasonal oscillations, assuming that the time unit is a month. Which ones, and how do you know? (As usual, we assume $E[\varepsilon(t) \,|\, \text{past } y] = 0$.) You should be able to answer this question several ways — recursive computation of the MAR, checking roots of the system polynomial, and looking at the shape of the spectral density.

(a) $y(t) = -.99y(t-1) - .9801y(t-2) + \varepsilon(t)$

For just this subquestion, I display all three ways to answer. The roots of $1 + .99L + .9801L^2$ are $-0.5051 \pm 0.8748i$, and $\arctan(.8748/-.5051)$ is $2\pi/3$. Since the absolute value of the roots is nearly 1, they induce a strongly persistent oscillation of period 3 months, which is of course also periodic with period twelve months, and thus a seasonal. The first 14 elements of the recursively computed inverse of the polynomial are: 1.0000, -0.9900, 0, 0.9703, -0.9606, 0, 0.9415, -0.9321, 0, 0.9135, -0.9044, 0, 0.8864, -0.8775, which is clearly oscillating at about the frequency 3. The Fourier transform of the lag operator coefficients is $|1 + .99e^{-i\omega} + .9801e^{-2i\omega}|$, whose squared absolute value is $2.9407 + 3.9206\cos(\omega) + 1.9602\cos(2\omega)$. The spectral density of y is proportional to the inverse of this, and a graph of it, below, shows that it has very sharp peaks at $\pm 2\pi/3$.



(b) $y(t) = .61182y(t-1) - .9801y(t-1) + \varepsilon(t)$

This polynomial has complex roots corresponding to period 5, which is not an annual period and not any even fraction of the 12 month annual period, so the model does not imply a near seasonal pattern.

(c) $y(t) = .99y(t-1) - .9801y(t-2) + \varepsilon(t)$

This polynomial has complex roots corresponding to a period 6, which is an even divisor of 12 and thus corresponds to a seasonal oscillation.

(3) Write down a multivariate linear autoregressive model involving x and y and in which y does not Granger-cause x, yet x fails to be strictly exogenous in any equation with $y(t)$ as left-hand-side variable. [Hint: The system will have to involve more than two variables; why?]

$x \,\mathrm{GCP}\, y$ implies there is a regression with y on the left and current and past x on the right in which x is strictly exogenous. (See the notes on Granger Causality.) Thus we need an example of a system in which $y \sim GC\, x$, yet $x \sim \mathrm{GCP}\, y$. In a two-equation system, these two are equivalent, which is why we need a third variable. Any system of the form

$$\begin{bmatrix} y(t) \\ x(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} a & b & c \\ 0 & d & e \\ f & g & h \end{bmatrix} \begin{bmatrix} y(t-1) \\ x(t-1) \\ z(t-1) \end{bmatrix} + \varepsilon(t)$$

will do, for example.

(4) Consider the following multivariate autoregression:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \begin{bmatrix} .3 & .2 & .3 \\ 0 & .7 & 0 \\ -.6 & .1 & .95 \end{bmatrix} \begin{bmatrix} y(t-1) \\ y(t-2) \\ y(t-3) \end{bmatrix} + \begin{bmatrix} \varepsilon_t(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \end{bmatrix} \tag{A3}$$

Are any of the variables or blocks of variables in this system GCP to other variables or blocks?

By putting y_2 at the end of the list of variables instead of the middle, the system can be seen to have a block structure with $y_2 \,\mathrm{GCP}\, \{y_1, y_2\}$.

(5) What is the shape of the time path of smoothed estimates of the state in a Kalman filtering setup when the plant equation is $\beta(t) = \beta(t-1)$ (i.e. the true state is constant)? What if $\beta(t) = .9\beta(t-1)$? (i.e., non-constant, but deterministic, state). What if $\beta(t) = \beta(t-1) + \nu(t)$ with non-zero variance for ν?

In the first case, with β constant, the smoothed estimate is the conditional expectation, given the whole sample, of a single parameter, the true constant value of β. Therefore the smoothed estimate is perfectly flat. In the second case, since we know that $\beta_t = .9^{t-T}\beta_T$, and there is no uncertainty about this knowledge, $E_T[\beta_t] = .9^{t-T}E_T[\beta_T]$, so the smoothed path will be an exponential path shrinking toward the final smoothed estimate of β_T. In the third case, there are no easily stated qualitative restrictions on the shape of the smoothed path. Once the parameter evolution is stochastic, in principle any path for the smoothed estimates is possible, regardless of the plant equation.

(6) Consider a regression model of the form

$$y(t) = \underset{1\times k}{X}(t)\beta + \varepsilon(t)\,, \quad t = 1, \dots, T\,,$$

where $\varepsilon(t) \mid \{X(s),\, s = 1, \dots, T\} \sim t(n, \sigma)$. The $t(n, \sigma)$ is the Student-t distribution with n degrees of freedom and scale factor σ. It is the marginal distribution of a random variable z generated in two stages, by first drawing s_t^2 from a Wishart distribution with n degrees of freedom and scale parameter σ^2, then drawing x from $N(0,1)$ and setting $z = xn/s_t^2$. (The Wishart distribution for this scalar case is the distribution of the sum of squares of n i.i.d. draws from a $N(0,1)$ distribution.) The posterior distribution for $\beta \,|\, \text{data}$ under a flat prior can be simulated by using Gibbs sampling to generate an artificial sample from the joint posterior distribution of β and $\{\sigma_t^2\}_1^T$. Explain in detail how this algorithm would work. How would it be used to evaluate the posterior probability that β_1, say, lies in a particular interval (a,b)? How could you assess whether the artificial sample is big enough to be giving accurate results for this probability?

The complications I thought I saw in this question during the review session are not really there, it turns out. However, in giving you all the detail about how to draw a t-distributed random variable I was misleading (and misled myself in the review session). The way to make $\varepsilon_t \sim t(n,\sigma)$ is to set $\theta_t = \frac{1}{2}\sigma_t^2/\sigma^2$ and $v_t = \varepsilon_t/\sigma_t$, then give a joint pdf to (v_t, θ_t) proportional to

$$\theta^{(n-1)/2} e^{-\theta} \cdot e^{-\theta v_t^2} \,.$$

It is not too hard to check that when s_t^2 and z are drawn in two steps as described in the problem statement, their joint pdf has this form. With this form, because β enters linearly in v, which appears as a square in the exponential term, the posterior is normal on β, conditional on the $\{\sigma_t\}$ sequence, and in fact will just be normal with the usual GLS estimate of mean and variance, using the inverses of the σ_t's to weight observations. The conditional pdf of $\theta \,|\, v_t$ is then that of a $\chi^2(p+1)$ divided by $2(1+v_t^2)$. Thus it is after all possible to do the Gibbs sampling here analytically. One starts with, say a constant σ_t equal to the OLS residual standard deviation, and draws a β from the OLS posterior. Then one uses the estimated residuals one by one to scale T successive draws from $\chi^2(p+1)$ to get a new $\{\sigma_t\}$ sequence. And so on.

(7) The Kalman filter has produced as a posterior on the final period parameter vector $\beta_T \sim N(\hat\beta_T, \Sigma_T)$ and as a posterior on the next-to-last period $\beta_{T-1} \sim N(\hat\beta_{T-1}, \Sigma_{T-1})$. The plant equation is $\beta_t = A\beta_{t-1} + \nu_t$, with $\mathrm{Var}(\nu_t) = \Omega$. Is this enough information for you to construct the conditional distribution of β_{T-1} given the full sample of data? Suppose the model for the data is $y(t) = \beta y(t-1) + \varepsilon(t)$, with $\varepsilon(t)$ normally distributed, with known variance matrix Ψ and independent of past y's. How would you construct the posterior distribution, given the sample data, of the yet-unobserved data $y(T+4)$? Would $\hat\beta_{T-1}, \Sigma_{T-1}$ be useful for this purpose, or is all the necessary information in $\hat\beta_T, \Sigma_T$?

It is enough information to construct the β_{T-1} distribution conditional on the full sample. This is exactly the main Kalman smoothing result, that it can be done recursively, with each step using just the smoothed distribution for β_{t+1} given the full sample and the filtered distribution for β_t given data up to time t. Constructing the requested distribution of $y(T+4)$ analytically would be a mess. But it would be easy to draw from it. [The problem statement should have said the model for

y is $y(t) = \beta_t y(t-1) + \varepsilon(t)$.] For each draw, draw initially from the posterior pdf of β_T given the sample. Then construct a draw of $y(T+1)$ by drawing from the distribution of ν_{t+1} and thereby generating a draw of β_{t+1} from the plant equation, drawing from the distribution of ε_{t+1} and thereby generating a draw of $y(T+1)$ from the observation equation, etc. until a draw for $y(T+4)$ is reached. This is then repeated lots of time to produce a Monte Carlo sample from the distribution of $y(T+4)$ that reflects uncertainty about β_T as well as uncertainty about future ε's.

(8) Suppose we wish to create an artificial sample from a posterior pdf on β that is proportional to $f(\beta)$, a function we know how to evaluate. We know how to draw from the pdf $g(\beta)$, which is a so-so approximation to f. One solution is to draw from g every time, using importance sampling to generate weights to accompany the artificial sample. An alternative is to draw from g every time, but to use the Metropolis-Hastings rule to accept or reject each draw, creating a Markov chain sequence of draws that needs no accompanying weights. If $f(\beta)/g(\beta)$ is very large in some region of the parameter space, even if $f(\beta)$ itself is small in the region, so the region has low posterior probability, problems will arise with these procedures. What are the problems? Are the problems any different, or any less, with one procedure rather than the other?

For importance sampling, the problem is that the large ratios of $f(\beta)/g(\beta)$ become large weights, so that the few observations with these large weights dominate the sample and effective sample size is very small. For Metropolis-Hastings, the acceptance rule will imply that once a draw with such a high $f(\beta)/g(\beta)$ is obtained, new draws will be rejected with extremely high probability, so this value of β, instead of being weighted very high, repeats extremely many times in the artificial sample. In this situation there may be a slight advantage for importance sampling, as with the Metropolis-Hastings procedure the number of repetitions is not only very high, it is very uncertain. It might be worth noting that pure Metropolis sampling, if the distribution of jump sizes keeps them fairly small, is less likely to get stuck in this situation. (Why?)