TAKEHOME FINAL EXAM

(1) Suppose the univariate stochastic process y is ARMA(2,2) of the following form:

$$y_t = 1.6974y_{t-1} - .9604y_{t-2} + \varepsilon_t - 1.6628\varepsilon_{t-1} + .9216\varepsilon_{t-2}\,, \tag{1}$$

where ε is i.i.d. $N(0,1)$.

(a) Show that the process is stationary and that ε_t is its innovation process (i.e. its one-step-ahead forecast error).

The roots of the AR polynomial in the lag operator are complex, both 1.02 in absolute value. Therefore that polynomial has a convergent one-sided inverse and the model implies convergence to a stationary process. The roots of the MA polynomial are also complex and of absolute value 1.0417, so that moving average operator also has a convergent one-sided inverse, implying that ε can be represented as a convergent linear combination of current and past values of y and thus that it is the innovation in y. Note that the phase angles of the roots (b in their representation as $r = e^{a+bi}$ is $2\pi/12$ to within 5 significant figures in both cases.
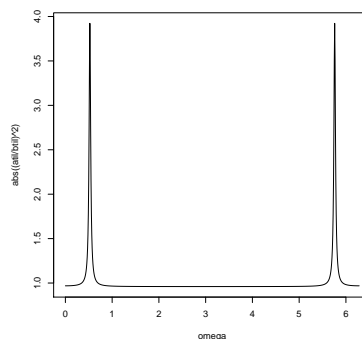
(b) Plot the spectral density of y.

If $y = c(L)\varepsilon$, with ε serially uncorrelated with unit variance, its spectral density is

$$\left| c(e^{-i\omega}) \right|^2\,.$$

In this example we have

$$c(e^{-i\omega}) = \left| \frac{1 - 1.6974e^{-i\omega} + .9604e^{-2i\omega}}{1 - 1.6638e^{-i\omega} + .9216e^{-2i\omega}} \right|^2\,.$$

In R or Matlab, which know how to do complex arithmetic, one simply calculates and plots this expression, which is by construction real-valued. Here's the plot.

Note the high spikes at $2\pi/12$ and $2\pi - 2\pi/12$. This is the spectral density of the sum of a white noise with variance 1 and another process, independent of it, that is a slowly varying sine wave of period 12.

(c) Show that if the data are monthly, this process will display strong seasonal variation. What kind of a seasonal pattern will it display?

It actually doesn't display a "strong" seasonal pattern. I had plotted the spectral density and noted the peaks at the sine wave of period 12, but did not think carefully enough about the fact that the peaks are so narrow that their integrals are small. This means that the variance of the data is dominated by the white noise component. There is a slowly varying sine wave in there, but it is very hard to see in plots of a simulated series. Simulating 6000 periods of realizations of this process and calculating an estimated autocorrelation function, one sees a clear, statistically significant seasonal pattern in the coefficients, but the absolute size of the autocorrelations is small.

To get a clearly visible sine wave in the realizations I would have had to make the peaks higher, by making the absolute value of the AR root closer to one. For example, 1.730319 on the first lag and -.998001 on the second would have worked.

(d) Explain why this example illustrates a potential pitfall of fitting an ARMA(2,2) to seasonally adjusted data.

With seasonally adjusted data, power at the seasonal frequencies is greatly reduced. Therefore in fitting a model that is only approximately correct, usual estimation methods will allow more approximation error at those frequencies. A low-order autoregression or a low-order MA model for a univariate time series can only produce a smooth spectral density across seasonal frequencies, so the lack of power at the seasonal in the data is not an issue. But though an ARMA(2,2) model may look like it does not contain any clear seasonal "mechanism", it can produce big spikes at seasonal frequencies, and it could therefore lead to estimates that imply spurious seasonal variation in the data or (perhaps worse) match the dips at the seasonal frequencies in the spectral density that arise from seasonal adjustment — in effect exploiting the two-sided nature of seasonal adjustment filters to imply spuriously accurate forecasts.

(2) Here's a version of the Solow growth model:

$$C_t + I_t = A_t K_{t-1}^{\alpha} \tag{2}$$

$$\frac{I_t}{C_t + I_t} = \theta_t \tag{3}$$

$$K_t = (1 - \delta)K_{t-1} + I_t \tag{4}$$

$$\log(A_t) = \log(A_{t-1}) + \varepsilon_t \tag{5}$$

$$\log\left(\frac{\theta_t}{1 - \theta_t}\right) = (1 - \rho)\bar{\theta} + \rho \log\left(\frac{\theta_{t-1}}{1 - \theta_{t-1}}\right) + \nu_t . \tag{6}$$

ε and ν are assumed i.i.d. normal, with unknown variances σ_ε^2 and σ_ν^2.

(a) Explain in detail how, given values for the Greek-letter parameters and an initial value K_0 for capital, one could use a computer to generate a simulated time path for all the variables in the model.

The two exogenous drivers for the model are the A (technology) and θ (savings rate) processes. A is non-stationary, so an initial value for it must be selected arbitrarily — for example, $A_0 = 1$. θ is most naturally regarded as stationary (otherwise it is eventually almost exactly one or zero almost all the time). So it makes sense to draw $\theta_0/(1-\theta_0)$ from its steady-state distribution $N\left(\bar{\theta}, \sigma_\theta^2/(1-\rho^2)\right)$. Then of course one has to use this value to solve for θ_0 itself.

With initial values in hand we can use the last two equations to recursively generate time paths for A_t and θ_t, having the computer draw shocks from the given normal distributions.

Then the first three equations can be solved to eliminate C and I, yielding

$$\theta_t A_t K_{t-1}^\alpha = K_t - K_{t-1}(1-\delta). \tag{†}$$

This can be solved recursively to deliver a time path for K.

Then (4) can be used to generate a time path for I_t, and finally (1) can be used to generate the path of C_t.

(b) If we have observed time series for C and K (and hence also I), calculating the likelihood for this model is harder than simulating it. Explain why.

We have a normal density for the two shocks ε and ν, and there are packaged routines to draw efficiently from these distributions. In the previous part we showed that there is then a non-iterative algorithm that generates a simulated path for the variables.

The likelihood also has a recursive structure, as usual in time series models, in that we can write the log likelihood times the prior as

$$\sum_{t=1}^T \log(p(C_t, K_t \mid C_{t-s}, K_{t-s}, s > 0; \alpha, \delta, \rho, \bar{\theta}, \sigma_\varepsilon^2, \sigma_\nu^2) + \log \pi(\alpha, \delta, \rho, \bar{\theta}, \sigma_\varepsilon^2, \sigma_\nu^2, A_0, \theta_0), \quad (*)$$

where π is the pdf of a prior. We can solve for ε_t, ν_t at any given date, given the history of C_t and K_t up to that date — we get A_t from (2) and θ_t from (3) and (4). But we can't find the likelihood by simply substituting functions of the observed data into $(*)$, because of the nonlinearity in the mappings. We need Jacobian terms.

To be precise, (2)-(4) imply the Jacobian between C, K and A_t, θ_t

$$\left| \frac{\partial(A_t, \theta_t)}{\partial(C_t, K_t)} \right| = \frac{1}{A_t K_{t-1}^{2\alpha}}.$$

And then to get the Jacobian for the full transformation from ε_t, ν_t to A_t, θ_t, we need to multiply the Jacobian above by

$$\left| \frac{\partial(\varepsilon_t, \nu_t)}{\partial(A_t, \theta_t)} \right| = \frac{1}{A_t} \frac{1}{(1-\theta_t)^2}.$$

So, since we have the Gaussian form for the conditional pdf's in $(*)$ and have an analytic form for ε_t, ν_t as functions of C_t, K_t (and past values and parameter values) at each date, we can construct the likelihood. It involves quite a bit more algebra than the simulation, and would be slower computationally, but this model is simple enough that the analytic form of the likelihood could be calculated quite quickly on a desktop or laptop.

Note the contrast with a rational expectations version of this model, in which agents choose I_t and C_t optimally at each date rather than saving a fixed fraction of income. In the rational expectations version there is no analytic form available (unless $\delta = 1$) for the mapping from shocks to choice variables. We can approximate the mapping itself numerically, but numerically approximating the Jacobian would be many times harder.

(c) Sometimes particle filtering provides a simple way to approximate the likelihood in dynamic models that are easily simulated. That is not so true in this model. Why?

We can fairly easily compute the exact likelihood here, so the gap in difficulty between simulating the model and evaluating the likelihood is not that great. Furthermore, the particle filter involves generating at each date draws of an imperfectly observed state, then reweighting those draws based on the observed data. But here the model implies the state variable is perfectly observed, so this strategy simply doesn't work. If the observable were just C, not K, the particle filter would be a help. In that case, we could at each date and for each particle make a draw for K_t from (†), then reweight the particles by using the conditional pdf of K_t given C_t, for which an analytic formula could be constructed. Proceeding this way would avoid the need to integrate K_t out of the joint density of K_t, C_t given history.

(d) Would the particle filter become easier to implement if we assumed that one or both of observed C and K contain i.i.d. normal measurement error? Explain your answer.

Yes, if both have measurement error, because then at each step of the particle filter we could draw particles from the mechanism in the original model, then reweight by the normal measurement error density. Of course this adds two parameters to the model (the measurement error variances), and changes its economic interpretation if the model implies non-trivial measurement error variances. If only one has measurement error, the particle filter is still no help.

(3) In a reduced form vector autoregression, orthogonalized by a triangular normalization, simulated draws from the posterior distribution of the impulse responses can be constructed without use of MCMC, because the posterior takes the form of a standard distribution.

    (a) Explain how this is done. (This is a standard result, so you can be brief.)

    Under a flat or conjugate prior, the covariance matrix of residuals has a Wishart posterior distribution, and conditional on that the regression coefficients have a joint normal distribution. Since these are standard distributions, we can draw from them directly (covariance matrix first, then coefficients) without MCMC or importance sampling.

    (b) In a structural VAR where the number of free parameters in the matrix A_0 of coefficients on current values is exactly $(n^2 + n)/2$ and lagged coefficients are unrestricted, it may be possible to construct posteriors on the impulse responses without MCMC by basing them on draws from the posterior on the reduced form parameters. But there are cases where this is not straightforward. Suppose the pattern of zero restrictions on A_0 in a 3 variable system is

$$\begin{bmatrix} x & x & 0 \\ x & 0 & x \\ 0 & x & x \end{bmatrix}. \tag{7}$$

How could you sample from the posterior for this model? Could your sampling be based on an initial set of draws from the reduced form posterior?

With this pattern of zeros in A_0, there are pairs of distinct A_0 matrices that both satisfy the zero restrictions and both imply the same reduced form covariance matrix $\Sigma = (A_0'A_0)^{-1}$. This implies that differences in postierior density across such pairs come entirely from any prior one might be imposing on A_0. This is not a problem for sampling from the posterior in itself — a posterior with multiple peaks properly reflects the fact that multiple parameter values are equally likely according to the data. It is a problem for the strategy of starting by drawing values for the reduced form covariance matrix, though, as now there is no unique way to produce an A_0 that corresponds to Σ.

Equally important is that with this pattern of zeros in A_0 there are Σ matrices such that no A_0 satisfying the restrictions can match Σ. So if we make draws from the reduced form posterior in the usual way, some of them will be Σ matrices that violate the zero constraints on A_0.

Nonetheless, it is possible to treat the prior as "flat" or conjugate (i.e. Wishart) over Σ values that can be represented as $(A_0'A_0)^{-1}$, with pairs of A_0 values implying the same Σ given equal weight. (That there are no more than two A_0 values matching a given Σ is true, though not obvious.) One would draw from the posterior on Σ, discard any draw that does not satisfy the restrictions, then check whether there are two A_0's that match the drawn Σ. If there are two, pick one with 50% probability, and otherwise just keep the draw. This prior would give odd results, and result in few draws being retained, if the estimated reduced form Σ itself failed to satisfy the contraints. Also, if model comparisons were to be made, a separate Monte Carlo calculation would be

required to determine what proportion of draws from the prior on Σ fail to satisfy the restrictions.

A better approach would be to impose a prior (say Gaussian), based on substantive reasoning about the application, on the elements of A_0, and to use an MCMC scheme to draw from the posterior on A_0. Because the distribution of the AR coefficients (both reduced form and structural) is of known Gaussian form conditional on A_0, we can easily integrate over the coefficient distribution analytically to calculate the marginal posterior density for each draw of A_0, using it to guide a random-walk Metropolis (for example) MCMC scheme. Then to get a draw from the joint posterior on A_0 and the coefficients, we can go through the MCMC sample of A_0 draws and for each one make a draw from the corresponding Gaussian distribution for the coefficients.

Of course this procedure is just what one would do in any over-identified structural VAR model.