Problem Set: Kalman Smoothing, Metropolis ARMA

1. *The Natural Rate*. Consider the model

$$p_t = a_t p_{t-1} + b_t U_{t-1} + g_t + e_t \tag{1}$$

$$\begin{bmatrix} a_t \\ b_t \\ g_t \end{bmatrix} = \begin{bmatrix} a_{t-1} \\ b_{t-1} \\ g_{t-1} \end{bmatrix} + n_t \tag{2}$$

With the elements of the vector $h_t' = \left[e_t \; n_t' \right]$ mutually independent and independent of all variables in the system dates $s < t$. We also assume joint normality for h. The variable p is the inflation rate (change in log of GDP deflator) and U is the (adult civilian) unemployment rate. At least if the variances of the elements of n are small enough, it is reasonable to interpret $-g_t / (1 - b_t)$ as the "natural rate of unemployment" – the rate that is consistent with zero inflation in the long run. Note that it might turn out that a stays close to one, in which case we instead have a "NAIRU", a "non-accelerating inflation rate of unemployment". You are to use this model to form, first, Kalman filtered estimates of the time series of parameters [a, b, g], then smoothed estimates of them based on the full time series. Use them to construct two (filtered and smoothed) time series of estimates of the natural rate. Use a prior that sets

$$\begin{bmatrix} a_0 \\ b_0 \\ g_0 \end{bmatrix} \sim N\left(\begin{bmatrix} .9 \\ -.1 \\ .5 \end{bmatrix}, \begin{bmatrix} .04 & .01 & -.01 \\ .01 & .01 & -.01 \\ -.01 & -.01 & .09 \end{bmatrix} \right). \tag{3}$$

Make the variance of n_t in (2) .004 times the variance matrix in (3). We will call this factor $d = .004$. Set s_e^2, the variance of $e(t)$ in (1), equal to the sample variance of $\Delta p(t)$. Because this is a regression model, not a complete model of the data, you will need to use as likelihood the p.d.f. of the data on p conditional on all values of U and on the initial value of p. Use quarterly data, taking it from CitiBase, which you can access from Statlab or via the departmental web page. Use the longest sample available on CitiBase (though don't use data before 1947 if it should be available.) Use the U data in percentage points (so recent unemployment is around 5) and the p data also in percentage points, at annual rates, i.e. 400 times the quarterly change in the log of the price index. Plot the natural rate series you find, together with error bands for them. Since the natural rate is a ratio of two parameters that are both normal, it has a complicated distribution. If the parameters g and b are estimated with high precision at each t, you can use the approximation $\mathrm{Var}((m+e)/(l+h)) = s_e^2 / l^2 + m^2 s_l^2 / l^4$. Otherwise, you will need to use Monte Carlo simulation to estimate the standard errors or quantiles of the distributions to get the error bands. Finally, check the sensitivity of your results to the degree of time variation, indexed by the parameter d above, and to s_e^2 by redoing the work for $d = .001$ and $d = .016$ and with s_e^2 also both four times larger and one fourth as large. [This is four additional cases. You don't need to trace out all 8 possible combinations of these variations in d and s_e^2.] Consider not only how the results change, but also how the size of the posterior p.d.f. changes in assessing whether it appears that it would be a good idea (which you are not supposed to pursue here) to proceed to formal Monte Carlo integration over a prior on one or both of these parameters.

2. Consider the ARMA(1,1) model

$$y(t) = ry(t-1) + e(t) + ae(t-1) \ . \tag{4}$$

Using US quarterly data on log of real GDP for 1947-the present (or whatever is available on Citibase – whichever is shorter), fit this model by producing a sample from the posterior joint p.d.f. for a and r given the sample. Here you are to use the unconditional likelihood and assume a flat prior. The marginal distribution for $y(t)$ here is

$$N\left(0, s_e^2\left(\frac{(ar+1)^2}{1-r^2}+1\right)\right) \ . \tag{5}$$

The conditional distribution of $e(0)$ given $y(0)$ is $N\left(\frac{s_e^2}{s_y^2}y(0), s_e^2\left(1-\frac{s_e^2}{s_y^2}\right)\right)$. Then the conditional distribution of $y(t)$ given past values of e and y can be read off for each t from (4). This allows us easily to write down the joint p.d.f. of the data for $t = 0,\ldots,T$ and $e(0)$, which we can then treat as the likelihood (and hence flat-prior posterior) in r, a, $e(0)$ and s_e^2. To generate an artificial sample from this posterior distribution, use the Gibbs-sampling idea of making draws one coordinate at a time, but draw from the exact conditional distribution only for s_e^2 (which has a standard inverse-gamma conditional distribution). For the other variables, use simple Metropolis sampling with a normal jump distribution centered on the last draw. As a starting point, take $r = .9$, $a = .8$, $e(0) = 0$, drawing from the distribution of s_e^2 conditional on these choices as the first draw. Make the standard deviations of the jump distributions .05 for r and a and .002 for $e(0)$. Make draws of 1000 complete sets of 4 parameters. Discuss whether the algorithm seems to have produced a reliable set of draws or not. Plot histograms for the marginal distributions of r and a.