

Christiano
FINC 520, Spring 2010
Homework 8, due May 30.

1. This question studies the Monte Carlo Markov Chain (MCMC) algorithm and the Laplace approximation. Because we do this in an example where we know the true distribution being approximated, we have a direct way to assess its performance. In addition, we can explore the wisdom of the recommendation that the parameter in the jump distribution be tuned so that the acceptance rate of the algorithm is in a neighborhood of 25 percent.

Consider the Weibull probability distribution function (pdf) for the random variable, x . In MATLAB, the density of the Weibull distribution associated with a specific value of the random variable, x , can be obtained by executing $[g] = wblpdf(x, a, b)$. Here, a and b are parameters of the Weibull density. Consider $a = 10$, $b = 20$. Graph this pdf over the grid, $[7, 11.5]$, with intervals 0.001 (i.e., graph g on the vertical axis, where $g = wblpdf(x, 10, 20)$, and x on the horizontal axis, where $x = 7 : .001 : 11.5$). Compute the mode of this pdf by finding the element in your grid with the highest value of g . In addition, you will require the appropriate information matrix. (Hint: the second derivative of a function, f , around a point x can be approximated as follows:

$$f''(x) = \frac{f(x + 2\varepsilon) - 2f(x) + f(x - 2\varepsilon)}{4\varepsilon^2},$$

for ε small. For example, you could set $\varepsilon = 0.000001$.) The jump distribution described in class was Normal, which can produce a candidate that is negative. However, a random variable with a Weibull distribution is non-negative. You could adjust the algorithm so that the jump distribution is the absolute value of the Normal described in class.

- (a) Set the number of random draws to $M = 1,000$ (this is actually a small number). Graph the density function implied by the MCMC approximation, the actual density function, and the one implied by the Laplace approximation. When graphing these functions, make sure they are appropriately scaled so that the area underneath

them is unity. (In the case of the MCMC you will have to draw using the output of the MATLAB histogram program using, say, 150 bins.) What value of the jump distribution parameter, k , produces roughly a 26 percent acceptance rate?

- (b) Show the effects of a very large and a very small value of k . How does the algorithm perform with these alternative values of k ? Does it look like the algorithm would converge to the right answer for values of k that produce a very low or high acceptance rate, as $M \rightarrow \infty$?
- (c) Notice how ‘choppy’ the histogram seems when you run it with $M = 1,000$. Try a much larger value of M , say $M = 100,000$, and show how the results smooth out. Above, I suggested that you use the absolute value of the Normal distribution as your jump distribution. Try instead using the Normal itself and then define the Weibull density of a negative candidate random variable as zero. How does this change affect the performance of the MCMC algorithm?
- (d) How does the Laplace approximation work in approximating Weibull?

2. Suppose you have time series data on output growth, Δy_t . Suppose we wish to estimate μ_y and the vector, $\gamma = \begin{pmatrix} \sigma_y & \rho_1 & \rho_2 \end{pmatrix}$:

$$\begin{aligned} \mu_y &= E\Delta y_t \\ \sigma_y^2 &= E[\Delta y_t - \mu_y]^2 \\ \rho_j &= \frac{E[\Delta y_t - \mu_y][\Delta y_{t-j} - \mu_y]}{E[\Delta y_t - \mu_y]^2}, \end{aligned}$$

for $j = 1, 2$. (Note in particular that it is σ_y and not σ_y^2 that needs to be estimated.)

- (a) Show how to set up the estimator for the 4×1 column vector of unknown parameters, $\beta = \begin{pmatrix} \mu_y & \gamma \end{pmatrix}$, as an exactly identified GMM estimator. Thus, you must find a GMM error column vector, $h_t(\beta)$, with the property, $Eh_t(\beta^0) = 0$, where β^0 denotes the unknown true value of β . Derive an explicit expression for

$$D' = \frac{\partial Eh_t(\beta)}{\partial \beta'} \Big|_{\beta=\beta^0},$$

and recall that $\sqrt{T}(\hat{\beta}_T - \beta^0)$ converges in distribution to $N(0, (D')^{-1}SD^{-1})$, where $\hat{\beta}_T$ is the GMM estimator.

(b) Explain how S might be estimated using data.

(c) Let

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix},$$

where S_{12} is 1×3 , S_{22} is 3×3 , and S_{11} is a scalar. The block structure of S is designed to be conformable with the structure of β in terms of μ_y and γ . Suppose $S_{12} = 0$. Show that $\sqrt{T}(\hat{\gamma}_T - \gamma^0)$ and $\sqrt{T}(\hat{\mu}_{y,T} - \mu_y^0)$ are asymptotically independent random variables (hint: recall that lack of correlation between Normal random variables implies that they are independent.)

(d) Based on your answer to (c), derive an expression for the limiting distribution of $\sqrt{T}(\hat{\gamma}_T - \gamma^0)$ in terms of S_{22} and the bottom 3×3 block of D . Write the distribution out explicitly in terms of the Normal distribution.

3. Consider the model economy in the previous homework. The solution to that model is a 4×4 matrix, B , where

$$z_t = Bs_t, \quad z_t = \begin{pmatrix} \pi_t \\ x_t \\ r_t \\ rr_t^* \end{pmatrix}.$$

Here, s_t is the vector of exogenous shocks having the following law of motion:

$$s_t = \begin{pmatrix} \Delta a_t \\ m_t \\ h_t^* \\ \mu_t \end{pmatrix} = \begin{bmatrix} \rho & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \rho_\mu \end{bmatrix} \begin{pmatrix} \Delta a_{t-1} \\ m_{t-1} \\ h_{t-1}^* \\ \mu_{t-1} \end{pmatrix} + \epsilon_t,$$

$$s_t = Ps_{t-1} + \epsilon_t, \quad E\epsilon_t\epsilon_t' = V,$$

where V is a diagonal matrix. The vector of observed variables is denoted Y_t . Suppose the econometrician has observations on output growth only:

$$Y_t = \Delta y_t.$$

The observer equation has the following representation:

$$Y_t = H'\xi_t + w_t, \quad Ew_t w_t' = R,$$

and law of motion for the state, $\xi_t = \begin{pmatrix} s_t & s_{t-1} \end{pmatrix}'$, is

$$\xi_t = F\xi_{t-1} + v_t, \quad Ev_t v_t' = Q.$$

Now, Y_t and w_t are scalars. Consider the following baseline values for the economic parameters of the model:

$$\theta = 0.75, \quad \beta = 0.99, \quad \kappa = 0.10, \quad \phi_\pi = 1.49, \quad \phi_x = 0.19.$$

Let the parameters of the stochastic processes take on the following values:

$$\delta = 0, \quad \lambda = 0.5, \quad \rho = 0, \quad \rho_\mu = 0.5, \quad R = 0.$$

Note that there is no measurement error, technology is a random walk and the monetary policy shock is *iid*. Also,

$$V = \begin{bmatrix} 0.0084^2 & 0 & 0 & 0 \\ 0 & 0.0007^2 & 0 & 0 \\ 0 & 0 & 0.4^2 & 0 \\ 0 & 0 & 0 & 0.003^2 \end{bmatrix}.$$

(a) Solve the model and compute the implied values of

$$\sigma_{\Delta y} = \left\{ E[\Delta y_t]^2 \right\}^{1/2}, \quad \rho_i = \frac{E[\Delta y_t - E\Delta y_t][\Delta y_{t-i} - E\Delta y_t]}{\sigma_{\Delta y}^2}, \quad i = 1, 2.$$

(b) In the previous question, you defined the GMM estimator of the above three statistics and determined their distribution. Here, we investigate the accuracy of the large sample assumption used to derive the distribution in a sample of size 200. In particular, use the model to generate 1,000 artificial data sets of 200 observations each: $\Delta y_1, \dots, \Delta y_{200}$. In each of the 1,000 data sets, draw $\epsilon_1, \dots, \epsilon_{200}$ independently from a $N(0, V)$, where $V = E\epsilon_t \epsilon_t'$. In addition, to initiate the simulations, you will need to do a random draw of ξ_1 . For this, you should use $N(0, W)$, where $W = E\xi_t \xi_t'$. In each of

the 1,000 artificial data sets, compute the GMM point estimates, $\hat{\sigma}_{\Delta y, T}$, $\hat{\rho}_{1, T}$, $\hat{\rho}_{2, T}$, as well as their sample standard deviations using the GMM formulas.¹

- i. Compute the mean and standard deviation of the point estimates, across the 1,000 artificial data sets. These calculations provide you with the actual small sample means and standard deviation of the GMM estimators of $\sigma_{\Delta y}$, ρ_1 , ρ_2 . Is the GMM estimator of $\sigma_{\Delta y}$, ρ_1 , ρ_2 unbiased? In each case, is the GMM estimator of the sample standard deviation of the estimator itself unbiased? What about its sampling uncertainty across the 1,000 artificial data samples? Is that sampling uncertainty small?
- ii. To assess the Normality result, compute the percent of times, across the 1,000 realizations, that the true values of $\sigma_{\Delta y}$, ρ_1 , ρ_2 lie outside 20 percent confidence intervals. In each artificial sample, the 20 percent confidence interval should be computed in the usual way. In the case of each parameter, the confidence interval is formed by the point estimate plus and minus the appropriate critical value times the GMM-estimated standard deviation. For the critical value, it is useful to know

$$prob[|x| > 1.28] = 0.20,$$

where x is $N(0, 1)$. Your calculations are expected to provide evidence that the GMM asymptotic theory works reasonably well. That is to say, if you are handed a set of point estimates, $\hat{\sigma}_{\Delta y, T}$, $\hat{\rho}_{1, T}$, $\hat{\rho}_{2, T}$, and you are told they are based on a sample of size 200, it is not unreasonable to suppose that the point estimates are realizations of a Normal distribution with mean equal to the true value of these parameters and variance consistently estimated using the GMM formulas.

- (c) We shall estimate, using Bayesian methods, the autocorrelation (i.e., λ) and innovation standard deviation, σ_{h^*} , of the preference

¹You may assume, as in the question above, that $S_{12} = 0$ and you may use the MATLAB routine, `se.m`, that is provided, using $l = 2$, $\theta = 1$, but make sure you confirm that that routine is doing the right thing.

shock, h_t^* . Let the prior distribution of λ be a beta distribution with mean equal to the true value of λ and standard deviation equal to 0.20. You may use the routine that is provided, `priordens.m`, to do the calculations (this routine computes the log pdf for various distributions).² The routine requires a lower and upper bound on the prior distribution. Let these bounds be 0 and 1. Let the prior distribution for σ_{h^*} be inverted gamma (type I), with mean equal to the true value and standard deviation 0.002. In addition, the lower and upper bounds on this variable should be 0 and 0.1, respectively. Graph these two prior distributions.

- (d) Do inference about the parameters using the standard Bayesian procedure, using one of the datasets you constructed for the earlier question with Monte Carlo simulation. You may find it convenient to include the command, `randn('seed',1)` at the start of your program, so that you use the same set of random numbers each time you run your program. This will be important when you compare runs based on different runs. You want to be sure that differences do not reflect random differences in simulated data. Use the Kalman filtering software from homework #7, to construct the Normal likelihood of the data. Use `priordens.m` for the log prior distribution. For the purpose of optimization, you may want to use the MATLAB multidimensional minimization algorithm, `fminsearch.m`. That algorithm uses as its third input argument a variable called `options`, which should be set by entering the following statement before the call to `fminsearch`:

```
options = optimset('Display','iter','TolFun',1e-13,'TolX',1e-13,'MaxIter',200);
```

The first input into `fminsearch` is the name of the function which takes as input values for the parameters and produces as output the corresponding posterior mode (not including, of course, the denominator constant). Suppose, for example, the function is called as follows, $x = \text{func}(x_0, p_1, p_2, \dots, p_N)$, where x_0 is the

²Note that the beta distribution is not symmetric. As a result, supposing that the mean of the prior coincides with the true value of the parameter has the effect of pushing the mode of the posterior distribution away from the true value of the parameter.

column vector containing values for the two parameters being estimated and $p_i, i = 1, \dots, N$ are other parameters that are required to evaluate the criterion. Then, the arguments of `fminsearch` are $(\textit{func}, x0, \textit{options}, p_1, \dots, p_N)$. Finally, it's probably a good idea to set the optimization criterion to the penalty value, $-.10e^{15}$, in case `fminsearch` attempts to evaluate the criterion at an impermissible value for the parameters. For example, you should restrict λ to lie inside the interval, $(0, 1)$, and the innovation standard deviation should be greater than zero.

- i. Compute and report the mode of the posterior distribution for the two parameters. Record the amount of computer time it takes to find the posterior mode by placing the command, `tic;`, just before the call to `fminsearch`, and `toc;` just after the call. Compute the Laplace approximation to the posterior density of the data, $p(Y)$. Plot the prior distribution for the two parameters and the posterior distribution using the Laplace approximation. Since the Laplace approximation to the posteriors is Normal, you can use MATLAB's `normpdf` function to do this. You can use the `priordens.m` program included with this homework to graph the prior density. To graph a density, you have to specify a range of values for the random variable. In the case of the autoregressive parameter, this is easy because it lives on the unit interval. You can use a grid, `.001 : .001 : .999`, for this. In the case of the innovation standard deviation, you should make the lower boundary, x_l say, one-half the prior mean and the upper boundary, x_u , twice the prior mean. Then, set $w = (x_u - x_l) / 800$, and make the grid `$x_l : w : x_u$` .
 - ii. Redo the calculations with a prior mean of $\lambda = 0.7$. What is new the posterior mean? I expect you to find little difference in the latter, reflecting that there is a lot of information in the data about λ . There is relatively little information in the data about the innovation standard deviation.
- (e) Redo part (d) using the moment-matching Bayesian limited information procedure described in class. In particular, replace the likelihood used in (d) of this question with the likelihood of $\hat{\gamma}_T$

used in question 2 (d) above. Record the time needed to compute the mode of the posterior distribution. Report all the results. You should see little difference. In addition, the computer time is cut substantially. The computer time required for the moment matching procedure is independent of T .

4. A variable of interest to policy-makers is the output gap, x_t . The main reason for this is that the gap indicates where the economy is in relation to its ideal position. More narrowly, the variable, x_t , is an important input into inflation forecasts (see one of the structural equations from homework 7). However, x_t is fundamentally a latent variable. It is a theoretical concept and cannot be read directly from some measured variable. If our theory is correct, then we can use it, along with the available data on output growth, to estimate the output gap. We can do this by solving the projection problem,

$$p[x_t | \{\Delta y_{t-j}, j = \pm 0, 1, 2, \dots\}] = \sum_{j=-\infty}^{\infty} f_j \Delta y_{t-j}.$$

Often, the HP filter is also used to estimate the output gap. We can compare the properties of the two measures using the frequency domain tools that we have developed in this course.

According to the orthogonality principle, the weights, f_j , satisfy

$$E \left(x_t - \sum_{j=-\infty}^{\infty} f_j \Delta y_{t-j} \right) \Delta y_{t-k} = 0,$$

for $k = \pm 0, 1, 2, \dots$. That is:

$$C_{x,y}(k) = \sum_{j=-\infty}^{\infty} f_j C_{y,y}(k-j)$$

where

$$C_{y,y}(l) \equiv E \Delta y_t \Delta y_{t-l}, \quad C_{x,y}(k) \equiv E x_t \Delta y_{t-k}.$$

- (a) Apply the convolution argument presented in class to establish that

$$f(z) = \frac{S_{x,y}(z)}{S_{y,y}(z)},$$

where

$$f(z) = \sum_{j=-\infty}^{\infty} f_j z^j.$$

(b) To see how to compute the weights on powers of z in $f(z)$, let

$$Z_t = \begin{pmatrix} \Delta y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \tau'_y \\ \tau'_x \end{bmatrix} \xi_t,$$

where τ'_y and τ'_x are 1×8 row vectors and ξ_t is the state, defined in question 2 above. The covariance generating function of Z_t , $S(z)$, is:

$$S(z) = \begin{bmatrix} S_{y,y}(z) & S_{y,x}(z) \\ S_{x,y}(z) & S_{x,x}(z) \end{bmatrix} = \begin{bmatrix} \tau'_y \\ \tau'_x \end{bmatrix} S_{\xi}(z) \begin{bmatrix} \tau_y & \tau_x \end{bmatrix},$$

where $S_{\xi}(z)$ is the covariance generating function of the state:

$$S_{\xi}(z) = (I - Fz)^{-1} Q (I - F'z^{-1})^{-1}.$$

Then,

$$f(z) = \frac{\tau'_x S_{\xi}(z) \tau_y}{\tau'_y S_{\xi}(z) \tau_y}.$$

To be in a position to compare the optimal filter weights, $f(z)$, with those of the HP filter, it is convenient to define the filter as it applies to the level data, y_t . Since $\Delta y_t = (1 - L)y_t$ the weights, $\tilde{f}(z)$, as they apply to the level of the data are:

$$\tilde{f}(z) = \left(\sum_{j=-\infty}^{\infty} f_j z^j \right) (1 - z) = \frac{\tau'_x S_{\xi}(z) \tau_y}{\tau'_y S_{\xi}(z) \tau_y} (1 - z).$$

Show how the integral formula can be used to recover the coefficients in $\tilde{f}(z)$. Use the Riemann approximation to that formula to implement it on the model analyzed above. Graph these coefficients.

- (c) Earlier in the course, we discussed the HP filter. In the frequency domain, this filter has the following representation:³

$$h(z) = -g_2 \frac{(1-z)(1-z)(1-z^{-1})(1-z^{-1})}{(1-g_1z-g_2z^2)(1-g_1z^{-1}-g_2z^{-2})},$$

where g_1 and g_2 are constants which are determined by the smoothing weight. This filter applies directly to the level data, and so it is comparable with $\tilde{f}(z)$.

- i. Prove that the phase angle of this filter is zero at all frequencies.
- ii. Compute the filter weights by applying the inversion formula. Graph these alongside the filter weights in $\tilde{f}(z)$. Note how different the two filters are. The HP filter provides a very poor estimate of the output gap in this model. There is a simple economic explanation which turns on the nature of potential output (recall, the output gap is the percent deviation between actual and potential output.) In the model, potential or, efficient, output is more volatile than actual. This is because the dynamics of the model are dominated by a preference shock whose effects are transitory. Because transitory disturbances induce people to save, economic activity tends to under-react relative to the efficient equilibrium. The implicit assumption of the HP filter is that potential output is a smooth version of actual output. This is why the HP filter and the optimal filter for extracting the gap in the model are negatively correlated. The same logic suggests that if ρ is large and the technology shock dominates the equilibrium of the model, then the two filters would resemble each other more closely. It is easily verified that this is indeed the case.

³See King, Robert G. and Sergio Rebelo. 1993. Low Frequency Filtering and Real Business Cycles. *Journal of Economic Dynamics and Control*. 17(1-2): 207-231.