Christiano
FINC 520, Spring 2007
Homework 7, due Monday, May 28.

1. Obtain Angus Maddison's (http://www.ggdc.net/maddison/) annual data on per capita output in India covering the period, 1884 to 2003 (the data are in an excel file on his web page, and have url

   http://www.ggdc.net/maddison/Historical_Statistics/horizontal-file_03-2007.xls).

   These data can be used to help assess Niall Ferguson's (see his book, *Empire*) view that British colonialism helped India by bringing law, good governance and a sense of 'fair play'. The British withdrew when India achieved independence on August 15, 1947. To evaluate Ferguson's hypothesis, we can compare Indian economic performance with and without British rule.

   (a) Estimate the mean growth rate, μ_1, over the period 1884 to 1947. Estimate the mean growth rate, μ_2, for the period 1948 to 2003. Treat the estimates of μ_1 and μ_2 as independent and test the null hypothesis that $\mu_1 = \mu_2$ (actually, Ferguson's hypothesis, presumably, is $\mu_1 > \mu_2$) versus the alternative that $\mu_2 > \mu_1$.

   (b) Actually, the British were in India long before 1884. They arrived in about 1600 and made it part of the British empire by the mid 1700's. Maddison's per capita gdp data are spotty before 1884. Compute the annual growth rate of per capita GDP from 1700 (Maddison's observation that is closest to 1750) to 1884. To get a sense of what was going on in India before the British, compute the mean annual growth rate from 1000 to 1700 using Maddison's data (use the formula, $(x_{t+n}/x_t)^{1/n}$, where x_t denotes an observation on per capita income in year t.) How does Ferguson's hypothesis fare in light of these data? Compute the ratio of Indian per capita output to UK per capita output over the period 1884 to 1947. Now how does Ferguson's hypothesis fare?

2. The Bayesian posterior distribution is:

$$f(\theta|y) = \frac{g(\theta)}{f(y)},\ g(\theta) \equiv f(y|\theta)\, f(\theta),$$

where y is the data set, $f(y)$ is the marginal distribution of y, θ is the set of parameters, $f(\theta)$ is the prior distribution, $f(y|\theta)$ is the likelihood function and $f(\theta|y)$ is the posterior distribution. Evidently, the mode of the posterior distribution, θ^*, can be found by simply maximizing $g(\theta)$ with respect to θ. Note, too, that the second derivative, $f_{\theta\theta}(y|\theta)$, of the posterior distribution at $\theta = \theta^*$ can be computed up to a constant by computing (possibly numerically), $g_{\theta\theta}(\theta^*)$. In practice, particularly in the standard case when θ has more than one element, numerical simulation algorithms are needed to obtain the posterior distribution, $f(\theta|y)$. An important class of algorithms for doing this is the Markov Chain Monte Carlo (MCMC) algorithms.

The idea is to generate a sequence of values of θ which, if the sequence is long enough, will cause θ to traverse every interval with the same frequency as is implied by the density, $f(\theta|y)$. Inputs into constructing the sequence include the mode of $g(\theta)$ as well as an estimate of $g_{\theta\theta}$ at the mode and the value of g at each step in the sequence.

The MCMC algorithm works like this. Let the sequence start with $\theta_0 = \theta^*$. Compute $g_{\theta\theta}$ at θ_0. Let $g_0 = g(\theta_0)$. Set $i = 1$ and use a random number generator to compute x_i:

$$x_i = \theta_{i-1} + k \times N(0, C),$$

where C is the unique lower triangular matrix with positive diagonal elements, having the property

$$CC' = -(g_{\theta\theta})^{-1}.$$

Also, k is a positive scalar discussed below. Form:

$$\lambda_i = \frac{g(x_i + \theta_{i-1})}{g(\theta_{i-1})},$$

and compare this to a random number, U, drawn from a uniform, [0,1]. Then,

$$\theta_i = \begin{cases} x_i + \theta_{i-1},\ U < \lambda_i \\ \theta_{i-1},\ U \geq \lambda_i \end{cases}.$$

In words, if the proposed step, x_i, in the sequence generates an increase in g, i.e., it takes the sequence into a high probability region according to $f(\theta|y)$ (recall, $f(\theta|y)$ is proportional to g), then the step, x_i, is definitely taken. If the proposed step in the sequence leads to a lower probability region of f (i.e., $\lambda_i < 1$), then that step is only taken with some probability. Continue in this way, for $i = 1, 2, ...\ , M$ generating a sequence, $\theta_1, \theta_2, ..., \theta_M$. The histogram of this sequence is the MCMC approximation to the density, $f(\theta|y)$. The key thing to note is that the algorithm produces a 'preference' for wandering in regions where the density, $f(\theta|y)$, is high. To a lesser extent, the algorithm will also wander in regions where $f(\theta|y)$ is lower. The parameter, k, should be selected so that the 'acceptance rate' is roughly 30 percent. Obviously, when θ is a scalar, then C could just as well be set to unity.

Consider a mixture of two normals, as in [22.3.5] in Hamilton, with $N = 2$. Let the first distribution be a standard Normal, and let the second have mean and variance 4 and 1.1, respectively. Consider $\pi_1 = 1$, .7, .5. Set M to $100,000$ in each of these three cases. Plot the histogram of the θ's generated by the MCMC algorithm, as well as the true density. Redo the calculations with $M = 1,000,000$.

Some computational notes: finding the mode of the mixture of normal distribution requires numerical optimization. Here is how you can do this.

(a) Compute a grid of values for theta: xx=-4:.001:8;

(b) Evaluate the density at each point on this grid: f=(mu/sqrt(2*pi*v1))*exp(-(xx-mu1).^2/(2*v1))+((1-mu)/sqrt(2*pi*v2))*exp(-(xx-mu2).^2/(2*v2));

Here, mu corresponds to π_1 and 1-mu corresponds to π_2. Also, v1 is σ_1^2, etc.

(c) Then, find the maximum on this grid, and an interval, (x1,x2) that brackets it: [Y,I]=max(f); x1=xx(I-100); x2=xx(I+100);

(d) Finally, use the minimization routine, fminbnd, to refine the estimate of the mode:

[X,FVAL,EXITFLAG] = fminbnd(@(xx)(-((mu/sqrt(2*pi*v1))*exp(-(xx-mu1).^2/(2*v1))... +((1-mu)/sqrt(2*pi*v2))*exp(-(xx-mu2).^2/(2*v2)))),x1,x2,opti
12,'Display','off'));

To better understand this command, type help fminbnd at the MATLAB prompt.

3. Consider the data on the long rate that you have used in previous homeworks.

   (a) Fit a 2 lag scalar ar representation to that (for this, you can use the software written for homework #4).

   (b) Test the null hypothesis that 2 lags is appropriate against the alternative that 5 is (see the discussion in the class for this, or, page 297 in the text). Does this test indicate that it is acceptable to use 2 lags, at the 10 percent significance level? Is the page 297 approach relevant for the case in which there is conditional heteroscedasticity in the residuals? If not, how would you do the test in a way that is not undermined by heteroscedasticity? (You need not implement this test.)

   (c) Continue working with the 2 lag specification. Graph the residuals. Graph the centered ± 2 year standard deviation of the residuals. Do you agree that there is notable heteroscedasticity in the residuals?

   (d) Estimate the coefficients in an ARCH(2) representation for the residuals from the 2 lag ar representation. Plot the residuals from this representation. Do they look like they have constant variance? Should they have constant (conditional) variance according to the ARCH(2) model?

   (e) Show that the two-step procedure you used to estimate the ar representation with 2 lags and ARCH(2) errors is a GMM procedure.