KALMAN FILTER EXERCISE

CHRISTOPHER A. SIMS
PRINCETON UNIVERSITY
SIMS@PRINCETON.EDU

There is data for US real GDP, chain-weighted, from 1947:I to 2006:2 on the course website in R binary format (use the `load()` command — it will put an R time series object named `gdp`, with attached dates, in your workspace) and in .xls and .csv formats. Using these data, logged, with the start date of your sample (not counting lagged initial values) 1952:I and the end date 2005:IV, do the following.

(a) Fit, by finding the posterior mode, a fifth order autoregression to the data, conditioning on the initial conditions (so that you in fact use data going back to 1948:IV). The model will have the form

$$y_t = \alpha + \sum_{s=1}^{5} \rho_s y_{t-s} + \varepsilon_t \,, \qquad \varepsilon \mid \{y_{t-1}, y_{t-s}, \ldots\} \sim N(0, \sigma^2) \,. \tag{1}$$

Use a prior with one factor that makes the means of the coefficients of the lagged variables [1,0,0,0,0] and their variances [.5,.4,.3,.2,.1]. The model should include a constant term, and the prior should include a second factor proportional to a $N(0,\sigma^2)$ pdf for $\bar{y} \cdot (\sum \rho_i - 1) - \alpha$, where $\bar{y}$ is the sample mean of the initial 5 lags of y. This prior is conditional on σ^2, so another factor will be the prior on σ^2. Choose this to make $1/\sigma^2$ exponential, i.e. use the pdf $\exp(-1/\sigma^2)/\sigma^4 \, d\sigma^2$. Note that though the likelihood-maximizing estimate for $\vec{\rho}$ and α can be found by OLS, the prior will make the posterior mode not coincide with the MLE, so that you will need to use an iterative maximization program to find the posterior mode. The OLS estimates are still worth computing, because they will be a good starting point for the optimization program.

(b) Repeat the exercise with the ARMA(5,1) model

$$y_t = \alpha + \sum_{s=1}^{5} \rho_s y_{t-s} + \varepsilon_t + \theta\varepsilon_{t-1} \,, \qquad \varepsilon \mid \{y_{t-1}, y_{t-s}, \ldots\} \sim N(0, \sigma^2) \,. \tag{2}$$

Use the same prior as before on $\vec{\rho}$, α, and σ^2, and use a $N(0,\sigma^2)$ prior (conditional on σ^2) for $\varepsilon_{1949:IV}$, the initial "state". For this model efficient likelihood evaluation will require the Kalman filter.

(c) Repeat the exercise with the ARMA(4,1) model described by 2 but with four instead of 5 lags on y. Same prior as before, except with variances on the four lags in the prior [.4,.3,.2,.1].

(d) Use the estimated coefficietns of each model to make a forecast 20 quarters ahead based on data through the end of the sample (2005:IV). Plot the actual logged gdp

Date: October 5, 2006.

series (including the two available 2006 values) and the three forecasts, all on the same graph. You need only show the actual values for the last 20 quarters of the sample.

You can use code I wrote for the maximization and for the Kalman filter. R and Matlab versions are on the course website. Matlab's default optimization routine does not try to use derivative information and is therefore very inefficient on functions (like this likelihood function) that are smooth. If you don't use my code (and better code is likely available) you should be sure that you use a routine that uses gradient information. It may calculate gradients numerically, so you don't have to code them, but it should not ignore gradient information.