KAGY NOTE

Jean-François Kagy in class today raised the question of whether it is possible to use just a two-dimensional state vector for the problem in the current 513 exercise. The answer is yes.

Why then would anyone consider using $s_t = [y_t, \dots, y_{t-k+1}, \varepsilon_t]$, as I suggested, or using Hamilton's suggestion of $s_t = \left\{(1 - \sum \rho_s L^s)^{-1} \varepsilon_{t-j}, j = 0, \dots, k\right\}$? (Hamilton uses ξ_t to refer to what I have been calling s_t) The Kagy form exploits the fact that I told you to condition on $y_0, y_{-1}, \dots, y_{-k+1}$. Thus even at the first $t = 1$ observation, these numbers are known and can be put into the "X" of Hamilton's observation equation. Equivalently, we can think of the observed data as being the sequence of values for $y_t - \alpha - \sum \rho_s y_{t-s}$, $t = 1, \dots, T$. With this setup, the state-space representation does not provide a framework for thinking about what to do if we don't know lagged values of y, but are presented with a single observation y_1. The other two setups, using higher-dimensional states, fully characterize the unconditional distribution of y_t in the cases where the model implies stationarity. Hamilton, gives formulas for the unconditional mean and variance of his state at the top of p.381. These will themselves be functions of the ρ_s parameters for Hamilton's setup and of the ρ_s parameters and θ if the lagged y's are in the state vector. The unconditional mean and covariance matrix for the state, when the state evolution equation is $s_t = A s_{t-1} + \varepsilon_t$, with $\mathrm{Var}(\varepsilon_t) = \Omega$, is the solution to the equation

$$W = \Omega + AWA' .$$

A useful matrix identity is $\mathrm{vec}(ABC) = (C' \otimes A)\vec{B}$. This allows us to write an explicit solution for W as

$$\vec{W} = (I - (A \otimes A))^{-1} \vec{\Omega} .$$

The matrix inversion does not work if A has unit eigenvalues, and if it has any eigenvalues exceeding one in absolute value the formula is meaningless, because there is no stationary covariance matrix. But it may in some applications be attractive to initiate the Kalman filter using this intial covariance matrix for the state, and then update the distribution for the state one observation at a time, starting with the first observation.

This approach would not be appropriate for this problem, however, as it is clear we want to allow for non-stationary behavior in log GDP. And once we have decided to condition on intial values for y, it is possible to reduce the state space dimension as Jean-François has suggested. The state is $s_t = [\varepsilon_t, \varepsilon_{t-1}]'$, the observations are $z_t = y_t - \alpha - \sum \rho_s y_{t-s}$ for $t = 1, \dots, T$, and the plant and observation equations are

$$\begin{aligned} s_t &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} s_{t-1} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \\ z_t &= \begin{bmatrix} 1 & \theta \end{bmatrix} s_t . \end{aligned}$$

Date: October 10, 2006.

It is important to note here that the Jacobian of the transformation from $y_1, \ldots, y_T$ to $z_1, \ldots, z_T$ is one, so that the pdf's for the z sample and the y sample, conditional on initial y's, are the same. Otherwise generating likelihood with the transformed "observations" would not be correct.

Even when we condition on initial lagged y's, though, a drawback to the Kagy approach is that for forecasting or likelihood evaluation it requires auxiliary equations that are not included in the state and plant equation system. The ρ coefficients do not even appear explicitly in the Kagy system as I have written it, so in that form one needs side equations to define the z_t's (and these have to be reevaluated every time $\vec{\rho}$ changes in likelihood evaluation). If one includes the lagged y's in Hamilton's x_t, and thereby makes the observation equation slightly more complicated, there is still nothing in the system that specifies that most of next period's x_t vector is just this period's shifted forward in time. In other words, for forecasting or likelihood evaluation, the system needs to be augmented with equations or code that form the z's from the y's and ρ's or else populates the x_t vector at each t by picking elements from $y_{-k+1}, \ldots, y_T$. With the other two setups, likelihood and forecasts arbitrarily many steps ahead can be calculated using the plant and observation equations alone.