MID-TERM EXAM ANSWERS

(1) Suppose a stock price p_t and the stock dividend δ_t satisfy these equations:

$$p_t + \delta_t = Rp_{t-1} + \eta_t \tag{1.1}$$

$$\delta_t = \gamma\delta_{t-1} + \phi p_{t-1} + \varepsilon_t\,, \tag{1.2}$$

where (ε_t, η_t) conditional on $\{\delta_s, p_s | s < t\}$ is distributed as $N(0, \Omega)$.

(a) Suppose we know R, γ, ϕ and Ω and have data for $p_t, t = 1, \ldots, T$ but no data on δ_t at any date t. How could you use the Kalman filter and/or smoother to estimate the time path of dividends? Be specific, including discussion of how you would handle initial conditions.

We can take the state vector to be $s_t = [p_t, \delta_t]'$. Then the two equations above become, in matrix notation,

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} s_t = \begin{bmatrix} R & 0 \\ \phi & \gamma \end{bmatrix} s_{t-1} + \begin{bmatrix} \eta_t \\ \varepsilon_t \end{bmatrix}\,,$$

or

$$\Gamma_0 s_t = \Gamma_1 s_{t-1} + \xi_t\,.$$

This becomes a standard state equation if we premultiply by Γ_0^{-1}, i.e.

$$s_t = As_{t-1} + \zeta_t\,, \text{ with}$$

$$A = \Gamma_0^{-1}\Gamma_1$$

$$\text{Var}(\zeta_t) = \Gamma_0^{-1}\Omega(\Gamma_0^{-1})'\,.$$

The observation equation is then $y_t = Hs_t$ with $H = [1\ 0]$. Application of the Kalman filter is now standard and would deliver a time series of estimates of the unobservable component of the state, δ_t. These would be partial-sample estimates, however, and to obtain estimates reflecting information in the full sample we would need to run the Kalman smoother. The Kalman filter only describes how to update a distribution based on new information. We have to start with a distribution for s_0, the pre-sample state. One natural choice is the unconditional

Date: November 18, 2006.

distribution implied by the model, assuming that the system admits a stationary distribution. For a dividend and stock price series, though, this might be unrealistic. Companies often grow steadily, or decline and go out of existence. So in this case it might be reasonable to postulate a very dispersed initial joint distribution for s_0, perhaps choosing the covariance matrix so it implies $E[\delta_0 \mid P_0] = (R-1)P_0$. There is no unique right answer to this initial conditions part of the question, though there are also many possible bad answers.

(b) Suppose that for two dates in the middle of the sample there is no data at all, though at all other dates there is data on p_t and not on δ_t, as before. Explain how to estimate both the time path of δ_t and the missing values of p_t using the Kalman filter.
Missing observations are easy to handle with the Kalman filter. These are just instances of observations in which H, the coefficient matrix in the observation matrix, happens to be zero. The Kalman filter has no trouble with H varying between the usual $[1\ 0]$ form and a $[0\ 0]$ form. At each step of the Kalman filter, a conditional mean of the state is formed, so these are estimates of the missing values of p_t. And as in the previous part, for many purposes we would want to use smoothed estimates to get the most accurate possible estimate of the missing values of p_t.

(c) How could you form estimates of γ, ϕ, R and Ω using these data on p?
The Kalman filter delivers the value of the pdf of the data for every set of values for the parameters listed. We could then maximize this likelihood value by using a hill-climbing algorithm that varies the parameters trying to find a likelihood peak. We could also use a prior over the parameters. Then we would multiply the likelihood value by the prior pdf at each parameter vector value to form the maximand for the hill-climbing routine. Finally, it would probably be best to form a posterior mean using some MCMC method rather than being satisfied with modal values of the posterior or likelihood. One way to do this relatively efficiently is to use the Kalman-filter-like backward recursion we discussed in class to sample from the distribution of the state vector for each parameter vector value. Then for each fixed state vector sequence the whole model becomes a pair of linear simultaneous equations. Because of the simultaneity, sampling

from this posterior is not trivial. An answer that recognized this was a good one.

(2) For each of these moving average models, in which in each case ε_t is i.i.d. $N(0,1)$, determine whether the representation is fundamental, and if it is not display the fundamental representation and determine the variance of the innovation.

(a) $x_t = \varepsilon_t - 1.1\varepsilon_{t-1} + 1.21\varepsilon_{t-2}$
This is obviously not fundamental, as its last coefficient exceeds the first. The two roots of the $C(L)$ MA polynomial are $(1 \pm \sqrt{3}i).5/1.1$, which both have absolute value $1/1.1 < 1$. Since both have to be inverted to arrive at the fundamental MA operator, it is therefore

$$1 - \frac{1}{1.1}L + \frac{1}{1.21}L^2 \doteq 1 - .9091L + .8263L^2 \,.$$

The innovation has to be scaled up by 1.1^2 and its variance is therefore $1.21^2 = 1.331$.

(b) $x_t = \varepsilon_t - 1.5\varepsilon_{t-1} + .5625\varepsilon_{t-2}$
Here there are two stable roots of $4/3$. So the representation is fundamental.

(c) $x_t = \varepsilon_t - \varepsilon_{t-1} - 1.21\varepsilon_{t-2} + 1.21\varepsilon_{t-3}$
This polynomial can be factored as $(1 - 1.21L^2)(1 - L)$. The second factor, having a root on the unit circle, is ok. The first has two roots, $\pm 1/1.1$), both of which have to be inverted. The result will be

$$\left(1 - \frac{1}{1.21}L^2\right)(1 - L) = 1 - L - .8263L^2 + .8263L^3 \,.$$

The innovation variance then has to be scaled up by the factor $1.21^2 = 1.331$.

(d) $x_t = \varepsilon_t - 2.5\varepsilon_{t-1} + \varepsilon_{t-2}$
One root is .5, the other 2. The fundamental representation MA operator is $1 - L + .25L^2$ and the innovation variance has to be scaled up by $2^2 = 4$.

(3) Suppose our model is

$$y_t = \varepsilon_t + \alpha\varepsilon_{t-1},\ t = 1, \ldots, T \tag{3.1}$$

$$\varepsilon_t \sim N(0, \sigma^2),\ \text{i.i.d. for } t = 0, \ldots, T \tag{3.2}$$

$$\varepsilon_t \text{ independent of } \{y_{t-s}, s > 0\},\ \text{all } t. \tag{3.3}$$

Our prior pdf for α is uniform on $(-1, 1)$ and for σ^2 it is the pdf $e^{-\sigma^2} d\sigma^2$ on the positive real line.

(a) Show the expression for prior times likelihood for this setup, treating ε_0 along with α and σ^2 as unknown parameters.
The pdf is

$$\varphi(\varepsilon_0; \sigma^2) \cdot (.5) \cdot \mathcal{I}_{(-1,1)}(\alpha) \exp(-\sigma^2) \prod_{t=1}^{\infty} \varphi(y_t - \alpha\varepsilon_{t-1}; \sigma^2), \tag{$*$}$$

where $\varphi(x; \nu^2)$ is the pdf of a normal distribution with variance ν^2 and $\mathcal{I}_{(a,b)}(x)$ is the indicator function for x being in the interval (a, b).

(b) A quick method for estimating this model that is sometiems proposed is to set $\varepsilon_0 = 0$, then estimate α and σ^2 by "iterated OLS". This is possible because once we know ε_{t-1} we can construct ε_t as $y_t - \alpha\varepsilon_{t-1}$ for each t. So we can start with an initial guess of α, construct corresponding ε's, estimate (3.1) by OLS to get a new α, construct a new ε sequence, etc. until we get convergence. At the end we will have not only estimates of α and σ^2, but, using the usual OLS formulas, a "posterior distribution" for α, σ^2 constructed as if the artificially constructed sequence of ε_t's were real data and this was indeed a standard linear regression model. The resulting posterior for α and σ^2 is "Normal-inverse-gamma", a standard form it is easy to sample from.

Explain how you could use this incorrect but perhaps not crazy posterior distribution for α and σ^2, together with the model's $N(0, \sigma^2)$ distribution for ε_0, to generate an independence Metropolis-Hastings MCMC sample from the correct posterior distribution on α, σ^2 and ε_0. Let

$$s_1 = \sum_{t=0}^{T-1} u_t^2(\hat{\alpha})$$

$$
\begin{aligned}
s_2 &= \sum_{t=1}^{T} u_t^2(\hat{\alpha}) \\
\hat{\alpha} &= \frac{\sum_{t=1}^{T} y_t \hat{\varepsilon}_{t-1}}{s_1} \\
\hat{\varepsilon}_t &= y_t - \hat{\alpha}\hat{\varepsilon}_{t-1} \quad t \geq 1 \\
\hat{\varepsilon}_0 &= 0 \,.
\end{aligned}
$$

The last three equations state that $\hat{\alpha}$ is the OLS estimator for α in a regression of y_t on $\hat{\varepsilon}_{t-1}$, while at the same time $\hat{\varepsilon}_t$ is the residual when $\hat{\alpha}$ is used to form the residual. The iterations described above converge when the last three equations hold simultaneously.

The independence M-H procedure goes as follows: Start with α, σ^2 values, possibly the converged estimates described above. Draw ε_0 from a $N(0, \sigma^2)$ distribution. Draw a new σ^2 from the inverse-gamma distribution implied by the "regression", i.e. an inverse gamma with scale parameter s_2 and degrees of freedom $T - 1$. Draw a new α from the normal distribution implied by the regression equation as the conditional distribution of $\alpha \mid \sigma^2$, i.e. a $N(\hat{\alpha}, (s2/s1)/T)$. Then decide whether to keep or reject this draw by forming the ratio of the true likelihood displayed in $(*)$ to the false pdf

$$\phi(\varepsilon_0; \sigma^2)\sigma^{-2(T-1)} \exp(-s1/\sigma^2)\phi(\hat{\alpha}, (s1/s2)/T) \,.$$

This ratio has to be compared to its value for the previous parameter values. If the ratio exceeds the previous value, the new draw is kept. If it is below the previous value, the new value is kept with probability equal to the ratio of the current ratio to the previous ratio, and if the new value is not kept the previous parameter values are repeated in the artificial sample.