Statistics 910, 2009

Final Examination

This is a take-home exam. I expect for each of you to work alone. You are permitted to use course textbook and course notes, but are not permitted to discuss the exam with *anyone*. If you have a question or think there's a problem with question, send me an e-mail. Your completed exam is due at my office in two days.

The exam questions are equally weighted. Please give complete answers; I will not give credit to answers lacking supporting material. Express all answers in your own words.

All time series in the questions are real-valued, with realizations (if relevant) from $t = 1, 2, \ldots, n$. To obtain the data needed in some questions use the scan command in R, as in

```
x1 <-scan(``http:// path /name_of_time_series.dat'')
```

with the path set to

```
www-stat.wharton.upenn.edu/~stine/stat910/exam_2009
```

1. Consider the two time series $X_{1,t}$ and $X_{2,t}$ that you can download from the course web page using the command

   ```
   x1 <-scan(``http:// path /x1.dat'')
   x2 <-scan(``http:// path /x2.dat'')
   ```

   These two time series are independent realizations of stationary stochastic processes, but are these two stochastic processes the same?

   Test the claim (null hypothesis) that $X_{1,t}$ and $X_{2,t}$ were generated by the same stochastic process. Assume that the realizations are independent and that the processes are stationary. If you need to make additional assumptions, state them clearly.

2. Suppose that the two time series $Y_{1,t}$ and $Y_{2,t}$ are independent realizations of the same AR(p) stochastic process with white noise variance σ^2. The process has mean zero, so $\mathbb{E}\, Y_{1,t} = \mathbb{E}\, Y_{2,t} = 0$.

   Suppose that we estimate the coefficients $\phi_1, \ldots, \phi_p$ using least squares from $Y_{1,t}$, obtaining $\hat{\phi}$. We then fit these estimated coefficients to the data in $Y_{2,t}$, obtaining the residual series

   $$e_t = Y_{2,t} - \hat{\phi}_1 Y_{2,t-1} - \cdots - \hat{\phi}_p Y_{2,t-p} \qquad (1)$$

How well can we expect estimates from one series to predict another realization from the same process?

(a) Show that $\mathbb{E}\, e_t^2 \approx \sigma^2(1 + p/n)$ to terms of order $1/n$.

(b) Suggest how this result can be used to form a criterion for choosing the order of autoregression.

3. An exponentially weighted moving average provides a prediction of the next observation when fit to a time series x_t. The prediction has the recursive form

$$\hat{x}_{t+1} = \hat{x}_t + \alpha(x_t - \hat{x}_t) \tag{2}$$

The predictions are initialized by setting the first value in the recursion to some constant, $\hat{x}_0 = c$, where $c = \mathbb{E}\, X_t$ (if that were known) or perhaps the first value of the time series.

(a) Would it be reasonable to form predictions of the time series

```
x3 <-scan(``http:// path /ewma.dat'')
```

using the exponential moving average (2)?

(b) In ARMA models, the presence of static initializations diminish geometrically. (For example, recall the effect of initializing an AR(1) model when simulating a series.) Can the same be said for the effect of the chosen initial value for exponential moving average?

4. A prominent economist was interested in the presence of 20-year cycles in certain economic time series. Because the data were noisy and contained higher frequency cycles, the economist began his analysis by smoothing the time series to remove these extraneous components. His first smoothing step was to apply a 5-year moving average. If we write the "raw" time series as x_t, then the data after this first step is

$$y_t = (x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})/5\ , \quad t = 3, \ldots, n-2. \tag{3}$$

The series y_t has $n' = n - 4$ observations. Next, the economist differenced the smoothed series, forming

$$z_t = y_{t+5} - y_{t-5}\ , \quad t = 6, \ldots, n' - 5\ . \tag{4}$$

The economist then analyzed the smoothed time series z_t and found evidence of strong 20-year cycles.

(a) Consider the following data that measure the historical price of wheat annually.

```
wheat <-scan(``http:// path /wheat.dat'')
```

Do you find evidence of a 20-year cycle in this time series? (Treat this series as x_t and do not smooth the series. You'll do that later in the question.)

(b) What's the effect of the smoothing operations applied by the economist to the presence of cycles in a time series? In particular, consider the gain and phase of the linear filter implied by these operations.

(c) Apply the economist's filtering operations to the wheat price series. (Form the series z_t with x_t set to the wheat prices.) Do you see evidence of a 20-year cycle in the filtered data?