Statistics 910, 2009

Solutions, Final Examination

1. There are a number of approaches for testing the equivalence of two stochastic processes based on independent realizations. The test should be invariant of time shifts; tests based on covariances or the periodogram have this property. Most tacitly assume normality; the covariances are sufficient only if the process is Gaussian. Also, don't forget to compare the means of the series.

   - Some methods for comparison in the frequency domain resemble the analogous problem of comparing the distributions of two random variables. Let $\hat{f}_1$ and $\hat{f}_2$ denote the periodograms of $X_{1,t}$ and $X_{2,t}$. If the variances of the two series are similar (if they differ, we're done), then we can normalize the spectral densities; assume then that $\text{var}(X_{1,t}) = \text{var}(X_{2,t})$. The difference between the cumulative sums $\sum_{j=1}^{m} \left(\hat{f}_1(\lambda_j) - \hat{f}_2(\lambda_j)\right)$ behaves like the Brownian bridge obtained by subtracting two CDFs if H_0 holds. Alas, these tests are known to have low power.

     Alternatively, the ratio of the periodograms under H_0 has an F distribution with 2 and 2 df, so the density of the ratio

     $$\hat{r}(\lambda) = \hat{f}_1(\lambda)/\hat{f}_2(\lambda) \tag{1}$$

     is $f_R(r) = (1+r)^2$ under H_0. (If you then take logs, you get a logistic density with mean 0 under H_0, but this does not generalize if you do any smoothing of the sdf.) You can fit a regression model of the form, for example,

     $$\mathbb{E}\,\hat{r}(\lambda) = \beta_0 + \beta_1\lambda + \beta^2\lambda^2 + \text{ etc} \tag{2}$$

     (or use the log of the ratio) by maximum likelihood and test whether the coefficients are zero. (If you use the log of the ratio, then β_0 is a scale shift, as if one generating process is a rescaled version of the other.) You don't have to use polynomials; a flexible adaptive smoother *that provides a test of* H_0 would work just as well (or better) so long as you get a test.

   - Parametric ARIMA models are another choice (none of you picked this choice). Based on the topics covered in class) one could identify an ARMA model for each (we are given that the processes are stationary). If the two realizations pick models of the same orders, then compare the estimated coefficients (and $\hat{\sigma}$ too). The best way to do that would be a likelihood ratio test, which basically boils down to an F-test of the residual sums-of-squares based on fitting separate models versus the sums-of-squares obtained by fitting common models to the two series (not the easiest to do with the basic R tools).

     If the realizations lead you to pick different types of models, you get an easy comparison if one chosen model nests inside the other (as an ARMA(2,1) inside of an ARMA(3,2)); fit the two and compare the estimates of the added parameters. This is basically the same as the LR test when the models match. If the models do not nest, you're left with a model selection problem.

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 $\mathbb{E}\, Y_{1,t} = \mathbb{E}\, Y_{2,t} = 0$.

   (a) e_t denotes the prediction error from using the coefficients obtained by fitting an AR(p) model to $Y_{1,t}$ to predict an independent realization $Y_{2,t}$. Let $\hat\phi$ denote the vector of estimates $\hat\phi = (\hat\phi_1, \ldots, \hat\phi_p)'$ and let $Y_{2,t}^s = (Y_{2,t}, Y_{2,t-1}, \ldots, Y_{2,s})'$ for which the covariance matrix is $\Gamma_p = \mathbb{E}\, Y_{2,t-1}\, Y_{2,t-p}$. With these conventions, the expected squared error is then

$$
\begin{aligned}
\mathbb{E}\, e_t^2 &= \mathbb{E}\left(Y_{2,t} - \hat\phi' Y_{2,t-1}^{t-p}\right)^2 \\
&= \sigma^2 + \mathbb{E}\left((\phi - \hat\phi)' Y_{2,t-1}^{t-p}\right)^2 \\
&= \sigma^2 + \mathrm{tr}\; \mathbb{E}\left((\phi - \hat\phi)(\phi - \hat\phi)'\right) \mathbb{E}\left(Y_{2,t-1}^{t-p}\; Y_{2,t-1}^{t-p\,\prime}\right) \\
&= \sigma^2 \left(1 + \tfrac{1}{n} \mathrm{tr}\; \Gamma_p^{-1}\, \Gamma_p\right) + o(1/n) \\
&\approx \sigma^2 (1 + p/n) \qquad (3)
\end{aligned}
$$

   This uses Property 3.9, page 133 of the text.

   (b) The implication is that MSE of prediction grows with the size of the model. To make this work, we need an estimate of σ. Pretend that future of $Y_{1,t}$ is uncorrelated with $\hat\phi$ (this is in fact approximately true to the order of accuracy needed here). The MLE for σ^2 (as produced, say, by the Kalman filter when fitting a model of order $m \ge p$ so there is no bias from omitted variables) is slightly biased (it does not adjust for degrees of freedom)

$$
\mathbb{E}\, \hat\sigma_m^2 = \mathbb{E}\, \frac{1}{n} \sum_{t=1}^{n} e_{m,t}^2 \approx \frac{n-m}{n} \sigma^2 \qquad (4)
$$

   when fitting an AR(m) using the full sequence $t = 1, \ldots, n$ with estimates $\hat\phi^m$ (the residual is $e_{m,t} = Y_{1,t} - \hat\phi^{m\prime} Y_{1,t-1}^{t-m}$ for $t > m$). If we use this in the expression obtained in (a) and correct for bias, we discover a familiar sort of penalty for the dimension of the model,

$$
\left(\hat\sigma_m^2 \frac{n}{n-m}\right)(1 + m/n) = \hat\sigma_m^2 \frac{n+m}{n-m} \approx \hat\sigma_m^2 (1 + 2m/n)\;. \qquad (5)
$$

   The penalty grows at a rate of twice the number of parameters. (An argument similar to this motivated Akaike to propose the "FPE", a predecessor to AIC, for selecting the order of an autogression.)

3. The question requires you to appreciate that an exponentially weighted moving average can be intrepreted as an integrated moving average. Start by considering the form of an exponentially weighted moving average (EWMA). The classical form looks a little different from the updating expression for the EWMA posed in the exam:

$$\hat{x}_{t+1} = \hat{x}_t + \alpha(x_t - \hat{x}_t) \tag{6}$$

The EWMA is usually defined as a weighted average of prior observations, with "exponentially" decaying weights. Given $0 \le \omega < 1$, define (ignoring that we have a finite realization of data)

$$\begin{aligned} s_{t+1} &= \frac{x_t + \omega x_{t-1} + \omega^2 x_{t-2} + \cdots}{1 + \omega + \omega^2 + \cdots} \\ &= \frac{x_t}{1 + \omega + \omega^2 + \cdots} + \omega \frac{x_{t-1} + \omega x_{t-2} + \cdots}{1 + \omega + \omega^2 + \cdots} \\ &= (1-\omega)\, x_t + \omega s_t \\ &= \alpha\, x_t + (1-\alpha) s_t \qquad (\alpha = 1 - \omega) \\ &= s_t + \alpha(x_t - s_t) \end{aligned} \tag{7}$$

as defined in (6) with $s_t = \hat{x}_t$.

A second expression embeds the EWMA in the ARIMA family. The easiest way (again, given what we have covered) is to write the EWMA using backshift notation. If $\hat{x}_t$ is the optimal predictor (in the sense of MSE), then the error of this predictor must be white noise. Hence,

$$x_t - (1-\omega)(x_{t-1} + \omega x_{t-2} + \omega^2 x_{t-3} + \cdots) = a_t \tag{8}$$

where $a_t \sim WN(0, \sigma^2)$. Now write this expression using backshift notation as

$$\begin{aligned} a_t &= x_t - \frac{1-\omega}{1-\omega B} x_{t-1} \\ &= \left(1 - \frac{B - \omega B}{1 - \omega B}\right) x_t \\ &= \left(\frac{1-B}{1-\omega B}\right) x_t \end{aligned} \tag{9}$$

or $(1-B)x_t = (1-\omega B)a_t$, an IMA(1,1) model.

(a) To determine if it is reasonable to form predictions of the time series $X_{3,t}$ using the exponential moving average (6), you simply need to verify that this process appears to be an IMA(1,1). You may be interested to know that this time series is in fact a sign-reversed, rescaled version of the global temperature data! I shortened it as well, cutting off some of the later data where the drift is stronger.

(b) Because the IMA(1,1) model is not stationary, we cannot ignore the initialization of the estimates. Plus, there is no such thing as a "stationary distribution" from which to draw the first values. The presence of the unit root implies that the variance of the process is growing, as in the case of a random walk. As a practical matter, however, the initial description of the EWMA makes it clear that the mean does eventually "ignore" the initialization conditions.

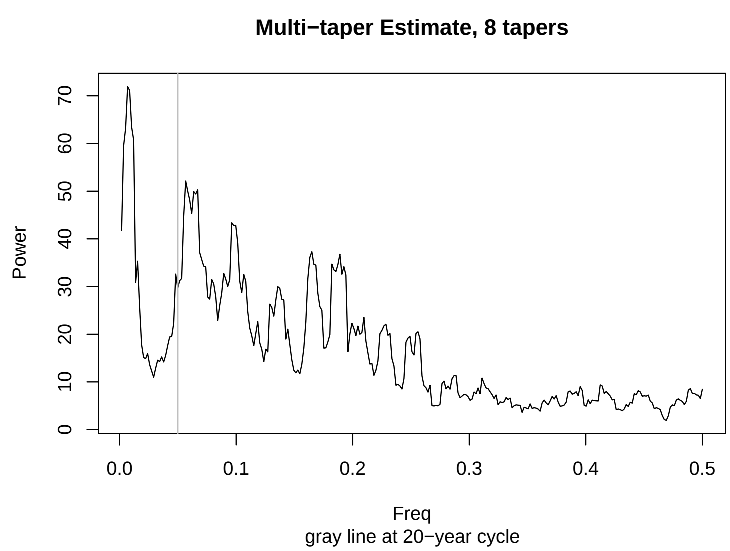
4. The prominent economist described in the question is none other than Simon Kuznets, a Nobel winner (1971) who helped establish the Statistics Department at Penn. Kuznets was interested in 20-year cycles in economic time series, but did some smoothing before starting the analysis, first forming the moving average

$$y_t = (x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2})/5 \;, \quad t = 3, \ldots, n-2. \tag{10}$$

and then differencing the series,

$$z_t = y_{t+5} - y_{t-5} \;, \quad t = 6, \ldots, n' - 5 \;. \tag{11}$$

(a) To look for evidence of a 20-year cycle in the wheat time series, consider the periodogram at frequencies near $\lambda = 1/20 = 0.05$. I used the multitaper estimates illustrated in class. These data do not have a very sharp peak in the spectral density, so leakage is not much of a problem and I can use all eight tapers (could probably use more). We could only use 6 tapers in the class example due to leakage in that illustration. The default R code pads the series when doing the FFT (after I removed the mean) to 600 cases before doing the analysis, so I used the option `fast=FALSE` to get 291 frequencies. The estimated spectral density then is the average of the 8 tapered estimates, with roughly a chi-squared distribution with 16 degrees of freedom. There's a big trough in the estimated spectral density near the 20-year frequency, with a little "bump" at this frequency.



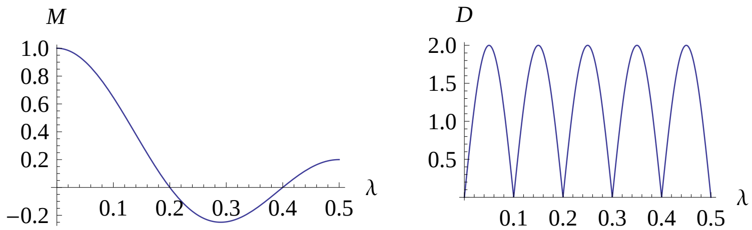
(b) Find the transfer function of the filter. Since the first smoothing filters is symmetric around 0, it is real valued:

$$M(\lambda) = \tfrac{1}{5} \sum_{t=-2}^{2} e^{2\pi\imath\,\lambda t} = 2(1+\cos 2\pi\lambda + \cos 4\pi\lambda)/5 \tag{12}$$

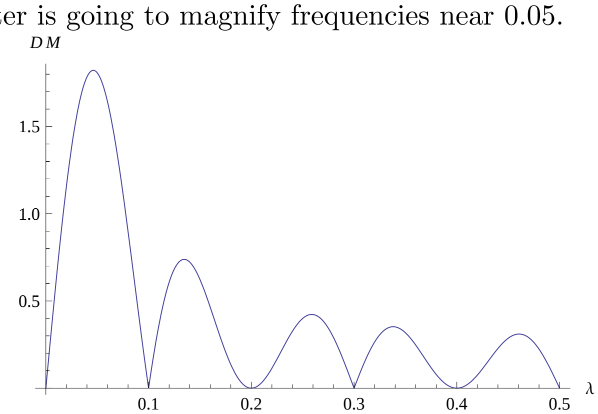
For the differencing $(x_{t+5} - x_{t-5})$ the filter is

$$D(\lambda) = 2\pi \sin 5\pi\lambda \tag{13}$$

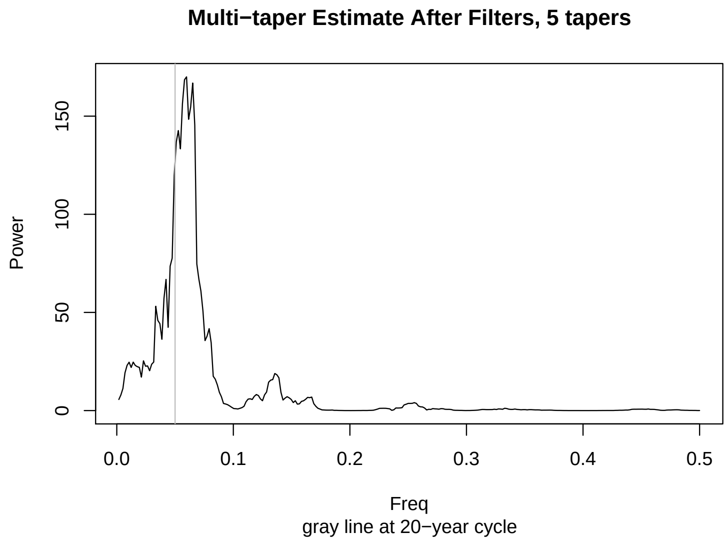
This figure shows the absolute value of the two filters (M on the left and D on the right).



The net effect of the smoothing in the frequency domain is then $M(\lambda)D(\lambda)$; you can see how this filter is going to magnify frequencies near 0.05.



(c) I again used the multi-taper estimators, this time with 5 tapers (otherwise these estimates start to obscure the "holes" introduced by the filters). The spectral density looks rather different, with some peaks and troughs where they were not before. Now you can see the small peak at the edge of the very large energy near frequency 0.06. Again, there's a little bump on the edge, but this time the edge of a very large peak introduced by the filtering. A less careful estimate will be contaminated by leakage.



So, with a good spectral estimator, I'd say that there is a little something happening at the 20-year cycle, but the filtering done by Kuznets insert a lot of power near this frequency.