ECON 6511: Advanced Applied Econometrics
Homework 2 Solutions

1. (Based on Wooldridge, Chapter 12, Problem 2) Let $\{e_t : t = -1, 0, 1, \ldots\}$ be a sequence of independent, identically distributed random variables with mean zero and variance one. Define a stochastic process by

$$x_t = e_t - \frac{1}{2}e_{t-1} + \frac{1}{2}e_{t-2}, \quad t = 1, 2, \ldots$$

   (a) Find $E(x_t)$ and $Var(x_t)$. Do either of these depend on t?
   Answer: $E(x_t) = E(e_t) - \frac{1}{2}E(e_{t-1}) + \frac{1}{2}E(e_{t-2}) = 0$ for $t = 1, 2, \ldots$. Also, because the e_t are independent, they are uncorrelated and so $Var(x_t) = Var(e_t) + \frac{1}{4}Var(e_{t-1}) + \frac{1}{4}Var(e_{t-2}) = 1 + \frac{1}{4} + \frac{1}{4} = \frac{3}{2}$ because $Var(e_t) = 1$ for all t.

   (b) Show that $Corr(x_t, x_{t+1}) = -\frac{1}{2}$ and $Corr(x_t, x_{t+2}) = \frac{1}{3}$.
   Answer: Because x_t has zero mean, $Cov(x_t, x_{t+1}) = E(x_t, x_{t+1}) = E[(e_t - \frac{1}{2}e_{t-1} + \frac{1}{2}e_{t-2})(e_{t+1} - \frac{1}{2}e_t + \frac{1}{2}e_{t-1})] = E(e_te_{t+1}) - \frac{1}{2}E(e_t^2) + \frac{1}{2}E(e_te_{t-1}) - \frac{1}{2}E(e_{t-1}e_{t+1}) + \frac{1}{4}E(e_{t-1}e_t) - \frac{1}{4}E(e_{t-1}^2) + \frac{1}{2}E(e_{t-2}e_{t+1}) - \frac{1}{4}E(e_{t-2}e_t) + \frac{1}{4}E(e_{t-2}e_{t-1}) = -\frac{1}{2}E(e_t^2) - \frac{1}{4}E(e_{t-1}^2) = -\frac{1}{2} - \frac{1}{4} = -\frac{3}{4}$; the third to last equality follows because the e_t are pairwise uncorrelated and $E(e_t^2) = 1$ for all t. $Corr(x_tx_{t+1}) = -\frac{3}{4}/\frac{3}{2} = -\frac{1}{2}$.
   Computing $Cov(x_t, x_{t+2})$ is even easier because only one of the nine terms has an expectation different from zero: $\frac{1}{2}E(e_t^2) = \frac{1}{2}$. Therefore, $Corr(x_t, x_{t+2}) = \frac{1}{2}/\frac{3}{2} = \frac{1}{3}$.

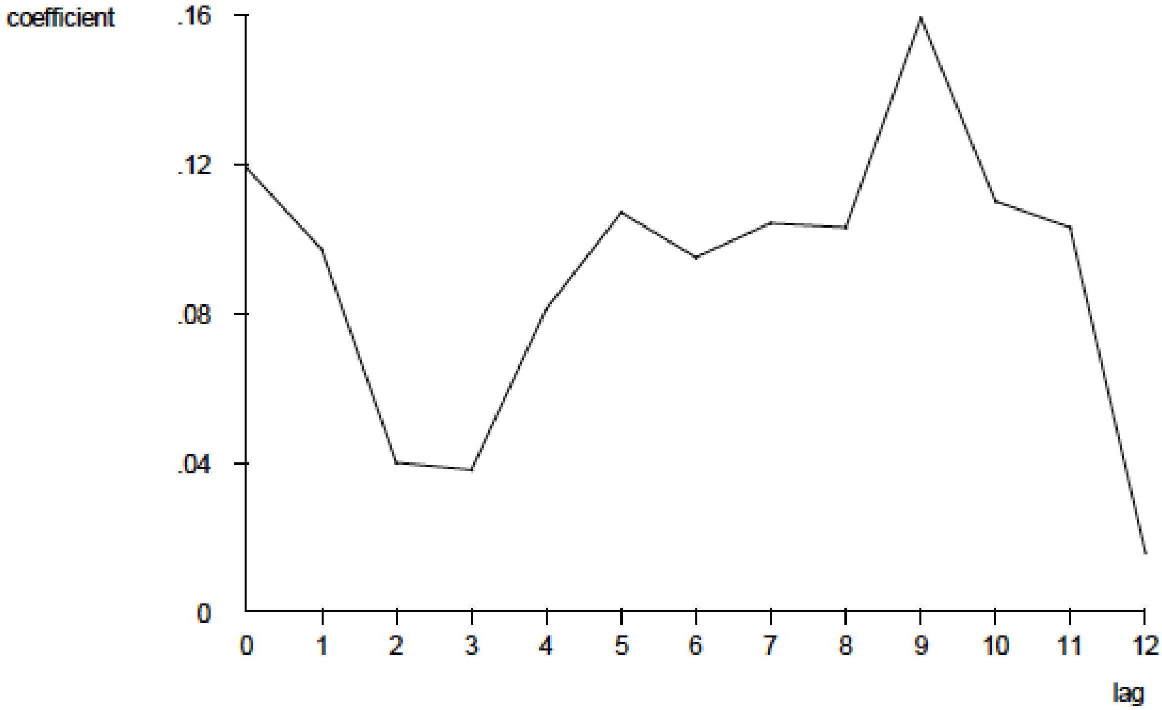
   (c) What is $Corr(x_t, x_{t+h})$ for $h > 2$?
   Answer: $Corr(x_t, x_{t+h}) = 0$ for $h > 2$ because, for $h > 2$, x_{t+h} depends at most on e_{t+j} for $j > 0$, while x_t depends on e_{t+j}, $j \leq 0$.

   (d) Is $\{x_t\}$ covariance stationary? Is it weakly dependent?
   Answer: Both $E(x_t)$ and $Var(x_t)$ are constant, and $Cov(x_t, x_{t+h})$ depends only on h and not t, so the process is covariance stationary. It is also weakly dependent since $Corr(x_t, x_{t+h})$ goes to 0 as h increases.

2. (Based on Wooldridge, Chapter 12, Problem 5) For the U.S. economy, let *gprice* denote the monthly growth in the overall price level and let *gwage* be the monthly growth in hourly wages. These are both obtained as differences of logarithms: $gprice = \Delta\log(price)$ and $gwage = \Delta\log(wage)$.

   (a) Using the monthly data in WAGEPRC.dta, estimate a distributed lag model:

$$\begin{aligned} gprice =& \alpha + \beta_1 gwage_t + \beta_2 gwage_{t-1} + \beta_3 gwage_{t-2} + \beta_4 gwage_{t-3} + \beta_5 gwage_{t-4} \\ &+ \beta_6 gwage_{t-5} + \beta_7 gwage_{t-6} + \beta_8 gwage_{t-7} + \beta_9 gwage_{t-8} + \beta_{10} gwage_{t-9} \\ &+ \beta_{11} gwage_{t-10} + \beta_{12} gwage_{t-11} + \beta_{13} gwage_{t-12} + u_t \end{aligned}$$



and sketch the estimated lag distribution. At what lag is the effect of *gwage* on *gprice* largest? Which lag has the smallest coefficient?

Answer: The graph gives the estimated lag distribution.The largest effect is at the ninth lag, which says that a temporary increase in wage inflation has its largest effect on price inflation nine months later. The smallest effect is at the twelfth lag, which hopefully indicates (but does not guarantee) that we have accounted for enough lags of *gwage* in the FLD model.

(b) For which lags are the t statistics less than two?

Answer: Lags two, three, and twelve have t statistics less than two. The other lags are statistically significant at the 5% level against a two-sided alternative. (We are assuming the CLM assumptions hold or the modifications from Lecture 2.)

(c) What is the estimated long-run propensity? Is it much different than one? Explain what the LRP tells us in this example.

Answer: The estimated LRP is just the sum of the lag coefficients from zero through twelve: 1.172. While this is greater than one, it is not a lot greater, and the difference from unity could be due to sampling error.

(d) What regression would you run to obtain the standard error of the LRP directly?

Answer: The model underlying and the estimated equation can be written with intercept α and lag coefficients $\beta_1, \beta_2, \ldots, \beta_{13}$. Denote the LRP by $\theta = \beta_1 + \beta_2 + \ldots + \beta_{13}$. Now, we can write $\beta_1 = \theta - \beta_2 - \beta_3 - \ldots - \beta_{13}$. If we plug this into the FDL model we obtain:

$$\begin{aligned} gprice_t =& \alpha + (\theta - \beta_1 - \beta_2 - \ldots - \beta_{13})gwage_t + \beta_2 gwage_{t-1} + \ldots + \beta_{13} gwage_{t-12} + u_t \\ =& \alpha + \theta gwage_t + \beta_2(gwage_{t-1} - gwage_t) + \beta_3(gwage_{t-2} - gwage_t) + \ldots \\ &+ \beta_{13}(gwage_{t-12} - gwage_t) + u_t \end{aligned}$$

Therefore, if we estimate this equation, we can obtain the coefficient and standard error on $gwage_t$ to estimate the LRP and its standard error.

3. Use the data in PHILLIPS.dta for this exercise.

   (a) Estimate an AR(1) model for the unemployment rate. Use this equation to predict the unemployment rate for 2004. Compare this with the actual unemployment rate for 2004. (You can find this information in a recent *Economic Report of the President.*)
   Answer: The estimated AR(1) model is:

   $$unem_t = 1.49 + 0.742 unem_{t-1}$$

   In 2003, the unemployment rate was 6.0, so the predicted unemployment rate for 2004 is $1.49 + 0.742(6) \approx 5.94$. From the 2005 *Economic Report of the President*, the U.S. civilian unemployment rate was 5.5. Therefore, the equation overpredicts the 2004 unemployment rate by almost half a percentage point.

   (b) Add a lag of inflation to the AR(1) model from part (a). Is inf_{t-1} statistically significant?
   Answer: When we add inf_{t-1} to the equation we get

   $$unem_t = 1.30 + 0.649 unem_{t-1} + 0.183 inft_{t-1}$$

   Lagged inflation is very statistically significant, with a t statistic of about 4.7.

   (c) Use the equation from part (b) to predict the unemployment rate for 2004. Is the result better or worse than in the model from part (a)?
   Answer: To use the equation from part (b) to predict unemployment in 2004, we also need the inflation rate for 2003. This is given in PHILLIPS.data as 2.3. Therefore, the prediction of *unem* in 2003 is $1.30 + 0.649(6) + 0.183(2.3) \approx 5.61$. While still too large, it is pretty close to the actual rate of 5.5 percent, and it is certainly better than the prediction from part (a).

4. (Wooldridge, Chapter 12, C2)

(a) Using the data in WAGEPRC.dta, estimate the distributed lag model from Problem 2 above. Regress $\hat{u}_t$ on $\hat{u}_{t-1}$ to test for AR(1) serial correlation.

Answer: After estimating the FDL model by OLS, we obtain the residuals and run the regression $\hat{u}_t$ and $\hat{u}_{t-1}$, using 272 observations. We get $\hat{\rho} \approx 0.503$ and $t_{\hat{\rho}} \approx 9.60$, which is very strong evidence of positive AR(1) correlation.

(b) Reestimate the model using iterated Cochrane-Orcutt estimation. What is your new estimate of the long-run propensity?

Answer: When we estimate the model by iterated C-O, the LRP is estimated to be about 1.110.

(c) Using iterated CO, find the standard error for the LRP. Determine whether the estimated LRP is statistically different from one at the 5% level.

Answer: We use the same trick from Problem 2 above, except now we estimate the equation by iterated C-O. In particular, write

$$\begin{aligned} gprice_t =& \alpha + (\theta - \beta_1 - \beta_2 - \ldots - \beta_{13}) gwage_t + \beta_2 gwage_{t-1} + \ldots + \beta_{13} gwage_{t-12} + u_t \\ =& \alpha + \theta gwage_t + \beta_2(gwage_{t-1} - gwage_t) = \beta_3(gwage_{t-2} - gwage_t) + \ldots \\ & + \beta_{13}(gwage_{t-12} - gwage_t) + u_t \end{aligned}$$

Where θ is the LRP and $\{u_t\}$ is assumed to follow an AR(1) process. Estimating this equation by C-O gives $\hat{\theta} \approx 1.110$ and $se(\hat{\theta}) \approx 0.191$. The t statistic for testing $H_0 : \theta = 1$ is $(1.110 - 1)/0.191 \approx 0.58$, which is not close to being significant at the 5% level. So the LRP is not statistically different from one.