BOSTON COLLEGE
Department of Economics
EC 228 Econometrics, Prof. Baum, Mr. Barbato, Spring 2003
Problem Set 7.
Answer sheet

Problem 12.14

(i) This is the model that was estimated in part (vi) of Computer Exercise 10.17. After getting the OLS residuals, $\widehat{\mu_t}$, we run the regression $\widehat{\mu_t}$ on $\widehat{\mu_{t-1}}$, t = 2,...,108. (Included an intercept, but that is unimportant.) The coefficient on $\widehat{\mu_{t-1}}$ is $\widehat{\rho} = .281(se = .094)$. Thus, there is evidence of some positive serial correlation in the errors (t ≈ 2.99). A strong case can be made that all explanatory variables are strictly exogenous. Certainly there is no concern about the time trend, the seasonal dummy variables, or *wkends*, as these are determined by the calendar. It seems safe to assume that unexplained changes in *prcfat* today do not cause future changes in the state-wide unemployment rate. Also, over this period, the policy changes were permanent once they occurred, so strict exogeneity seems reasonable for *spdlaw* and *beltlaw*. (Given legislative lage, it seems unlikely that the dates the policies went into effect had anything to do with recent, unexplained changes in *prcfat*.

(ii) Remember, we are still estimating the β_j by OLS, but we are computing different standard errors that have some robustness to serial correlation. Using Stata 7.0, I get $\widehat{\beta}_{spdlaw} = .0671, se(\widehat{\beta}_{spdlaw}) = .0267$ and $\widehat{\beta}_{beltlaw} = -.0295, se(\widehat{\beta}_{beltlaw}) = .0331$. The *t* statistic fro *spdlaw* has fallen to about 2.5, but it is still significant. Now, the *t* statistic on *beltlaw* is less than one in absolute value, so there is little evidence that *beltlaw* had an effect on *prcfat*.

(iii) For brevity, I do not report the time trend and monthly dummies. The final estimate of ρ is $\widehat{\rho} = .289$:

$$\underset{}{prcfat} = \underset{(.102)}{1.009} + \ldots + \underset{(.00500)}{.00062}\ wkends - \underset{(.0055)}{.0132}\ unem + \underset{(.0268)}{.0641}\ spdlaw - \underset{(.0301)}{.0248}\ beltlaw$$

$$n = 108,\ R^2 = .641$$

There are no drastic changes. Both policy variable coefficients get closer to zero,and the standard errors are bigger that the incorrect OLS standard errors [and, coincidentally, pretty close to the Newey-West standard errors for OLS from part (ii)]. So the basic conclusion is the same: the increase in the speed limit appeared to increase *prcfat*, does not have a statistically significant effect.

Problem 12.15

(i) Here are the OLS regression results:

$$\widehat{\log(avgprc)} = -\underset{(.115)}{.073} - \underset{(.0014)}{.0040}\ t - \underset{(.1294)}{.0101}\ mon - \underset{(.1273)}{.0088}\ tues + \underset{(.1257)}{.0376}\ wed + \underset{(.1257}{.0906}\ thurs$$

$$n = 97,\ R^2 = .086$$

The test for joint significance of the day-of-the-week dummies is F = .23, which gives *p*-value = .92. So there is no evidence that the average price of fish varies systematically within a week.

(ii) The equation is

$$\log(avgprc) = -\underset{(.190)}{.920} - \underset{(.0014)}{.0012}\ t - \underset{(.1141)}{.0182}\ mon - \underset{(.11121)}{.0085}\ tues + \underset{(.1117)}{.0500}\ wed + \underset{(.1110)}{.1225}\ thurs$$
$$+ \underset{(.0218)}{.0909}\ wave2 + \underset{(.0208)}{.0474}\ wave3$$

$$n = 97,\ R^2 = .310$$

Each of the wave variables is statistically significant, with *wave*2 being the most important. Rough seas (as measured by high waves) would reduce the supply of fish (shift the supply curve back), and this would result in a price increase. One might argue that bad weather reduces the demand for fish at a market, too, but that would reduce price. If there are demand effects captured by the wave variables, they are being swamped by the supply effects.

(iii) The time trend coefficient becomes much smaller and statistically insignificant. We can use the omitted variable bias table from Chapter 3, Table 3.2 (page 92) to determine what is probably going on. Without *wave*2 and *wave*3, the coefficient on *t* seems to have a downward bias. Since we know the coefficients on *wave*2 and *wave*3 are positive, this means the wave variables are negatively correlated with *t*. In other words, the seaswere rougher, on average, at the beginning of the sample period. (You can confirm this by regressing *wave*2 on*t* and *wave*3 on *t*.)

(iv) The time trend and daily dummies are clearly strictly exogenous, as they are just functions of time and the calendar. Further, the height of the waves is not influenced by past unexpected changes in log(*avgprc*).

(v) We simply regress the OLS residuals on one lag, getting $\widehat{\rho} = .618, se(\widehat{\rho}) = .081, t_{\widehat{\rho}} = 7.63$. Therefore, there is strong evidence of positive serial correlation.

(vi) The Newey-West standard errors are $se(\widehat{\beta}_{wave2}) = .0234$ and $se(\widehat{\beta}_{wave3}) = .0195$. Given the significant amount of AR(1) serial correlation in part (v), it is somewhat surprising that these standard errors are not much larger compared with the usual, incorrect standard errors. In fact, the Newey-West standard error for $\widehat{\beta}_{wave3}$ is actually smaller than the OLS standard error.

(vii) The Prais-Winsten estimates are

$$\begin{aligned}\log(avgprc) = & -\underset{(.239)}{.658} - \underset{(.0029)}{.0007}\ t + \underset{(.0652)}{.0099}\ mon + \underset{(.0744)}{.0025}\ tues + \underset{(.0746)}{.0624}\ wed + \underset{(.0621)}{.1174}\ thurs + \underset{(.0174)}{.0497}\ wave2 \\ & + \underset{(.0174)}{0323}\ wave3 \\ & n = 97,\ R^2 = .135\end{aligned}$$

The coefficient on *wave*2 drops by a nontrivial amount, but it still has a *t* statistic of almost .3. The coefficient on *wave*3 drops by a relatively smaller amount, but its *t* statistic (1.86) is borderline significant. The final estimate of ρ is about .687.

Problem 15.4

(i) The state may set the level of its minimum wage at least partly based on past or current economic activity, and this could certainly be part of μ_t. Then $gMIN_t$ and μ_t are correlated, which causes OLS to be biased and inconsistent.

(ii) Because $gGDP_t$ controls for the overall performance of the U.S. economy, it seems reasonable that $gUSMIN_t$ is uncorrelated with the disturbances to employment growth for a particular state.

(iii) In some years, the U.S. minimum was will increase in such a way so that it exceeds the state minimum wage, and then the state minimum wage will also increase. Even if the U.S. minimum wage is never binding, it may be that the state increases its minimum wage in response to an increase in the U.S. minimum. If the state minimum is always the U.S. minimum, then $gMIN_t$ is exogenous in this equation and we would just use OLS.

Problem 15.6

(i) Plugging (15.26) into (15.22) and rearranging gives

$$\begin{aligned}\gamma_1 &= \beta_0 + \beta_1(\pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2) + \beta_2 z_1 + \mu_1 \\ &= (\beta_0 + \beta_1\pi_0) + (\beta_1\pi_1 + \beta_2)z_1 + \beta_1\pi_2 z_2 + \mu_1 + \beta_1 v_2,\end{aligned}$$

and so $\alpha_0 = \beta_0 + \beta_1\pi_0$, $\alpha_1 = \beta_1\pi_1 + \beta_2$, and $\alpha_2 = \beta_1 v_2$.

(ii) From the equation in part (i), $v_1 = \mu_1 + \beta_1 v_2$

(iii) By assumption, μ_1 has zero mean and is uncorrelated with z_1 and z_2, and v_2 has these properties by definition. So v_1 has zero mean and is uncorrelated with z_1 and z_2, which means that OLS consistently estimates the α_j. [OLS would only be unbiased if we add the stronger assumptions

$E[\mu_1|z_1, z_2] = E[v_2|z_1, z_2] = 0.]$

Problem 15.15

(i) The equation estimated by OLS, omitting the first observation, is

$$\widehat{i3}_t = \underset{(0.47)}{2.37} + \underset{(.091)}{.692} \text{ inf }_t$$

$$n = 48,\ R^2 = .555.$$

(ii) The IV estimates, where inf_{t-1} is an instrument for inf_t, are

$$\widehat{i3}_t = \underset{(0.65)}{1.50} + \underset{(.143)}{.907} \text{ inf }_t$$

$$n = 48,\ R^2 = .501$$

The estimate on inf_2 is no longer statistically different from one. (If $\beta_1 = 1$, then one percentage point increase in inflation leads to a one percentage point increase in the three-month T-bill rate.)

(iii) In first differences, the equation estimated by OLS is

$$\Delta\widehat{i3}_t = \underset{(.186)}{.105} + \underset{(.073)}{.211} \Delta \text{inf }_t$$

$$n = 48,\ R^2 = .154$$

This is much lower estimate than in part (i) or part (ii).

(iv) If we regress Δinf_t on Δinf_{t-1} we obtain

$$\widehat{\Delta\text{inf}}_t = \underset{(.325)}{.088} + \underset{(.1266)}{.0096} \Delta \text{inf }_{t-1}$$

$$n = 47,\ R^2 = .0001$$

Therefore, Δinf_t and Δinf_{t-1} are virtually uncorrelated, which means that Δinf_{t-1} cannot be used as an IV for Δinf_t.

Problem 15.17

(i) Sixteen states executed at least one prisoner in 1991, 1992, or 1993. (That is, for 1993, *exec* is greater than zero for 16 observations.) Texas had by far the most executions with 34.

(ii) The results of the pooled OLS regression are

$$\widehat{mrdrte} = -5.28 - 2.07d93 + .128exec + 2.53unem$$

$$n = 102,\ R^2 = .102,\ \overline{R^2} = .074$$

The positive coefficient on *exec* is no evidence of a deterrent effect. Statistically, the coefficient is not different from zero. The coefficient on *unem* implies that higher unemployment rates are associated with higher murder rates.

(iii) When we difference (and use only the changes from 1990 to 1993), we obtain

$$\Delta\widehat{mrdrte} = \underset{(.209)}{.413} - \underset{(.043)}{.104}\ \Delta exec - \underset{(.159)}{.067}\ \Delta unem$$

$$n = 51,\ R^2 = .110,\ \overline{R^2} = .073$$

The coefficient on $\Delta exec$ is negative and statistically significant (p-value $\approx .02$ against a two-sided alternative), suggesting a deterrent effect. One more execution reduces the murder rate by about .1 so 10 more executions reduce the murder rate by one (which means one murder per 100,000 people). The unemployment rate variable is no longer significant.

(iv) The regression $\Delta exec$ on $\Delta exec_{-1}$ yields

$$\widehat{\Delta exec} = \underset{(.370)}{.350} - \underset{(0.17)}{1.08}\ \Delta exec_{-1}$$

$$n = 51,\ R^2 = .456,\ \overline{R^2} = .444$$

which shows a strong negative correlation in the change in executions. This means that, apparently, states follow policies whereby if executions were high in the preceeding three-year period, they are lower, one-for-one, in the next three-year period.

Technically, to test the identification condition, we should add $\Delta unem$ to the regression. But its coefficient is small and statistically very insignificant, and adding it does not change the outcome at all.

(v) When the differenced equation is estimated using Δexec_{-1}as an IV for $\Delta exec$, *we obtain*

$$\widehat{\Delta mrdrte} = \underset{(.211)}{.411} - \underset{(.064)}{.100}\ \Delta exec - \underset{(.159)}{.067}\ \Delta unem$$

$$n = 51,\ R^2 = .110,\ \overline{R^2} = .073$$

This is very similar to when we estimate the differenced equation by OLS. Not surprisingly, the most important change is that the standard error on $\widehat{\beta}_1$ is now larger and reduces the statistical significance of $\widehat{\beta}_1$

Problem 16.1

(i) If $\alpha_1 = 0$ then $\gamma_1 = \beta_1 z_1 + \mu_1$, and so the right-hand-side depends only on the exogenous variable z_1 and the error term μ_1. This then is the reduced form for γ_1. If $\alpha_1 = 0$, the reduced form for γ_1 is $\gamma_1 = \beta_2 z_2 + \mu_2$. (Note that having both α_1 and α_2 equal zero is not as interesting as it implies the bizarre condition $\mu_2 - \mu_1 = \beta_1 z_1 - \beta_2 z_2$.)

If $\alpha_1 \neq 0$ and $\alpha_2 = 0$, we can plug $\gamma_1 = \beta_2 z_2 + \mu_2$ into the first equation and solve for γ_2 :

$$\beta_2 z_2 \mu_2 = \alpha_1 \gamma_2 + \beta_1 z_1 + \mu_1 \qquad \text{or}$$

$$\alpha_1 \gamma_2 = \beta_1 z_1 - \beta_2 z_2 + \mu_1 - \mu_2$$

Dividing by α_1(because $\alpha_1 \neq 0$)*gives*

$$\gamma_2 = (\beta_1/\alpha_1)z_1 - (\beta_2/\alpha_1)z_2 + (\mu_1 - \mu_2)/\alpha_1$$

$$\equiv \pi_{21} z_1 + \pi_{22} z_2 + v_2,$$

where $\pi_{21} = \beta_1/\alpha_1, \pi_{22} = -\beta_2/\alpha_1$ and $v_2 = (\mu_1 - \mu_2)/\alpha_1$. Note that the reduced form for γ_2 generally depends on z_1 and z_2 (as well as on μ_1 and μ_2).

(ii) If we multiply the second structural equation by (α_1/α_2) and subtract it from the first structural equation, we obtain

$$\gamma_1 - (\alpha_1/\alpha_2)\gamma_1 = \alpha_1\gamma_2 - \alpha_1\gamma_2 + \beta_1 z_1 - (\alpha_1/\alpha_2)\beta_2 z_2 + \mu_1 - (\alpha_1/\alpha_2)\mu_2$$

$$= \beta_1 z_1 - (\alpha_1/\alpha_2)\beta_2 z_2 + \mu_1 - (\alpha_1/\alpha_2)\mu_2$$

or

$$[1 - (\alpha_1/\alpha_2)]\gamma_1 = \beta_1 z_1 - (\alpha_1/\alpha_2)\beta_2 z_2 + \mu_1 - (\alpha_1/\alpha_2)\mu_2$$

Because $\alpha_1 \neq \alpha_2, 1 - (\alpha_1/\alpha_2) \neq 0$, and so we can divide the equation by $1 - (\alpha_1/\alpha_2)$ to obtain the reduced form for $\gamma_1 : \gamma_1 = \pi_{11} z_1 + \pi_{12} z_2 + v_1$, where $\pi_{11} = \beta_1/[1 - (\alpha_1/\alpha_2)], \pi_{12} = -(\alpha_1/\alpha_2)\beta_2/[1 - (\alpha_1/\alpha_2)]$, and $v_1 = [\mu_1 - (\alpha_1/\alpha_2)\mu_2]/[1 - (\alpha_1/\alpha_2)]$.

A reduced form does not exist for γ_2, as can be seen by subtracting the second equation from the first:

$$0 = (\alpha_1 - \alpha_2)\gamma_2 + \beta_1 z_1 - \beta_2 z_2 + \mu_1 - \mu_2;$$

because $\alpha_1 \neq \alpha_2$, we can rearrange and divide by $\alpha_1 - \alpha_2$ to obtain the reduced form.

(iii) In supply and demand examples, $\alpha_1 \neq \alpha_2$ is very reasonable. If the first equation is the supply function, we generally expect $\alpha_1 > 0$, and if the second equation is the demand function, $\alpha_2 < 0$. The reduced forms can exist even in cases where the supply function is not upwardsloping and the demand function is not downward sloping, but we might question the usefulness of sucg models.

Problem 16.7

(i) Attendance at women's basketball may grow inways that are unrelated to factors that we can observe and control for. The taste for women's basketball may increase over time, and this would be captured by the time trend.

(ii) No. The university sets the price, and it may change price based on expectations of next years's attendance; if the university uses factors that we cannot observe, these are necessarily in the error term μ_2. So even though the supply is fixed, it does not mean that price is uncorrelated with the unobservables affecting demand.

(iii) If people only care about how this year's team is doing, $SEASPERC_{t-1}$ can be excluded from the equation once $WINPERC_t$ has been controlled for. Of course, this is not a very good assumption for all games, as attendance early in the season is likely to be related to how the team did last year. We eould also need to check that $1PRICE_t$ is partially correlated with $SEASPERC_{t-1}$ by estimating the reduced form for $1PRICE_t$.

(iv) It does make sense to include a measure of men's basketball ticket prices, as attending a women's basketball game is a substitute for attending a men's game. The coefficient on $1MPRICE_t$ would be expected to be negative. The winning percentage of the men's team is another good candidate for an explanatory variable in the women's demand equation.

(v) It might be better to use first differences of the logs, which are then growth rates. We would then drop the observation for the first game in each season.

(vi) If a game is sold out, we cannot observe true demand for that game. We only know that desired attendance is some number above capacity. If we just plug in capacity, we are understanding the actual demand for tickets. (Chapter 17 discusses censored regression methods can be used in such cases.)