Problemløsninger PS3 – BA-BMECV1031U

1 (a) The simple OLS regression suggests that the coefficient on *cunem* is statistically different from minus 2. The Breusch Godfrey LM test of order 1 serial correlation gives a X^2 statistic of 0.081 and p-value of 0.78. Thus, there is no evidence for first order serial correlation (at 8% level).

(b) The simple regression $\hat{u}_t^2$ on $\Delta unem_t$ gives a slope coefficient of about .168 with $t = 1.59$, and so there is little evidence for heteroscedasticity (p-value 0.11%). [The F-test comes to the same conclusion.] The variance of the error does not appear to be larger when the change in unemployment is larger.

(c) The heteroskedasticity-robust standard error is about .213 and the Newey-West standard error is 0.237, compared with the usual OLS standard error of .165. So, the robust standard errors are more than 30% larger than the usual OLS one. Of course, a larger standard error leads to a wider confidence interval for β_1. As a result the coefficient on *cunem* is no longer statistically different from -2 (-2 is contained in the 95% confidence intervals).

These somewhat contradicting consequences of the standard error corrections and the findings of the tests in parts a) and b) can be explained by the small sample size (less than 100 observations). The tests conducted are large sample tests.

2 We can reason this from the equation 12.4 on page 395 in the textbook, because the usual OLS standard error is an estimate of $\sigma/\sqrt{SST_x}$. When the dependent and independent variables are in level (or log) form, the AR(1) parameter, ρ, tends to be positive in time series regression models. Further, the independent variables tend to be positive correlated, so $(x_t - \bar{x})(x_{t+j} - \bar{x})$ – which is what generally appears in the equation for var(beta_1) when the $\{x_t\}$ do not have zero sample average – tends to be positive for most *t* and *j*. With multiple explanatory variables the formulas are more complicated but have similar features.

If $\rho < 0$, or if the $\{x_t\}$ is negatively autocorrelated, the second term in in the relevant equation could be negative, in which case the true standard deviation of $\hat{\beta}_1$ is actually less than

$\sigma/\sqrt{SST_x}$.

3 (a) There is substantial serial correlation in the errors of the equation, and the OLS standard errors often underestimate the true standard deviation in $\hat{\beta}$. This makes the usual confidence interval for β and *t* statistics invalid.

(b) We can use the Newey-West method to obtain an approximately valid standard error or we could use a differenced version of the model because the estimated rho is quite close to +1.

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

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

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

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

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

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

This is a much lower estimate than in part (a) or part (b).
[Remark: An explanation for the systematic differences in results is that *inf* in the level model of part (a) is not exogenous (correlation with the error term) and the differencing of part (b) can mitigate this issue. For this reason, panel and time series models are often estimated in differences. Something that you will see and understand in follow up courses on Panel or Time Series Econometrics.]

5 a) The average value for $unem98$ is 0.172. The average value for $unem96$ is 0.312. Thus, a much lower fraction of workers was unemployed in 1998 as compared to 1995 (the year in which $unem96$ measures). This may be due to job training, but there are many other possible explanations for this (e.g. an improved job market in 1998 compared to 1995).

(b) We estimate $\widehat{unem}98 = \underset{(0.014)}{0.163} + \underset{(0.024)}{0.026} train$. This suggests that participation in the training program increases the likelihood of being unemployed in 1998. Of course, it is more likely that we have reverse causality and those who are unemployed are more likely to participate in job training. In any event, the estimate is not significantly different from 0.

(c) The estimated training effect is now -0.121 and it is statistically significant at the 1% level (standard error = 0.024). This estimate suggests that participating in job training reduces the likelihood of being unemployed, as we would hope. The change in sign and significance is due to the inclusion of $earn96, educ, age,$ and $married$. These variables can help control for non-random differences between those people who participate in job training and those who do not. We hope that conditioning upon these variables gets us closer to random assignment for training, allowing for identification of the causal effect of job training on unemployment.

(d) The estimated average treatment effect increases in magnitude to -0.123. This is a small change from the estimate in part c. The standard error estimate is 0.030, which is also larger than what was estimated in part c, but does not change the results of any significance tests.

(e) The interaction terms are not jointly significant, with $F_{4,1120} = 0.43$ when testing the null hypothesis that the coefficients on the interactions are all equal to 0. We fail to reject the null.

(f) Obtain $\widehat{unem98}_1$ by regressing $unem98$ on $earn96, educ, age, married$ for observations for which $train = 1$. Run the same regression for $train = 0$ to obtain $\widehat{unem98}_0$. Calculate $\hat{\tau}_{ura} = n^{-1}\sum\left(\widehat{unem98}_{1i} - \widehat{unem98}_{0i}\right) = -0.123$, which is precisely the ATE we estimated in part d. The regression in part d is much more convenient in obtaining a standard error.