

Problem set IV

In this assignment we use the Card-Krueger minimum wage data. In lecture 11 the final equation is a model for a conditional dif-in-dif estimator. This model has a full set of chain dummies but no intercept.

- a. Estimate the parameters of this model and compare the results with those reported in lecture 11. Explain any differences.

In the model we assume a random component error term with 8 clusters defined by chain and treatment status. We want to compute the clustered standard errors. Convince yourself that this is a model **with only cluster specific independent variables**, so that you can apply the relevant formulas from lecture 11.

- b. Compute the number of observations per cluster and the the variance of the cluster specific error and the idiosyncratic error. Estimate the within cluster correlation of the random errors.
- c. Compute the OLS correction factor based on the average cluster size and use it to compute corrected standard errors.
- d. Use the variance formula in lecture 11 to compute the standard errors and compare the OLS, the corrected OLS and the correct standard errors.
- e. Estimate the model in a. but adding treatment and chain dummy interaction terms. Can you estimate the clustered standard errors in this model?
- f. A possible alternative to clustering is the bootstrap. Use the nonparametric bootstrap to estimate the standard errors in the model in a. and compare your results. Note that in STATA bootstrap standard errors are an option (see p. 1514 of the STATA manual where the VCE options for the REGRESS command are mentioned; see also Chapter 20 for further explanation).

Some remarks on using STATA. By now all of you will be using DO-files of STATA commands as explained in Chapter 13 of the Get Started Manual. This assignment involves computations using STATA results. There are many ways to do the same calculation in STATA, so this is just a suggestion. You may have found a more convenient way to get the same result.

First, you have to create a new variable that contains the OLS residuals of the regression in a. This can be done with the command **predict residuals, r** where residuals is the new variable with the OLS residuals. You also may want to create the **matrix Z**. The columns of this matrix are the 8 vectors of observations

of the cluster indicators. These indicators themselves can be generated as the product of the Penn/New Jersey dummies and the chain dummies. Check that the sum of the columns is equal to 1. To create this matrix you can use the horizontal concatenation explained on p. 251 of the Programming Manual.

In b. you can use the Z matrix to calculate the \tilde{e} vector on p. 14 of lecture 11. The two formulas below that expression can be used directly to get the answer to b. Finally the calculation of $Var(\hat{\theta})$ in d. is another matrix calculation. Matrix calculations are further explained in Chapter 14 of the User Guide.