Ph.D. course: Applied Econometrics for Researchers

Workshop 2

1. Stata introduction: Tutorials, installation, data sets, log-file, do-file, help files.

Stata tutorials are can be found on the Stata webpage:

http://www.stata.com/links/video-tutorials/

For Workshop 2, you might find the following tutorials helpful:

Simple linear regression in Stata

Basic scatterplots in Stata

Histograms in Stata

Descriptive statistics in Stata

- 2. **Motivation:** We continue modeling the relationship between a firm's product innovation and its use of external sources of knowledge. We also consider the potential role of internal R&D and ask if there are moderating effects of the use of internal R&D.
- 3. **Data:**

Use the do-file WORKSHOP2_INITIAL.do to build up your program. This file will merge three different data files and run the first part of the analysis.

4. Variables:

Introduced in the previous workshop. See the UK CIS questionnaire for more information.

5. Basic regression (without interaction term):

Re-run the linear regression from Workshop 1 with *prodnew* as DV and *extsource* and *rdintpct* as explanatory variables controlling for *inconst* and *lempl00*.

- 5.1. Are the controls jointly significant?
- 5.2. Comment on the goodness-of-fit of the regression.
- 5.3. Do a test to see if the regressors included (the explanatory variables and the controls) jointly explain a significant part of the variation in the DV.

The distinction between explanatory variables and controls is not always clear-cut. They enter the model as regressors in a completely parallel fashion. Explanatory variables are those on which we hypothesize theoretically.

Controls are mainly included in the model to make sure that the explanatory variables are not picking up "other effects". For example, firms of different sizes likely differ in their use of external knowledge sources. Moreover, firm size is also potentially related to innovation (our dependent variable). We therefore want to control for size when we determine the effect of external knowledge sources on innovation, trying to keep "all else equal."

- 5.4. Do you see evidence of such size effects? *Hint: Calculate the correlation matrix between the regressors of the model.* Are there any other "large" correlations between the regressors?
- 5.5. Run an OLS regression for a restricted model that leaves out *inconst* and *lempl00*. *Hint:* To make sure that both regressions use the same sample of observations, run the extended model and "fix" the sample as follows:

```
g insample = e(sample)
```

reg prodnew extsource rdintpct if insample

- 5.6. What happens to the coefficient of *extsource* compared to the previous model? Is this what you would expect?
- 5.7. Use the formula for the F-test (formula 4.37 in Wooldridge) to manually re-calculate the F-test of the restricted model against the unrestricted model that we considered above. Check that you get the same answer.

If there is any doubt that the error term of the model is homoscedastic, one can base the tests on robust test-statistics.

5.8. Use the robust option with regress and re-do the above tests. Do you get different conclusions? In that case, which test would you prefer?

6. **Regression with interaction term:**

- 5.9. Consider *rdintpct* as a potential <u>moderator</u> and check whether it significantly moderates the relationship between *extsource* and *prodnew*. Generate the interaction between *extsource* and *rdintpct* and enter it as an extra variable in the regression model. Keep the two controls in the model as well.
- 5.10. Use this regression to answer the following research question:

"Does internal R&D moderate the impact of external sources of knowledge on innovative performance?"

- 5.11. What hypothesis test do you consider?
- 5.12. Would your answer be different if the question were stated as:
 - 6.1. "Does internal R&D *positively* moderate the impact of external sources of knowledge on innovative performance?"
- 5.13. What hypothesis test do you consider? *Hint: What is the alternative hypothesis for this test?*
- 5.14. Do results change if you center the variables *rdintpct* and *extsource* around their means? How do you read the results in this case? *Hint: Use the command "keep if e(sample)* == 1" *to make sure that you are demeaning the variables over the regression sample.*
- 5.15. Run the interacted model using Stata notation ## for automatically generating the main and interaction effects and indicating c. for the kind of variables you are dealing with. Compare to your previous results.
- 5.16. Use the Stata command margins to calculate the marginal effect of external sources of knowledge on innovative performance. Illustrate by the Stata command marginsplot how this effect depends on internal R&D.
- 5.17. What is your answer for the above research question?

>>>>>	Follow-up on question 6 in class	<<<<<<<<
-------	----------------------------------	----------