Professor: Pam Paxton
TA: Nicholas Reith

SOC 385L Lab Assignment 8

Simultaneous Equation Models: Nonrecursive Models with Instrumental Variables Estimated with OLS, and 2SLS

For this lab assignment, we will be learning various ways of calculating Simultaneous Equation Models, specifically a nonecursive model. The model is taken from the example used in Professor Paxton's class notes, and in Chapter 2 of her co-authored book on Nonrecursive Models,[1] for which a pdf copy is included on Blackboard with this assignment under the file name **39916_Chapter2.pdf**. This model of voluntary associations and generalized trust, on pages 18-21, is considered nonrecursive because it has both a reciprocal relationship between the two endogenous variables, and also the disturbances of its two equations are correlated. Both of the models are considered identified (just identified or overidentified) by the order condition, because of the instrumental variables in the model, which mean that each equation has at least one unique parameter not included in any other equation. The data for this example are taken from the 1993 and 1994 waves of the General Social Survey, and can also be downloaded from Blackboard as **Lab8.dta**.

Coding and Questions:

1. Begin with the usual setup, a do file, a log file, change directory, clear all, set more off, and open the dataset for Lab 8. Describe the data and get an idea of the variables it contains.

2. Read the example on pages 18-21 of the pdf for Chapter 2 from Professor Paxton's book.

3. See **Figure 2.4** of the just identified path model of Voluntary Associations and Generalized Trust. Note the two instrumental variables without which, this model would be underidentified or unidentified. Write the two equations for the model in **Figure 2.4** and provide any additional information needed to fully understand the model (i.e. correlations of errors).

4. See **Figure 2.5** of the over-identified path model of Voluntary Associations and Generalized Trust. Note the additional two instrumental variables that give us additional information in this model. Write the two equations for the model in **Figure 2.5** and

[1] Paxton, Pamela M. (Marie), John R. (Robert) Hipp, and Sandra T. Marquart-Pyatt. *Nonrecursive Models: Endogeneity, Reciprocal Relationships, and Feedback Loops*. 1st ed. Sage Publications, Inc, 2011.

provide any additional information needed to fully understand the model (i.e. correlations of errors).

5. Let's estimate the models in **Figures 2.4 and 2.5** ignoring simultaneity for a moment, in other words, performing a normal OLS regression. **Be sure to interpret both regression outputs for each model.** (2 models x 2 equations = 4 regressions).

   NOTE:
   But let's think for a moment about why OLS is inappropriate. Simultaneity is a reality of the models (**2.4** and **2.5**) as they are specified in the path diagrams. What this means is that there is a reciprocal effect of association and trust that is not modeled with single equation OLS. Thus, we will use a technique called Two-Stage Least Squares. Here is a simple explanation and an example of Two-Stage Least Squares and how it works, taken from this website:[2]

   Standard linear regression models assume that errors in the dependent variable are uncorrelated with the independent variable(s). When this is not the case (for example, when relationships between variables are bidirectional), linear regression using ordinary least squares (OLS) no longer provides optimal model estimates. Two-stage least-squares regression uses instrumental variables that are uncorrelated with the error terms to compute estimated values of the problematic predictor(s) (the first stage), and then uses those computed values to estimate a linear regression model of the dependent variable (the second stage). Since the computed values are based on variables that are uncorrelated with the errors, the results of the two-stage model are optimal.

6. Now, let's run a 2 stage least squares regression with one instrumental variable per equation for the model in **Figure 2.4**. Note that 2 equations means 2 separate regressions here as well. **Be sure to interpret the results of these two regressions and postestimation commands.**

7. We may not be satisfied with the model in **Figure 2.4** for a number of reasons you may discover in your interpretations. So, let's now calculate the model the same way with Two-Stage Least Squares, but instead include two instrumental variables per endogenous variable, for a total of four instruments as in **Figure 2.5**. **Be sure again to interpret the results of both regressions and postestimation commands.**

8. Given this information, would you respecify the model? If so, how?

[2] *http://pic.dhe.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Fidh_tsls.htm*

9. Look at the dataset and see if you can think of one more instrumental variable for each endogenous variable in our model. Run the model from **Figure 2.5** again adding these variables into the model. Run it both as 2SLS (with postestimation tests for endogenous variables, and exogeneity of instrumental variables) and as 3SLS. **Be sure to interpret, and compare your model's fit to the previous models (i.e. the AIC and BIC from 3SLS postestimation).**

NOTES:

a) The Stata commands **ivregress** allow the use of instrumental variables and various types of 2SLS, 3SLS and other estimators for simultaneous equations. These date from approximately Stata 10, though they have been improved continually.

b) It is also possible to do 3 stage least squares, which is has the same steps as 2SLS, but inserts a third step in the middle, which uses the results of the first stage to estimate the error of the model. It is a better fit in data with a higher correlation of the errors of the simultaneous equations in the model. The command **reg3** in Stata is the way to do this. See *http://sfb649.wiwi.hu-berlin.de/fedc_homepage/xplore/tutorials/xaghtmlnode22.html* for more info on the method of estimation.

c) A newer and more comprehensive, but also more complex, package in Stata is the SEM (Structural Equation Modelling) Package, available since Stata 12, which also allows estimation of Simultaneous Equation Models, as these are a sub-class of Structural Equation Models. The advantage with SEM, is that it also allows latent variables, factor analysis and other more complex models. If this is something you are interested in, please talk to Nicholas or Pam, look at Pam's Sage green book cited on page 1, and consider her class on SEM whenever it is offered again.