

Replication Exercise
ECO 60303
Due in Hardcopy, December 8, 2010 by 5:00pm

Bill Evans
Fall 2010

In 1982, the *Journal of Money Credit and Banking (JMCB)* adopted a policy that required authors to provide data and computer programs upon request to researchers. The goal of the policy was to allow scholars to more easily replicate published work. In 1986, the *American Economic Review* published an article from the *JMCB Data Storage and Evaluation Project* where the authors attempted to replicate multiple articles published in the *JMCB* with data obtained through their replication policy. The results of the exercise were less than encouraging for our profession in that the authors concluded "...inadvertent errors in published empirical articles are a commonplace rather than a rare occurrence." (Dewald, Thursbury and Anderson, 1986).

As a result of the *JMCB* data replication project, the *American Economic Review* adopted a policy requiring authors to provide data and computer code to interested researchers. Since then, the *AER* policy has been expanded and authors are required to upload computer code and data on a server prior to publication so that others can easily replicate their work. This same policy has been adopted by the *Journal of Political Economy*, the *Review of Economics and Statistics*, the *Review of Economic Studies* and all of the new *AEA* journals.

Replication can take many forms. In the physical sciences where laboratory experiments are common, replications can include trying to reproduce the results of an existing experiment. In the social sciences where experiments are not as frequent, replication usually takes one of two forms. The first exercise examines the external validity of results by re-estimating the same model on different samples. Another type is where researchers attempt to reproduce a set of results from the same sample. In this assignment, I want you to perform this latter exercise.

Replications exercises are an important part of the scientific process in the physical sciences but they are less commonplace in social sciences where primary data collection through experimentation is not a critical element of most research programs. More importantly, it is hard to get replication exercises published in journals so authors are naturally discouraged from starting them. In many cases, however, replication exercises have detected some significant limitations and outright errors in work. Some of the more famous recent examples in our profession are: McCrary, *American Economic Review*, 2002; Rothstein, *American Economic Review*, 2007; Foote and Goetz, *Quarterly Journal of Economics*, 2008; Minarik, *Quarterly Journal of Economics*, 1984; and my personal favorite, Kahn and Udry, *American Sociological Review*, 1986.

Replications are also a great pedagogical tool. For those new to a data set or a computer program, they are a great way to learn about a subject matter since one has a fixed target to work towards.

Your assignment is to select an article that utilized publicly available data and attempt to replicate the results of the analysis. You are to return to original sources for data and start the replication exercise from scratch. Download the data and write your own program to extract the analysis samples. Start small. Try to match sample sizes and means first before jumping to regression models. In many cases, the authors may have different samples, different data sets, etc. Focus on one set of results rather replicating all the numbers.

Do not pick articles where the authors collected original data for the project, such as Card and Krueger's work on the minimum wage or Bertrand and Mulinathian's work on names. Use data sets like the Census Public Use Micro Samples or the Current Population Survey.

I do not care what article you choose to replicate but you need to inform me of your choice and since this is a small class, there should be no duplicates. Please let me know of your selection by **Wednesday, October 6th**. Do not contact authors for data or programming code. Violation of this directive will be considered a violation of the Academic Code of Honor. If at the end of the semester, you still cannot replicate their results, then you can contact the authors – this usually means the author did something wrong or the write-up was wrong.

Your write-up should include up to 15 pages of double-spaced text (1 inch margins, 12 point font) that outlines the goal of the original paper, what data set was used, a short description of the analysis sample and any sample selection restrictions employed by the author, plus a short description of your attempts to replicate the results. Highlight key decisions the author may have made (how they selected respondents, deleted outliers, etc.) and whether the results were sensitive to these key assumptions. You should also describe and estimate one additional model you think the authors should have included in their analysis. Explain why they should have estimated this model. You should provide as many tables as necessary comparing your results to those published in the journal and these do not count against the 15 page limit. You should also attach to the paper the programming code (not output) that produced your results. Hardcopies of the assignment should be turned into me by 5:00pm on December 8th. The text should not be an outline but rather written as if you were submitting the paper to a journal. I will subtract points for poor exposition and grammar.

You are to work on the project alone but you can ask colleagues for programming advice or for some background information about data set. Also, be smart about the papers you try to replicate. Pick a paper from your field of interest or one as close to your dissertation as possible. Don't replicate the work of someone in our department. Also, the personal and professional return from replicating results published in a third-tier

journal are substantially lower than if you look at papers published in the top five journals. Hunt rhinos, not possums.

Some publicly available data sources commonly used in economics. This list is heavily weighted towards micro data sets:

ICPSR (www.icpsr.org) is a user supported data repository at the University of Michigan that contains thousands of data sets. Most major universities subscribe to ICPSR and data is downloadable from the web site after registration.

Census Public Use Micro Samples (1 and 5 percent cross-sectional samples of the U.S. in census years) www.ipums.org (requires registration but it is free)

Current Population Survey (Monthly labor market survey of about 60K households) http://www.nber.org/data/cps_index.html.

Annual Demographic File (March) from the Current Population Survey (The March CPS contains detailed questions about work experience and insurance coverage in the previous year). The data have been harmonized across years and can be found at www.ipums.org.

National Health Interview Surveys: (Annual survey of 60k households, designed to measure the stock of health in the U.S.) <http://www.cdc.gov/nchs/nhis.htm> and ICPSR

National Health and Nutrition Examination Surveys (Detailed survey of the health of about 14K people including physical exams and many diagnostic tests) <http://www.cdc.gov/nchs/nhanes.htm>

Mortality and Natality detail data (Census of all birth and death certificates in the U.S.) <http://www.nber.org/data/>

National Longitudinal Survey of Youth, 1979 and 1997 (longitudinal data of young adults) <http://www.bls.gov/nls/>

Panel Study of Income Dynamics (PSID) (Longitudinal data set of 5000 families beginning in 1969) <http://psidonline.isr.umich.edu/>

Indonesian Family Life Survey (IFLS) (A longitudinal data set of 7200 families from Indonesia. Recent waves have interesting biomarker data). <http://www.rand.org/labor/FLS/IFLS/>

Survey of Consumer Finance (SCF) (Cross sectional surveys done in 1989, 1992, 1995, 1998, 2001, and 2004 By the Federal Reserve Board of Governors, designed to measure the balance sheet, pensions, wealth and income of US households) <http://www.federalreserve.gov/pubs/oss/oss2/2004/scf2004home.html#scfdata2004>