

## Regression Analysis Final Project Information

DSO 529, Spring 2017

Due date: Thursday, April 27, 2017

The goal of this project is to get hands-on practice with multiple regression analysis, in the context of time series and/or panel data analysis. A final report will be turned in on the due date and will be graded based on various objectives. You should utilize as many methods from the second half of the course as fit your data. I expect you to thoroughly analyze your data and check for model violations, taking corrective measures, if necessary.

Select a topic of interest – perhaps something business related, or maybe something fun and interesting to you. You must be able to obtain enough data to make the analysis worthwhile. Each group or individual is responsible for finding his/her/their own data set. For websites with large amounts of publicly available, please see:

<http://pages.stern.nyu.edu/~jsimonof/classes/datalink.html> and

<http://databank.worldbank.org/data/home.aspx>. Also, here is a link to the USC library's database website: <http://libguides.usc.edu/finhelp>. If your group is having significant difficulty locating worthy data, come talk to me so that we can try to pool resources. I have also added a rather lengthy list at the end of this document with other econometric websites listed. If your group is having significant difficulty locating worthy data, come talk to me so that we can try to pool resources.

The dataset that you start with should have *at least 50 observations and at least five independent variables at the start*. I realize that during the analysis process you may legitimately choose to delete some variables.

Because we've covered some advanced topics in this course, your project should include at least one of the following:

1. Variable transformations to alleviate heteroscedasticity and non-linearity,
2. Dummy variables, interaction effects, and outlier detection,
3. Time series datasets that include at least two time series and an attempt to regress them against each other, checking for stationarity and cointegration, or
4. Pooled or panel data from at least two time periods.

Your data may not lend itself well to all of these topics, but try to find a dataset that does need some work done to it! ☺

### **Final Report**

The **final report** should consist of the following, which should be written in a report format, not just a series of steps and conclusions:

## Project Information

- A short cover-page statement about the problem (akin to an executive summary)
- Source of the data
- Listing of the variables in the database, including the designation of the dependent variable, as well as a fairly complete description of each
- Descriptive Statistics
  - Means, standard deviations, and correlations of the variables
  - One or more scatterplots involving the dependent variable
- The proposed regression objective
  - [This will typically be the exploration of a relationship. It could also involve prediction of new data points.]
- Initial regression model, including output and residual plot
- Sketch of regression work
  - This includes, but is not limited to, elimination of variables, differencing of variables, transformations such as logarithms, among others. Basically, describe the steps (perhaps in a table format) you took in the analysis and why you took them, as if you were writing a journal.
- Final fitted regression model, including output and residual plot
- Interpretation of your model choice and intuition about what the output is telling you - coefficient values,  $R^2$  value, F-statistic, Durbin-Watson statistic (along with appropriate cutoff values or p-values), etc. [Also: If the  $R^2$  is not very high, why do you think that is?] Explain the final model in plain English in the context of your data and tell me if it seems reasonable.
- Relevant example to interpret the regression equation, either from your actual dataset or a separate, but meaningful set of values.
- Post-regression analysis
  - What recommendations would you have for improving the model?
  - [Also: Were you satisfied with your results? If not, are there any ideas or suggestions you have about why they were not satisfactory?]

The final report will *probably* be about 10 pages in length. It's hard to give a precise description of the length, because the layout of a report is strongly dependent on the number of plots and on their sizing. Please hand in only **relevant** output plots and charts – I don't need to see everything that you tried but discarded!!

NOTE: You will NOT be graded based on the values you get in your final output! Yes, the objective is to obtain the best values possible, but depending on your data, you may end up with results you don't like. I want you to learn how to analyze the data, and that's not necessarily evident in the numbers!

### **Additional Econometric Websites:**

- *Economic Statistics Briefing Room*: An excellent source of data on output, income, employment, unemployment, earnings, production and business activity, prices and money, credits and security markets, and international statistics. <http://www.whitehouse.gov/fsbr/esbr.htm>
- *Federal Reserve System Beige Book*: Gives a summary of current economic conditions by the Federal Reserve District. There are 12 Federal Reserve Districts. [www.federalreserve.gov/FOMC/BeigeBook/2008](http://www.federalreserve.gov/FOMC/BeigeBook/2008)
- *National Bureau of Economic Research (NBER) Home Page*: This highly regarded private economic research institute has extensive data on asset prices, labor, productivity, money supply, business cycle indicators, etc. NBER has many links to other Web sites. <http://www.nber.org>
- *Panel Study*: Provides data on longitudinal survey of representative sample of U.S. individuals and families. These data have been collected annually since 1968. <http://www.umich.edu/~psid>
- *The Federal Web Locator*: Provides information on almost every sector of the federal government; has international links. [www.lib.auburn.edu/madd/docs/fedloc.html](http://www.lib.auburn.edu/madd/docs/fedloc.html)
- *WebEC: WWW Resources in Economics*: A most comprehensive library of economic facts and figures. [www.helsinki.fi/WebEc](http://www.helsinki.fi/WebEc)
- *American Stock Exchange*: Information on some 700 companies listed on the second largest stock market. <http://www.amex.com/>
- *Bureau of Economic Analysis (BEA) Home Page*: This agency of the U.S. Department of Commerce, which publishes the *Survey of Current Business*, is an excellent source of data on all kinds of economic activities. [www.bea.gov](http://www.bea.gov)
- *Business Cycle Indicators*: You will find data on about 256 economic time series. <http://www.globalexposure.com/bci.html>
- *CIA Publication*: You will find the *World Fact Book* (annual). [www.cia.gov/library/publications](http://www.cia.gov/library/publications)
- *Energy Information Administration (Department of Energy [DOE])*: Economic information and data on each fuel category. <http://www.eia.doe.gov/>
- *FRED Database*: Federal Reserve Bank of St. Louis publishes historical economic and social data, which include interest rates, monetary and business indicators, exchange rates, etc. <http://www.stls.frb.org/fred/>
- *International Trade Administration*: Offers many Web links to trade statistics, cross-country programs, etc. <http://www.ita.doc.gov/>
- *STAT-USA Databases*: The National Trade Data Bank provides the most comprehensive source of international trade data and export promotion information. It also contains extensive data on demographic, political, and socioeconomic conditions for several countries. <http://www.stat-usa.gov/>

## Project Information

- *Bureau of Labor Statistics*: The home page contains data related to various aspects of employment, unemployment, and earnings and provides links to other statistical Web sites. <http://stats.bls.gov>
- *U.S. Census Bureau Home Page*: Prime source of social, demographic, and economic data on income, employment, income distribution, and poverty. <http://www.census.gov/>
- *General Social Survey*: Annual personal interview survey data on U.S. households that began in 1972. More than 35,000 have responded to some 2500 different questions covering a variety of data. [www.norc.org/GCS+Website](http://www.norc.umd.edu/GSS+Website)
- *Institute for Research on Poverty*: Data collected by nonpartisan and nonprofit university-based research center on a variety of questions relating to poverty and social inequality. <http://www.ssc.wisc.edu/irp/>
- *Social Security Administration*: The official Web site of the Social Security Administration with a variety of data. <http://www.ssa.gov>
- *Federal Deposit Insurance Corporation, Bank Data and Statistics*: <http://www.fdic.gov/bank/statistical/>
- *Federal Reserve Board, Economic Research and Data*: <http://www.federalreserve.gov/econresdata>
- *U.S. Census Bureau, Home Page*: <http://www.census.gov>
- *U.S. Department of Energy, Energy Information Administration*: [www.eia.doe.gov/overview\\_hd.html](http://www.eia.doe.gov/overview_hd.html)
- *U.S. Department of Health and Human Services, National Center for Health Statistics*: <http://www.cdc.gov/nchs>
- *U.S. Department of Housing and Urban Development, Data Sets*: <http://www.huduser.org/datasets/pdrdatas.html>
- *U.S. Department of Labor, Bureau of Labor Statistics*: <http://www.bls.gov>
- *U.S. Department of Transportation, TranStats*: <http://www.transtats.bts.gov>
- *U.S. Department of the Treasury, Internal Revenue Service, Tax Statistics*: <http://www.irs.gov/taxstats>
- *Rockefeller Institute of Government, State and Local Fiscal Data*: [www.rockinst.org/research/sl\\_finance](http://www.rockinst.org/research/sl_finance)
- *American Economic Association, Resources for Economists*: <http://www.rfe.org>
- *American Statistical Association, Business and Economic Statistics*: [www.amstat.org/publications/jbes](http://www.amstat.org/publications/jbes)
- *American Statistical Association, Statistics in Sports*: <http://www.amstat.org/sections/sis/>

## Project Information

- *European Central Bank, Statistics:* <http://www.ecb.int/stats>
- *World Bank, Data and Statistics:* <http://www.worldbank.org/data>
- *International Monetary Fund, Statistical Topics:* <http://www.imf.org/external/np/sta/>
- *Penn World Tables:* <http://pwt.econ.upenn.edu>
- *Current Population Survey:* <http://www.bls.census.gov/cps/>
- *Consumer Expenditure Survey:* <http://www.bls.gov/cex/>
- *Survey of Consumer Finances:* <http://www.federalreserve.gov/pubs/oss/>
- *City and County Data Book:* <http://www.census.gov/statab/www/ccdb.html>
- *Panel Study of Income Dynamics:* <http://psidonline.isr.umich.edu>
- *National Longitudinal Surveys:* <http://www.bls.gov/nls/>
- *National Association of Home Builders, Economic and Housing Data:*  
<http://www.nahb.org/page.aspx/category/sectionID=113>
- *National Science Foundation, Division of Science Resources Statistics:* <http://www.nsf.gov/sbe/srs/>
- *Economic Report of the President:* <http://www.gpoaccess.gov/eop/>
- *Various Economic Data Sets:* <http://www.economy.com/freelunch/>
- *The Economist Market Indicators:* <http://www.economist.com/markets/indicators>
- *Statistical Resources on the Military:* <http://www.lib.umich.edu/govdocs/stmil.html>
- *World Economic Indicators:* <http://devdata.worldbank.org/>
- *Economic Time Series Data:* <http://www.economagic.com/>