Due Date

The new due date for empirical paper is Wednesday, March 24 at the beginning of class.

Description

For the empirical exercise, you will use corporate governance, financial, and macroeconomic data at the country level to develop a statistical model that can be used to determine which governance characteristics are of most importance and which ones do not matter for the development of capital markets. Most importantly, you will have to explain the economic intuition behind your results and draw public policy recommendations supported on your statistical evidence. Then, based on your econometric work, you will draft a paper describing your data, methodology, and results. Finally, the paper should compare the results you obtained with the stylized facts published in the literature.

Data

I will make available on the course website an Excel and a STATA files that include the data that you will use for the empirical exercise. I will also post a “manual” file that

*Please send any questions, comments, or suggestions to mgamboa@fas.harvard.edu. I also want to thank Joe Aldy and Jeremy Tobacman for allowing me to base this empirical exercise on their own previous Ec-970 experiences. This document is available at: http://www.courses.fas.harvard.edu/~ec970mg/Assignments/Empirical/
includes the description of the variables in those files. The dataset consists of a cross-
section of 49 countries—those same countries analyzed in the paper “Law and Finance,”
by La Porta et al. (1998). The remaining variables are drawn from the World Development
Indicators published by the World Bank for the year 2000. The STATA dataset includes
a description for each of these variables.

Analysis

The empirical paper should go beyond just plotting data points and telling a story.
The report must involve statistical analysis and the reporting of regression coefficients
and its standard errors. Standard errors, in particular, give us a sense of the uncertainty
in the data. Large standard errors mean that many different regression specifications
and thus, coefficients, are consistent with the data, a result that weakens any statements
about your modeling of the world.\footnote{Recall that we can tell whether a coefficient is statistically significant by computing a $t$-statistic. In particular, for the hypothesis that a coefficient is statistically different from zero, we obtain the $t$-statistic by dividing the regression coefficient by its standard error:

$$H_0 : \hat{\beta} = 0 \quad t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

As a rule of thumb, if $t \geq 2$, we say that the coefficient, $\hat{\beta}$, is statistically different from zero at the 5 percent level.}

In particular, you will need to think about:

1. **Left-Hand-Side (LHS) Variables or Dependent Variables:**
   These are the variables you want to explain with your regression, e.g., stock market valuation to GDP, trading volume to GDP, interest rates, real exchange rate depreciation, economic growth, etc.

2. **Right-Hand-Side (RHS) Variables or Independent Variables:**
   These are the variables you think will explain or affect the LHS variable. You need to have an economic reason to put in a certain variable—think about the economic causality, e.g., enhanced shareholder protection should imply a higher stock market valuation.
   In general, you will want to run the following equation:
   \[
   Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + \epsilon_i
   \]
   where $Y_i$ is the LHS variable for country $i$ that you want to explain and $X_{k,i}$ is the $k$-th RHS explanatory variable for country $i$.

3. **Control Variables (a kind of RHS variables):**
   Sometimes you may want to filter-out the effect of country-specific characteristics. For instance, you may want to control for the “natural” effect that the wealth of a
country has on the stock market valuation. The controls most commonly used in
corporate governance regressions include the log of GDP per capita, the change in
GDP or GDP growth, and country fixed-effects in the case of a panel.\textsuperscript{2}
In this case, you will want to run the following equation:

\[ Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + \Gamma + \varepsilon_i \]

where \( \Gamma \) is a matrix of controls.

4. **Hypotheses Variables (a kind of RHS variables):**
These are the variables you want to focus on and test their statistical significance
on the LHS variable. For example, think about the effect of creditor protection on
the size of the debt market.

5. **Interaction Terms (a kind of RHS variables):**
Interaction terms serve to refine the analysis of your hypotheses. Suppose your
hypothesis is that creditors’ rights only work in German legal origin countries, where
enforcement is much stronger than in French legal origin countries and creditors are
well-protected. You may want to test this hypothesis by including an interaction
term like, \( (G_i \cdot CR_i) \), where \( G \) is a dummy variable taking on the value of 1 for all
countries that are of German legal origin and 0 otherwise, and \( CR \) is the score of
the creditors’ rights index.
In this case, you will want to run the following equation:

\[ Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + \alpha (G_i \cdot CR_i) + \Gamma + \varepsilon_i \]

where \( \alpha \) is the relevant coefficient on the interaction term.

For the empirical exercise, you should be able to properly interpret regression esti-
mates and its standard errors. Moreover, you will have to compute \( t \)-statistics for your
coefficients and test the null hypothesis of whether they are statistically different from
zero. Furthermore, you should build a story (sort of a theory) around your statistical
analysis and support your conclusions with the data analysis. **All regressions in your
paper must represent one or several underlying hypotheses that you are willing to test.** Always think about the economic intuition behind your regression analysis
and always avoid data mining.\textsuperscript{3}

Also, for this empirical exercise you should keep in mind possible endogeneity or
reverse causality problems. For example, if you are trying to explain the size of the stock
market (LHS) with a corporate governance characteristic, say, shareholder protection
(RHS), you should also keep in mind the possibility that larger stock markets induce
better shareholder protection because in more liquid market there is also more investor

\textsuperscript{2}A panel is a dataset that includes country and year observations. For instance, a panel of a cross-
section of countries would include data on various countries for several years.

\textsuperscript{3}Data mining refers to the act of running regressions—“sometimes called spurious regressions”—
without knowledge of the data and the underlying hypotheses. That is, you can always find a statisti-
cally significant relations between any two variables, but it is much harder to come up with a coherent
hypothesis to support the relation.
awareness and activism. In this case, you will have to instrument shareholder protection with an exogenous variable, for instance, legal origin. You can do this in STATA by using the “ivreg” command, which stands for Instrumental Variables (IV).

Recommendations

Keep in mind that the main purpose of the empirical exercise is to practice thinking and writing clearly about an empirical question. First, you have to narrow down your topic and define your research question, a focused question, which you can later translate into a single regression. You should carefully describe the regression equation you plan to use—this is usually called the regression specification. Secondly, you should identify the dependent variable (LHS) and independent variables of interest (RHS). Then, you must explain how to interpret the LHS-RHS relationship—why they might be related and why.

The variables you use can be taken directly from the dataset or, alternatively, you can construct your own. These data lend themselves naturally to analysis with instrumental variables (IV) regression due to endogeneity problems. Therefore, you should implement your econometric strategy using the “ivreg” command in STATA, instrumenting with legal origin. As you write down your regression equation, with a clear dependent (LHS) variable, and a clear independent (RHS) variable of interest, you should consider whether you need to include other “control” variables on the right hand side. In addition, it is important to include a constant unless you have a very good reason not to; STATA will automatically include a constant unless you tell it otherwise.

Also, you might consider whether you want to simultaneously look at the whole sample or just a subset. You should anticipate how you will interpret the output from your regression. When you run the regression, how will you use the output to answer your question? STATA output includes many numbers: it reports an $R$-squared, an $F$-test, point estimates for coefficients on the independent variables, $t$-tests, $p$-values, and confidence intervals for each coefficient. Each of these pieces of output can be used to answer different commonly-asked econometric questions; which data you use depends on which question you are asking.

Finally, to evaluate the effect of the independent variable $X$ on the dependent variable $Y$ though, it is not enough just to look at the point estimate for the coefficient on $X$. To characterize the effect of $X$, you need to ask two follow-up questions: “is the effect statistically significant?” and “is the effect economically substantial?” You can address statistical significance using the $t$-statistic. Assessing economic impact is trickier and depends on the actual economic circumstances. A good start is often to report the number of standard deviations by which $Y$ changes when you change $X$ either by a standard

---

4 Presumably, since legal origins were determined approximately 300 years ago, they are exogenous enough to be considered good instruments. A good instrument has the characteristic of being correlated with the RHS variables but uncorrelated with the LHS variable.

5 Legal origin will be the instrument used for empirical exercise. For the term paper, you can come up with your own instrument.
deviation or from 0 to 1.

Format of the Empirical Paper

The empirical paper should be typed in Times Roman 12-point font, double-spaced, with 1” margins. The page limit for this assignment is 7 pages. Tables and figures should be attached at the end of the paper, as appendices, and will not count towards the page limit. Pages beyond the number allocated for will not be taken into account. In general, a paper should include an Introduction part, a Body/Literature Review part, a Results part, and a Conclusions part. This may be a useful structure for your papers but it is not the only one allowed. You should also refer to your “Writing Economics” manual for specific details on citations and structure.

Additional Resources

On STATA Software

You can download and install Stata for keyaccess use for free from here: http://www.fas.harvard.edu/cgi-bin/software/download.pl

For a general introduction to using Stata see Chapters 1 and 2 from here: http://www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm

Also, make sure you understand this page: http://www.ats.ucla.edu/stat/stata/output/reg_output.htm

On Regression Analysis

James Stock and Mark Watson, Introduction to Econometrics. (This is the textbook for Economics 1123.) Chapter 4, on univariate regression, is available on the web: http://www.aw.com/info/stock_watson/Chapter4.pdf

---

6 Please refer to the sample paper that I have posted on the course website.