Prof. Murillo Campello

EMPIRICAL CORPORATE FINANCE

Empirical Assignment #1

Introduction

Rajan and Zingales's (1995) "What Do We Know about Capital Structure? Some Evidence from International Data," in the *Journal of Finance*, is a reference paper for the study of corporate capital structure. The authors use a "catch-all" regression specification that puts together variables representing multiple theories. To gauge the relative importance of the theories they consider, Rajan and Zingales estimate a regression similar to the following:

$$Leverage_{i,t} = \beta_0 + \beta_1 CashFlow_{i,t} + \beta_2 Tangibility_{i,t} + \beta_3 Size_{i,t} + \beta_4 Q_{i,t} + \varepsilon_{i,t} \tag{1}$$

where all variables are defined below.

Later in the course, students will be asked to discuss how the above empirical model estimate the importance of theories such as asymmetric information, limited contractability, pecking-order, trade-off, underinvestment, market timing, and others. Right now, the objective of this exercise is to "challenge" the intuition gained from simple regression models.

Tasks

The objective of this assignment is to familiarize students with the basics of standard empirical analysis in Corporate Finance. Your task is to estimate various versions of Eq. (1) for manufacturing firms in COMPUSTAT (or Datastream) for as many years as possible. In the appendix below, you find a description of the construction of the variables needed and some basic data filtering criteria. Present your work in the following way (tables), sequentially adding each change listed below as you perform the tests:

1) Detailed summary statistics for the variables presented. Do any of the variable distributions (and their economic meaning) cause concerns for the proposed estimation?

2) Correlations between all the variables. Do we already have an idea of how the actual regression analysis will look like? Any concerns about colinearity? Do you have concerns about direction of causation for leverage?

3) A simple OLS regression model as presented in Eq. (1) above. Interpret the economics of the results following the approach used in Campello and Giambona's (2013) "Real Assets and Capital Structure," in the *Journal of Financial and Quantitative Analysis*. Which regressor has the largest effect on leverage and appears to be the most economically important driver?

4) An OLS-FE version of Eq. (1), with year, industry, and firm effects. Do these fixed effects capture a lot of variation in the distribution of leverage? Is endogeneity solved in this way? Do results change a lot? Do your economic inferences change?

5) A version of Eq. (1) where all variables are "winsorized" at the 5% and 95% cutoffs. Another where variables are trimmed at those values. This approach is touted as a way to address the issue of outliers. Do results change a lot?

6) A version of Eq. (1) where the r.h.s. variables are lagged one period? This approach is touted as a way to address endogeneity by making the drivers of the l.h.s. variable "pre-determined." Is endogeneity solved in this way? Do results change a lot?

7) Just for fun, introduce for each firm it's "initial leverage" as suggested by Lemmon, Roberts and Zender's (2008) "Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure," in the *Journal of Finance*. Include this variable in the leverage regressions. What do you find?

8) Describe your views about the robustness of the benchmark empirical model of capital structure. Are you concerned that the economics of the results may change seemingly small changes in the empirical methodology? Do you trust causal inferences made with this model?

Logistics

The assignment is to be performed in groups of 2 people maximum. Please turn in a written report (printed with relevant tables and text) as assigned in the course schedule or discussed in class (no e-mails please). Replication of the results is expected to be straightforward. Your report has to look professional. As such, write a coherent text addressing each of the tasks above and do not simply cut and paste output from statistical software in you paper.

Appendix

This appendix describes data selection and variable construction procedures generally found in the empirical corporate finance literature. The data is from the annual COMPUSTAT. The software language used is Stata. Feel free, however, to use any approach you want to collect data and construct variables. Researchers need not follow any cookbook recipe for their discoveries.

Data Screening. Each observation selected into the sample had to satisfy the following criteria:
1. Firm-year observations are uniquely identified by COMPUSTAT's gvkey-fyear combinations:
(indfmt == "INDL" & datafmt == "STD" & popsrc == "D" & consol == "C")
2. Drop missing company identifiers and missing years
(gvkey != .)
(fyear != .)
3. Keep firms with fiscal years between, for example, 1980 and 2014
(fyear >= 1980 | fyear <= 2014)
4. Keep only manufacturing firms
(sic > 1999 & sic < 4000)
5. Require non-missing information on assets, investment, cash flows, and stock market data
(at !=. & capxv !=. & sppe !=. & ib !=. & dp !=. & prcc_f != . & csho != .)

Variable Computation. Variable names and commands:
Leverage: gen Leverage = (dlc + dltt) / at
Cash Flow: gen CashFlow = (ib + dp) / L.ppent
Tangibility: gen Tangibility = ppent / at
Size: gen Size = log(at)
Q: gen Q = (at + (prcc_f*csho) - ceq - txdb) / at