Homework 1: Computing Firm-Level TFP with Compustat
Due On January 18th, 5:00pm

The goals of this homework are to introduce you to Compustat and to illustrate the many choices you have to make when working with firm-level microdata. Specifically, we will use Compustat to estimate firm-level total factor productivity (TFP) and study how TFP is correlated with other firm outcomes. You are free to use any statistical package you like for this problem set, but Yulia will only be able to answer questions about `Stata`.

You will notice that the instructions in the problem set are very explicit at the beginning but become less explicit as the problem set goes on. This is intentional. In class, we will discuss the different choices that you were forced to make and how they impacted your results.

1. **Download Data** You recently received an email with instructions on how to download Compustat from WRDS. Please follow those instructions to register for WRDS and navigate to Compustat – Capital IQ – North America. We will be working with the monthly-updated Fundamental Annual database.

   (a) Choose date range 1964 - 2017.

   (b) Choose `gvkey` for company codes.

   (c) Select `Search the Entire Database`.

   (d) In the `Screening variables` section, uncheck the industry format `FS` and `CAD` in the currency tab. This will help us avoid duplicate observations.

   (e) In the `Identifying Information` section, select `Stock Exchange Code (EXCH)` and `Foreign Incorporation Code (FIC)`. This will allow us to identify firms based in the U.S.

   (f) Finally, include the following variables in the query: `SIC, PPEGT, PPENT, AQC, AT, SALE, EMP, COGS`.

   We will also need to download price deflators to convert nominal variables to real variables. We will use the NIPA tables, located at: `http://www.bea.gov/iTable/index_nipa.cfm`. Download `GDP and Personal Income, Table 1.1.9. Implicit Price Deflators for Gross Domestic Product` for 1964 - 2017. Merge the series for the GDP deflator (`line 1`) and the series for the nonresidential fixed investment good deflator (`line 9`) into the Compustat data.

2. **Data Description** Once you have downloaded the data and loaded into your statistical software, the first step is to understand what data you have. Please write out a brief data dictionary with a short (two sentence at most) definition of each variable in the workspace.

3. **Sample Selection and Variable Construction** Before analyzing the data we will do some sample selection and variable construction.

   (a) **Selecting firms**

      i. We will only consider firms incorporated in the United States. Please exclude observations with `FIC` not equal to `USA`.

      ii. We will not include utilities firms because these are heavily regulated and therefore assumptions about cost minimization or profit maximization are unlikely to hold. Please exclude observations with `SIC` codes between 4900 - 4999.

      iii. We will not include financial firms because their balance sheets are much different from other firms. Please exclude observations with `SIC` codes between 6900 - 6999.

      iv. Finally, we will only study firms who do business in U.S. Dollars. Please exclude observations with `CAD` not equal to `USD`.

   (b) **Computing the capital stock** Although `PPEGT` is a measure of the book value of the capital stock, we will construct our own measure of capital using the perpetual inventory method. Please do the following for each firm.

      i. Initialize the capital stock using the first available entry of `PPEGT`.

      ii. With the initial value of capital in hand, we now iterate forward on capital using the accumulation equation $k_{it} = (1 - \delta)k_{it-1} + i_{it}$. However, instead of constructing a measure of *gross* investment i_{it}, we will construct a measure of *net* investment $i_{it} - \delta k_{it-1}$. In particular, compute net investment as $i_{it} - \delta k_{it-1} = \texttt{PPENT}_{it} - \texttt{PPENT}_{it-1}$.

      iii. Sometimes, there are missing observations of `PPENT` within a given firm's investment spell. Replace these missing observations with a linear interpolation of the neighboring values of `PPENT`.

      iv. Finally, we want a measure of the real capital stock, so we need to deflate net investment by the investment goods deflator.

   (c) **More sample selection** With this constructed capital stock in hand, our dataset is nearly complete. We will just perform a couple more sample selection.

      i. Exclude observations in which acquisitions are larger than 5% of the value of total assets. A lot of things change about the firm after a large acquisition and we don't want to pick this up in our estimation of TFP.

      ii. Exclude observations with negative sales or employment, which are obviously measurement error.

   (d) **Data description** Now that we have arrived at our final sample for analysis, we need to describe some basic properties of the data. How many observations do we have? What is the average number of firms per year in the sample? Has that

been trending over time? How many sectors are in the sample? What is the average number of firms per sector? Please report anything else you think it would be useful to know.

4. **TFP Estimation** We will now estimate firm-level TFP as the residual of the equation

$$\log(y_{it}) = \mu_i + \mu_t + \alpha_s^k \log(k_{it-1}) + \alpha_s^n \log(n_{it}) + \varepsilon_{it}, \text{ where} \tag{1}$$

   (a) y_{it} is real sales

   (b) k_{it-1} is our constructed capital stock (note that in Compustat capital is recorded as end of period, so capital used in production is k_{it-1})

   (c) n_{it} is employment

   (d) μ_i is a firm fixed effect, meant to capture permanent differences across firms

   (e) μ_t is a time fixed effect, meant to capture aggregate changes common to all firms

   Estimate 1 by OLS. Please report some results of this regression. What is the R^2 of the regression? What is the average estimated α_s^k and α_s^n across sectors? What do the fixed effects μ_t look like over time – do they seem reasonable? Please report anything else you think it would be useful to know.

   The residuals ε_{it} of 1 are our measure of the firm-specific component of TFP. In class, we discussed some issues with this interpretation. What are some things you are worried about in interpreting ε_{it} as productivity? These can be either issues with measurement or issues with the econometric specification of 1.

5. **Analysis of TFP** The last thing we will now do is analyze the properties of our measured firm-level TFP, ε_{it}.

   (a) It is common in quantitative models to assume firm-level TFP follows an AR(1) process

$$\varepsilon_{it} = \rho \varepsilon_{it-1} + \omega_{it}, \text{ where} \tag{2}$$

   ω_{it} is an innovation to the process. Estimate 2 by OLS and report the results of your estimates of the autocorrelation ρ and the variance of the innovations. Do you think they are reasonable?

   (b) Models of optimizing firm behavior would imply that firms with high productivity will also tend to hire more workers and invest more in capital. Let's test this prediction in our data.

i. Construct net hiring at the firm $h_{it} = \texttt{EMP}_{it} - \texttt{EMP}_{it-1}$. How is this correlated with firm-level TFP? With the innovations to firm-level TFP you estimated in 2?

ii. How is net investment correlated with firm-level TFP and the innovations to firm-level TFP from 2?

iii. Construct a measure of gross investment by adding δk_{it-1} back into net investment, assuming $\delta = 0.03$ for all firms. Do things change if you look at gross vs. net investment?