The raw data files are downloaded from WRDS database. Three data files need to be downloaded separately.

1. EXECUCOMP: Executive compensation data file containing historical data on executive compensation for top executive of each public firm. We use the data from earliest available date to track the past experience of CEOs.
2. CRSP & COMPUSTAT merged file: The data file contains information of fundamental accounting variables for firms. It is readily merged to the CRSP data file, thus there is no need to download a separate file to obtain information on stock prices used to calculate market capitalization of the firms.
3. COMPUSTAT SEGMENTS: The historical segments data is used to calculate the scope measure. The file provides segment reporting.

**CEO Experience:**

Many CEOs have been the CEO in their current company for multiple years in a row before 2019. The CEO experience is defined as:

1. Serving in a non-executive role at the current company prior to being promoted to CEO. This is identified by observing the CEO at the current company in previous years with *ceoann* variable being blank. Then, the CEO's experience is considered to be at the current firm in the last year which the CEO observed as a non-CEO executive.
2. Serving at another firm as a CEO or a non-CEO executive. This is identified by observing the CEO at another firm in EXECUCOMP. Then, the CEO's experience is defined to be at the firm, which is different from current firm, in the latest year observed in the data.

There are CEO-firm observations in 2019 that have missing experience according to the above definition.

* Founder CEOs: They appear as the CEO of the company from the earliest records in EXECUCOMP. 24 CEOs are identified as founders.
* Experience in non-US firm: EXECUCOMP tracks executive of the SP1500 listed US companies. A number of CEOs have experience in large public firms based in Europe. The other data files contain information about these firms, but not EXECUCOMP. The past experience of these CEOs is identified from online sources, e.g., Bloomberg Executive Profiles. Gvkey and they last year of employment is then recorded in a separate spreadsheet.
* A number of CEOs do not have a career history in public firms. Although, it is possible to identify the company and years in which the CEO was working at these private companies, important information about private firms will be missing from the other data files. These CEOs are not included in the sample
* A number of CEOs were active in their current firm prior to becoming CEO in lower-level positions. This is why they were not included in the top 5 executives reported annually by the company.

Except for the founder CEOs and those with a carrier history entirely in private firms, the CEOs experience is identified by gvkey of the company of their employment and the last year of employment. The data file is then merged with the executive compensation dataset.

Out of 495 firm-CEO pairs in S&P500 identified in the data in 2019:

* 21 are founded by the CEOs. These firms and their CEOs are not going to be part of the sample.
* 7 are matched to CEOs that do not have past experience in public companies. These company-CEO observations are not going to be included in the sample.
* 466 CEOs have past experience EXECUCOMP dataset which is either (i) non-CEO executive role in the same company; or (ii) CEO or non-CEO executive role in another company. There is a gvkey (unique firm identifier) and a year variable to identify the company and the (last) year at which the CEO was active in that firm prior to becoming CEO at the current company. This could point to the current company of the CEO if the CEO appears in a non-CEO executive role in Execucomp in previous years.
* Out of these 466 remaining firms, there are 24 firms-CEO pair that either (i) the firm does not appear in the segment file in 2019; or (ii) the firm that CEO worked at previously does not appear in the segment file in that year.   These are mostly banks and financial institutions. They do not appear in the raw segment file. This means we do not have the scope measure for these observations. I can go through the financial statements and find the revenue by segments. But I am not including them in the sample for now.

We are left with 442 firm-CEO observations with scale and scope measures in the data. The scale measure for a firm is the number of employees in 2018, and for the CEO, it is the number of employees of the firm where the CEO played an executive role prior to becoming the CEO at the current company.

Out of 442 firm-CEO observations, 76 have past experience at a company other than their current one.

The correlation between the number of employees of the firm and the number of employees for the CEO's experience is 0.9. The correlation between the scope measure of the firm and CEO's scope measure is 0.58.

The high correlation in employee number is probably due to the fact that 366 out of 442 of these CEOs have past experience at their current company as an executive.

Here's a scatter plot of firm's scale measure versus CEO's scale measure:

Here's the same for scope's measure: