Title: "CEO Compensation: Facts"
Authors: Gian Luca Clementi and Thomas Cooley
Manuscript number: RED-D-23-00159
Guest editor: Nezih Guner

# 1 Introduction

This document provides detailed information about the STATA codes used to compute the statistics displayed in the paper titled "Executive compensation: Facts," by Gian Luca Clementi and Tom Cooley. Calculations were executed on a Windows machine using STATA v.15. The execution time is negligible.

In Section 2, we provide an account of the .do files.

The paper uses proprietary data from EXECUCOMP, COMPUSTAT, and CRSP, all maintained by WRDS at the University Pennsylvania. The authors could access the data as employees of the Stern School of Business at New York University, which has contract with WRDS. The terms of Stern's user contract do not allow the authors to make the raw data available on the journal's website. In section 3, we describe the raw data taken in input by the codes. Anyone with access to EXECUCOMP, COMPUSTAT, and CRSP, will easily assemble the datasets needed to perform the calculations.

The codes also produce results that were not included in the published version of the paper. Datasets and do files that are not required to produce the published results are marked with an asterisk.

# 2 Description and sequencing of the DO files

In this section we describe the purpose of each .do file, the data they take in input, and the output they produce. In order to guarantee a smooth run, the .do files should be executed respecting the order in which they are listed.

1. **cpi.do** – It loads the PCE index produced by the BEA directly from FRED @ St. Louis Fed. It produces standard monthly series of price level, 12-month and 1-month inflation according to the solar year convention and makes them available in *modified_cpi*. It also computes monthly series that can be used to determine yearly series of the price level and 12-month inflation according to the fiscal year convention (for all firms, no matter their chosen fiscal year) price deflator and inflation. These series are made available in *cpi.dta* and *modified_cpi.dta*.

2. **wages.do** It loads data on private-sector wages produced by the BLS, directly from FRED @ St. Louis Fed. If produces time series of real wages and makes them available in *wages.dta*

3. **sp500.do** – Uses data for monthly value-weighted returns for NYSE and NASDAQ companies, inclusive of dividends. It makes the returns available with solar year convention in *sp_varianza.dta*, which will be used by *varianza.do*. It also produces series of 12-month cumulative returns with fiscal year convention and makes it available in *s&p.dta*. It needs in input: *s&p_raw.dta*.

4. **int_rate.do*** – It loads data on 3- and 5-year interest rates directly from FRED @ St. Louis Fed for the purpose of applying the Black-Scholes formula. The output is *int_rate.dta**.

5. **conversion.do** It loads EXECUCOMP data downloaded from WRDS, does some cleaning, and generates the STATA datasets *compensation_raw.dta*, *executive_raw.dta* and *company_raw.dta* for use by other do files. It needs in input: *execucomp_22.dta* and *company_22.dta*.

6. **stock.do** It needs in input: *map.dta*, *stock_raw.dta*

    - It produces a bridge between the identifiers PERMCO and GVKEY, needed for the merge of Compustat and Execucomp with CRSP. The output is the file *chiavi.dta*.
    - It merges stock price data from CRSP data, downloaded into *stock_raw.dta*, with compensation data

- It produces time-series of volatility for the purpose of correcting sensitivity estimates for the variance of shareholder returns. The output is in *stock.dta*, stock_dollar.dta for the volatility of stock returns and dollar returns, respectively.
- It computes the max and mean stock price for all companies, for every fiscal year, using monthly data from CRSP. The output is *price_moments.dta*
- It produces stock prices at the end of the three latest fiscal years and a factor needed to adjust the number of options. The output is in *options.dta**

7. **yield.do*** It loads stock market data from CRSP and produces a measure of the dividend yield used in the valuation of options via the Black & Scholes fomula. The output is *bs_yield.dta**. It needs in input: *volat_raw.dta**

8. **volat.do*** It loads price data from CRSP, made available in *volat_raw.dta* and computes estimates of stock return volatility for the purpose of evaluating options via the Black–Scholes formula. The output is *bs_volat.dta**. It needs in input: *volat_raw.dta*, *chiavi.dta*

9. **varianza.do** It computes systematic and idiosyncratic risk needed for the sensitivity regressions. The output is *varianza.dta*.

10. **ceo.do** It combines data from *compensation_raw.dta*, *executive_raw.dta* and *company_raw.dta*, does more cleaning and produces *ceo.dta* as output

11. **survival.do** It uses information in *ceo.dta* to estimate CEOs' conditional survival rates. The output is *survival.dta*

12. **compensation.do** It merges datasets produced by modules above and computes compensation measures and descriptive statistics. It needs in input: *ceo.dta*, *price_moments.dta*, *options.dta**, *bs_volat.dta**, *bs_yield.dta**, *int_rate.dta**. It produces in output *ready_compensation.dta*

13. **stake.do** It produces Figure 2. It needs in input: *compensation.dta* and cpi.dta

14. **size_figure.do** It produces Figures 1, 3, 8 (left panel). It needs in input: *s&p.dta*, *ready_compensation.dta*, *cpi.dta*

15. **descriptive_split.do** It produces Figures 4, 5, 6, 7, 9 (right panel). It needs in input: *s&p.dta*, *ready_compensation.dta*, *cpi.dta*, *wages.dta*

16. **growth_split.do** It produces Figures 8 (right panel) and 9 (left panel). It needs in input: *s&p.dta*, *ready_compensation.dta*, *cpi.dta*, *wages.dta*

17. **regressioni_HL.do** It produces Figures 10, 11, as well as Tables 1 and 2. It needs in input: *ready_compensation.dta*, *varianza.dta*, *cpi.dta*

18. **size_table.do** It produces Tables A.3, A.4, A.5. It needs in input: s&p.dta, ready_compensation.dta, cpi.dta.

# 3 Description of the raw data files

In this section, we describe which raw variables (from EXECUCOMP, COMPUSTAT, and CRSP) should be included in the datasets that constitute input to the do files described above.

- **map.dta** This data is downloaded from CRSP/Compustat merged. It must contain the variables GVKEY, LPERMCO, DATADATE, FYEAR
- **company_22.dta** This data is downloaded directly from the EXECUCOMP dataset – ANNUAL COMPENSATION on WRDS. Table named COMPANY FINANCIAL AND DIRECTOR COMPENSATION. Here is the list of variables that this file must contain:

  - SALES, ASSETS, MKTVAL, DIVYIELD, PRCCF, PRCC, SHRSOUT, EMPL, AJEX, BS_VOLAT, BS_YIELD, GVKEY, CUSIP, YEAR, FYR

- **execucomp_22.dta** This is data downloaded directly from the EXECUCOMP dataset – ANNUAL COMPENSATION on WRDS. Here is the list of variables that this file must contain:
  - OLD_DATAFMT_FLAG is a dummy that takes value 1 when the recording standard is old, and takes value if the recording standards are new
  - OTHANN & OTHCOMP. OTHANN is the old variable. OTHCOMP is the new variable. However, when a record has OTHANN, has also OTHCOMP and OTHANN$\leq$OTHCOMP. Therefore I get rid of OTHANN.
  - RSTKGRNT & STOCK_AWARDS_FV. The former is the value of stock granted according to the old standards. STOCK_AWARDS_FV is the new measure, that potentially is computed differently
  - LTIP & NONEQ_INCENT. NONEQ_INCENT has the same definition as LTIP, with the new standards
  - SHROWN_EXCL_OPTS is the number of shares owned by the executive, excluding the options that are about to expire. By inspection, I verified that it only includes common stock. That is: it does not includes restricted stock holdings. In the new version, there is no dummy PINCLOPT.
  - OPTION_AWARDS_RPT_VALUE & OPTION_AWARDS_FV. The former coincides with SOPTVAL in the previous version of the dataset. OPTION_AWARDS_FV is the name of the variable for the records with the new standards
  - OPTION_AWARDS_NUM. This coincides with the old SOPTGRNT
  - OPTION_EXER_NUM. This coincides with the old SOPTEXSH
  - OPT_EXER_VAL. This coincides with the old SOPTEXER
  - OPT_UNEX_EXER_NUM. This coincides with the old UEXNUMEX
  - OPT_UNEX_UNEXER_NUM. This coincides with the old UEXNUMUN
  - OPT_UNEX_UNEX_EXER_EST_VAL. This coincides with the old INMONEX
  - OPT_UNEX_UNEX_UNEXER_EST_VAL. This coincides with the old INMONUN
  - STOCK_UNVEST_VAL. This coincides with the old RSTKHLDV
  - STOCK_UNVEST_NUM. This coincides with the old RSTKHLD
  - SALARY & BONUS
  - ALLOTHTO, ALLOTHPD, YEAR, GVKEY, CUSIP, NAICS, SIC, CONAME, CEOANN, TITLEANN, EXECDIR, PCEO, EXECID, PAGE, EXEC_LNA, EXEC_FNA, BECAMECE, JOINED_C, REJOIN, LEFTOFC, LEFTCO, RELEFT

- **s&p_raw.dta** This is data from CRSP (Indices and Deciles - Monthly). It must contain the variables CALDT, VWRETD

- **stock_raw.dta** This is data from CRSP MONTHLY STOCKS. It must be from 1985, because the calculations of volatilities require 60 month window around each date. Here is the list of variables that this file must contain:
  - CUSIP, PERMNO, TICKER, COMNAM, SHROUT (Shares outstanding), PRC (price), RET (holding period return), CFACPR (cumulative factor to adjust price), CFACSHR (cumulative factor to adjust shares)

- **volat_raw.dta** This is data from CRSP MONTHLY STOCKS. It must be from 1985, because the contribution of volatilities require 60 month window around each date. Here is the list of variables that this file must contain:
  - CUSIP, PERMNO, TICKER, COMNAM, DISTCD, SHROUT (Shares outstanding), PRC (price), RET (holding period return), RETX, CFACPR (cumulative factor to adjust price), CFACSHR (cumulative factor to adjust shares)