

BoardEx and ExecuComp Project Documentation (100322)

Summary of Main Processing Pipeline

	Task	Comments
1	Get ISINs and Company Names for Firms of Interest	
1.1	Start with a table of gvkeys and ISINs ("firms of interest"), and save the list of gvkeys as a .txt file.	
1.2	Search on WRDS using the gvkeys with no ISINs to get the ISINs. Multiple ISINs usually appear for every gvkey. These ISINs were added to the ISINs of interest.	
1.3	Search on WRDS using all gvkeys to get the company names. These company names were added as an additional variable to the firms of interest.	
2	Download ExecuComp Data from WRDS	
2.1	The "ExecuComp Data" comprises 15 tables. For tables that can be queried using gvkeys, the list of gvkeys was provided and the data downloaded. For tables that cannot be queried using gvkeys, the full dataset was downloaded on SAS Studio.	<ul style="list-style-type: none"> The date range was set to be as close as possible to 2000-current. All files are in .sas7bdat.
3	Download BoardEx Data on Box	
3.1	The "BoardEx Data" comprises 116 tables: 4 regions (NA, Europe, ROW, UK), each with the same set of 29 tables.	<ul style="list-style-type: none"> The WRDS account does not have access to BoardEx data on WRDS. Instead, the data is downloaded from a Box folder maintained by Booth, named 'BoardEx_Rest_of_World_2021'. All files are in .xlsx.
3.2	Open "Board Summary.xlsx" for all 4 regions, and collate a table of CompanyIDs, ISINs and Company Names. After removing duplicates (i.e. entries where all 3 fields are the same), there are 37323 entries.	<ul style="list-style-type: none"> Checks described below.
3.3	First check if the ISINs of interest match the BoardEx ISINs. While most boardex ISINs appear only once, a few appear twice, and 5 such ISINs matched the ISINs of interest. 1 case had the 2 ISINs refer to different firms (the wrong one was replaced with the correct one), while the other 4 cases had the 2 ISINs refer to the same firm (the second ISIN was added).	
3.4	The unmatched gvkeys were then filtered out. Fuzzy matching was done between the company names from WRDS and the company names in BoardEx. The top 2 matches in BoardEx were saved, and manual comparison was done to check for matches. The gvkeys of the matched firms were then linked to the company IDs in BoardEx.	
3.5	Save the final list of company IDs as a .xlsx file.	
4	Filter BoardEx files based on the company IDs.	
4.1	Filter BoardEx files based on the company IDs. The processing is done on Mercury, sped up with multiple job submission	

	and parallelized code. The code will do nothing for files without “company ID” as a variable.	
4.2	For each file type (e.g. Europe Compensation), combine the filtered segments into a bigger file.	

Details of ExecuComp Tables

Table	Status
ANNCOMP	Web Query, Non-Empty Result
CODIRFIN	Web Query, Non-Empty Result
COLEV	SAS Studio
COPEROL	SAS Studio
DEFERREDCOMP	Web Query, Non-Empty Result
DIRECTORCOMP	Web Query, Non-Empty Result
EXNAMES	SAS Studio
EX_BLACK	SAS Studio
EX_HEADER	SAS Studio
LTAWDTAB	Web Query, Non-Empty Result
OUTSTANDINGAWARDS	Web Query, Non-Empty Result
PENSION	Web Query, Non-Empty Result
PERSON	SAS Studio
PLANBASEDAWARDS	Web Query, Empty Result
STGRTTAB	Web Query, Non-Empty Result

Details of ExecuComp Tables

- Of the 29 tables, there are 13 tables for Directors, 13 analogous tables for SMDE (Senior Management, Disclosed Earners), and 3 separate tables (which should be for both Directors and SMDEs).
- The total size of all 29 tables is ~150GB. Big tables are split into multiple files.

Checks for Boardex “Board Summary.xlsx”

- All company IDs are unique.
- 11864 entries don’t have an ISIN. 35 ISINs appeared twice. All the other ISINs appeared once.
- The “ISIN” of a few entries is actually a list of multiple ISINs, e.g. “DE0006211162, DE0006211170, ... US0727352025”.
- After checking the 35 ISINs against the ISINs of interest, there are 5 matches. The actions taken are described as follows:

ISIN	Company ID*	Company Name	Action Taken
AU000000 AFG9	7.86E+09	ALLCO FINANCE GROUP LTD (Record Investments Ltd prior to 07/2006) (De-listed 09/2009)	Initial Match, Removed
AU000000 AFG9	2.36E+10	AUSTRALIAN FINANCE GROUP LTD (AFG)	Added

US00770C 1018	2103245 8	ADVANCED EMISSIONS SOLUTIONS INC (ADA-ES Inc prior to 07/2013) (De-listed 02/2015)	Initial Match, Retained
US00770C 1018	2.56E+10	ADVANCED EMISSIONS SOLUTIONS INC	Added
US045327 1035	1.71E+11	ASPEN TECHNOLOGY INC	Initial Match, Retained
US045327 1035	5.44E+08	ASPEN TECHNOLOGY INC (De-listed 02/2008)	Added
US41043F2 083	2.03E+08	HANGER INC (Hanger Orthopedic Group Inc prior to 06/2012) (De-listed 02/2016)	Initial Match, Retained
US41043F2 083	3.07E+11	HANGER INC	Added
US529898 1086	1.71E+11	LIBBEY INC (De-listed 06/2020)	Initial Match, Retained
US529898 1086	2.57E+09	LIBBEY INC (De-listed 04/2009)	Added

- All company names are unique, except for “CAPITAL TRUST LTD” which appeared twice. However, there are companies with duplicate entries, under names like “COMVERSE TECHNOLOGY INC (De-listed 02/2007)” and “COMVERSE TECHNOLOGY INC (De-listed 02/2013)”. These entries will have different company IDs but the same ISIN.