

Company and time identifiers

The dataset contains a large number of companies and identifies them by the variable **GVKEY**, a unique **six-digit ID** for all companies. GVKEY is developed by Compustat® and serves as its primary company identifier. For most applications in the dissertation, you don't have to know which company a GVKEY is associated with. However, in case you do need to know what individual firms are, you can check the variable CONM, which contains companies' full names.

The dataset spans over several years and indicates time by the variable **FYEAR** ("fiscal year"). A combination of GVKEY and FYEAR allows you to uniquely refer to each row (or observation) in the dataset. The dataset is sorted first by GVKEY and then by FYEAR. So all observations for a specific company are located adjacent to each other and ordered chronologically.

The dataset contains several other time identifiers. The variable **DATADATE** is the **fiscal year end date**, effectively, 31 December of each year in FYEAR. FYEAR and DATADATE can be used interchangeably.

Data sources and variable definitions

Accounting variables are from Compustat®, analysts' forecasts are from I/B/E/S, and betas and stock prices are from CRSP.

Compustat® collects accounting data, among others, directly from firms' financial statements. Most data items in Compustat® correspond to financial statement items. For this reason, the format of financial statements can be used to help you locate and understand data. A folder on Moodle called "Description of Compustat data" provides description of each variable, its measurement and location on financial statements. Following variables were obtained from Compustat:

gvkey datadate tic conm exchg cik costat fyear ajex act aqc at bkvlp capx ceq che csho cshpri dlc dp dpc dvc epspx ib invt ivch lct ni ppenb sale siv spi txt xad xint xrd prcc_c mkvalt sic rmum rank au auop auopic ap fincf gdwl intan ivncf oancf ppegt rect

Balance and income statement variables are measured in **\$ millions**. For auditor codes definitions, auditor opinion as well as other variable definitions please refer to the Compustat user guide. Note that Compustat® per share data are unadjusted, while I/B/E/S and CRSP data are adjusted following stock split/dividend. To be consistent, you should adjust Compustat® data for stock split/dividend as shown during the module. It is essential for you to understand all variable definitions and continually refer to these definitions when reporting and discussing results. *Failure to provide sufficient description of variables used in the analysis when reporting and discussing results will make it impossible for an examiner to understand what has been done. In such cases, examiners have no other option but to award a low mark.*

I/B/E/S (short for the *Institutional Brokers' Estimation System*) gathers and summarises analysts' forecasts and recommendations from a broad cross-section of equity analysts. The forecast data provided for your dissertation are "consensus forecasts". More specifically, they are the mean forecast computed using all available forecasts four months after the fiscal year end (for example, in April 2013 for a firm with a fiscal year end 31st December 2012). Analysts forecasts earnings and other financial measures over a range of horizons, say one-year-ahead, two-year-head or even further. Following variables were obtained from I/B/E/S:

ib_tic EPS_1 EPS_2 EPS_3 EPS_4 EPS_5 EPS_g DPS_1 DPS_2 DPS_3 DPS_4 DPS_5
DPS_g

All variables were renamed and therefore their names do not correspond directly to those reported in I/B/E/S (this was done because I/B/E/S does not follow a certain coding system for variable names). Therefore, when you describe your data in the dissertation, it would be misleading to state that you obtained variable “EPS_5” from I/B/E/S. Instead, you may state that the data on t+5 forecast of EPS was obtained from I/B/E/S. You may then explain that you chose to assign a name EPS_5 to this variable (or some other name, say EPS_{t+5}). ib_tic is the ticker provided by I/B/E/S. This ticker was created by I/B/E/S and thus does not correspond to ticker symbols allocated to firms by stock exchanges. EPS_1, EPS_2, ... is consensus (mean) analyst forecast of earnings per share (\$) for fiscal year t+1, t+2, ... EPS_g is the EPS long-term growth rate (%) forecasted by analysts. Similar, DPS_1, DPS_2, ... is consensus (mean) analyst forecast of dividends per share (\$) for fiscal year t+1, t+2, ...; and DPS_g is the DPS long-term growth rate (%).

CRSP is The Center for Research in Security Prices, which provides long-term market data for all U.S. quoted companies. This database covers information on stock prices, returns and stock betas. Following variables were obtained from CRSP:

permno date exchcd siccd cusip ticker comnam tsymbol naics trdstat,
and stock price “prc” four months after the fiscal year end (this variable was labeled “prc4”)

Most of the above variables are company identifiers or provide alternative industry groupings (based on NAICS). “prc4” is price per share at the end of fourth month after the fiscal year end (for example, in 30th April 2013 for a firm with a fiscal year end 31st December 2012). This is the date on which you will be carrying out the valuation. This price was used by numerous previous studies to assess the accuracy of valuation models. “beta” is the annual stock beta calculated using CRSP data on monthly firm returns and the value-weighted return index for the aggregate market. The variable “n” reports the number of previous months used to estimate firm beta (60 months were used to estimate beta; however, when fewer months were available in CRSP, beta was estimated using the data on 24-59 months). Please note that the dataset reports firm-specific beta in each of the sample years. The estimation strategy for beta generally follows Francis et al. (2000, p. 52) with two major exceptions: 1. Monthly returns are used instead of the daily returns; 2. The dataset reports firm-specific rather than industry-specific beta. Similar to other studies and Francis et al., you may want to use industry beta in your analysis, which can be computed as the mean (or median) of firm-specific annual betas for each two-digit SIC industry group and in each year.

In addition to firm beta, you will require information on the market risk premium and the risk-free rate to calculate the cost of equity capital. You can obtain information on risk free rates from public data sources (the FED, internet) and make an assumption about the market risk premium (similar to studies used in the module). Alternatively, the following website provides estimates of risk free rates and market premiums for each year:

<http://people.stern.nyu.edu/adamodar/>. Please read the website for how to reference these data.

Not all variables are available for all firm-years. Missing data may be due to a range of factors, the most common of which are that (a) the item is not available on Compustat® or (b) no analysts’ forecasts are available for forecast horizon k .

The nature of a variable determines its timeline, relative to the valuation date. Financial data are lagged behind share prices and analyst forecasts. In the following example, all accounting items from income statements or cash flow statements are for the fiscal year 2013 (ended 31st December 2013). All balance sheet items are as of 31st December 2013. Share prices (prc4) are collected on 30th April 2014, which is when valuation takes place. Analyst consensus forecasts was taken on 15th April 2014 to assure that EPS_1 refers to consensus analyst forecast of EPS for the fiscal year 2014.

conm	gvkey	fyear	act	at	ni	oancf	epspx	eps_1	prc4
INTL BUSINESS									
MACHINES CORP	006066	2013	51350	126223	16483	17485	15.06	17.9	196.47

Data selection

The observations (firm-years) in the dataset “valuation_data” are selected based on the following criteria:

1. U.S. firms publicly traded firms with data available on Compustat, CRSP and I/B/E/S;
2. Total assets from Compustat, analyst one-year ahead (EPS_1) forecasts from I/B/E/S, and CRSP price four months after the fiscal year end are not missing.

3. Using the data

The datasets allows you to conduct large sample empirical analyses. However, you are not required to use all the data in the dissertation.

Part of the ‘art’ of a good piece of empirical analysis is the sample selection process, which is driven almost exclusively by the research question(s) that are being investigated. Sample size, therefore, is not an important consideration when it comes to undertaking your empirical analysis, although it should be large enough to make your data analysis valid. Remember, it is important to select an appropriate sample that enables you to best answer your research question(s). Accordingly, the dataset provided merely represents the starting point from which to begin analysis. **Students therefore are encouraged to think carefully about the precise objectives of their empirical analysis BEFORE undertaking any detailed empirical analyses.**

4. Collecting additional data

The datasets described above contain sufficient observations and variables to enable you to address a wide range of research questions for the 703 dissertation. However, should you wish to supplement these data with additional variables and/or firms, you can do on your own by using accounting and financial databases such as Compustat®, I/B/E/S, and CRSP data. You require primary identifiers for major datasets to simplify the merging between multiple sources. This identifiers are provided as a part of the dataset described above. Specifically, GVKEY allows you to retrieve data from Compustat®; PERMNO allows you to retrieve stock prices/returns from CRSP; TIC_IB (is the same as TICKER on I/B/E/S) allows you to retrieve additional forecast data from I/B/E/S. Note, however, that the process of collecting and organising raw data is complex and time consuming. Only those students who are confident about their ability to successfully complete the dissertation and score a high mark

should attempt to collect additional data. We recommend that the majority of students restrict their analysis to the datasets provided.

Below are examples of Stata code that can assist you in merging additional data from Compustat or I/B/E/S.

Merge data from Compustat:

1. Download data from Compustat in Stata format; open your Compustat file in Stata.
2. `duplicates drop gvkey datadate, force`
3. `save new_name_comp, replace`
4. `use valuation_data, clear`
5. `duplicates drop gvkey datadate, force`
6. `merge 1:1 gvkey datadate using new_name_comp`
7. The table will appear that will show you how many observations were merged.
8. `drop _merge`

Merge data from I/B/E/S:

1. Download data from I/B/E/S in Stata format; open your I/B/E/S file in Stata.
2. `ren ticker ib_tic`
3. Choose relevant time variable.
4. `duplicates drop ib_tic your_time_variable, force`
5. `save new_name, replace`
6. `use valuation_data, clear`
7. Choose the same time variable as above.
8. `duplicates drop ib_tic your_time_variable, force`
9. `merge 1:1 ib_tic your_time_variable using new_name`
10. The table will appear that will show you how many observations were merged.
11. `drop _merge`

Appendix: Variable names

permno	PERMNO
date	Names Date
exchcd	Exchange Code
siccd	Standard Industrial Classification Code
cusip	CUSIP
ticker	Ticker Symbol
comnam	Company Name
tsymbol	Trading Symbol
naics	North American Industry Classification System
trdstat	Trading Status
beta	Beta
n	Number of previous months used for the estimation of beta (maximum = 60, minimum = 24)
prc4	Stock price 4 months after fiscal year end
ib_tic	IBES Ticker Symbol
EPS_1	Analyst forecast of EPS in year t+1
EPS_2	Analyst forecast of EPS in year t+2
EPS_3	Analyst forecast of EPS in year t+3
EPS_4	Analyst forecast of EPS in year t+4
EPS_5	Analyst forecast of EPS in year t+5
EPS_g	Analyst forecast of the long-term growth rate
DPS_1	Analyst forecast of DPS (dividends per share) in year t+1
DPS_2	Analyst forecast of DPS (dividends per share) in year t+2
DPS_3	Analyst forecast of DPS (dividends per share) in year t+3
DPS_4	Analyst forecast of DPS (dividends per share) in year t+4
DPS_5	Analyst forecast of DPS (dividends per share) in year t+5
DPS_g	Analyst forecast of the long-term dividend growth rate
gvkey	Standard and Poor's Identifier
datadate	Data Date
tic	Ticker Symbol
conm	Company Name
exchg	Stock Exchange Code
cik	CIK Number
costat	Active/Inactive Status Marker
fyear	Data Year - Fiscal
ajex	Adjustment Factor (Company) - Cumulative by Ex-Date
act	Current Assets - Total
aqc	Acquisitions
at	Assets - Total
bkvlp	Book Value Per Share
capx	Capital Expenditures
ceq	Common/Ordinary Equity - Total

che	Cash and Short-Term Investments
csho	Common Shares Outstanding
cshpri	Common Shares Used to Calculate Earnings Per Share - Basic
dlc	Debt in Current Liabilities - Total
dp	Depreciation and Amortization
dpc	Depreciation and Amortization (Cash Flow)
dvc	Dividends Common/Ordinary
epspx	Earnings Per Share (Basic) - Excluding Extraordinary Items
ib	Income Before Extraordinary Items
invt	Inventories - Total
ivch	Increase in Investments
lct	Current Liabilities - Total
ni	Net Income (Loss)
sale	Sales/Turnover (Net)
siv	Sale of Investments
spi	Special Items
txt	Income Taxes - Total
xad	Advertising Expense
xint	Interest and Related Expense - Total
xrd	Research and Development Expense
prcc_c	Price Close - Annual - Calendar
mkvalt	Market Value - Total - Fiscal
sic	Standard Industry Classification Code
rank	Rank - Auditor
au	Auditor
auop	Auditor Opinion
auopic	Auditor Opinion - Internal Control
ap	Accounts Payable - Trade
fincf	Financing Activities - Net Cash Flow
gdwl	Goodwill
intan	Intangible Assets - Total
ivncf	Investing Activities - Net Cash Flow
oancf	Operating Activities - Net Cash Flow
ppeg	Property, Plant and Equipment - Total (Gross)
rect	Receivables - Total