

Baker and Wurgler 2002 – Replication Project

0. Before you begin

- You will need a WRDS account and a skew3 account.
- Readings, resources and training materials for this exercise are posted in the following web page:
<http://www.kellogg.northwestern.edu/researchcomputing/workshops/summer2004.htm>
- Work through the examples in the “Data retrieval from CRSP and Compustat using WRDS's server” handout and become familiar with the contents of the Compustat and CRSP dataset collections.

1. Data required

The list of IPOs used in Baker and Wurgler (needed to create the Compustat sample) will be made available in the WRDS server in the following path: /projects/nwu/pledema/. The Graham et. al. data will be available in skew3 in /users/staff/ple531/graham/. [more details forthcoming as soon as they become available]

- Original lists of IPO dates (CUSIP and date) before any screens (received from authors). Sources: Jay Ritter (1968 through 1995); SDC for 1970 through 1998; SDC was used when the information was not available in Ritter's data. A SAS file with a list of IPOs that could be matched to Compustat is available in WRDS. For details, refer to <http://www.kellogg.northwestern.edu/researchcomputing/workshops/summer2004.htm>
- Measure of marginal corporate tax rate from Graham, Lemmon and Schallheim (1998). Get from John Graham.
- Compustat data items required: 6, 8, 10, 12, 13, 25, 35, 36, 56, 79, 181, 199, gvkey, dnum, yeara. Additional items (for other tables or variable construction from cash flow statement – se footnote 5): 14, 21, 46, 108, 111, 114, 115, 301.

2. Basic dataset construction

- Exclude firms with SIC code 6000-6999 (financial sector)
- IPO year is the first year in which Compustat reports market value. Regressions in “IPO time” (create IPO time variable); there are 4 cross-sections of interest: IPO+1, IPO+3, IPO+5, IPO+10.
- Second sample: All-firms, 1970-1999

3. Variable construction

<i>Variable mnemonic</i>	<i>Name in paper</i>	<i>Computation</i>
be	Book Equity, <i>BE</i>	Total Assets [data6] – Total Liabilities [data181] – Preferred Stock [data10] + Deferred Taxes [data35] + Convertible Debt [data79]. If Preferred Stock

<i>Variable mnemonic</i>	<i>Name in paper</i>	<i>Computation</i>
		[data10] is missing, replace it with Redemption Value of Preferred Stock [data56]
bd	Book Debt	Total Assets [data6] – Book Equity [be]
bl	Book Leverage, $\frac{D}{A}$	Book Debt / Total Assets [data6] * 100
me	Market Equity, ME	Common Shares Outstanding [data25] * Price [data199]
ml	Market Leverage, $\frac{D}{A}$	Book Debt / (Total Assets [data6] – Book Equity + Market Equity) * 100
e	Net Equity Issues, e	Δ Book Equity – Δ Balance Sheet Retained Earnings [data36])
e2a	Net Equity Issues, $\frac{e}{A}$	e / Total Assets [data6]
re2a	Newly Retained Earnings, $\frac{\Delta RE}{A}$	Δ Retained Earnings [data36] / Total Assets [data6]
d	Net Debt Issued, d	Residual Change in Assets = Δ Total Assets – e – ΔRE
d2a	Net Debt Issued, $\frac{d}{A}$	d / Total Assets [data6]
m2b	Market-to-Book, $\frac{M}{B}$	(Total Assets [data6] – Book Equity + Market Equity) / Total Assets [data6]
ppe2a	Asset tangibility, $\frac{PPE}{A}$	Net Plant, Property and Equipment [data8] / Total Assets [data6]
ebitda2a	Profitability, $\frac{EBITDA}{A}$	Earnings before Interest, Taxes and Depreciation [data13] / Total Assets [data6]
size	Size, $\log(S)$	$\log(\text{Net Sales [data12]})$
efwa_m2b	External Finance Weighted-Average Market-to-Book, $\left(\frac{M}{B}\right)_{efwa}$	$\sum_{s=0}^{t-1} \frac{e_s + d_s}{\sum_{r=0}^{t-1} e_r + d_r} \cdot \left(\frac{M}{B}\right)_s$ - see page 12 of paper

Additional variables used in the paper:

<i>Variable mnemonic</i>	<i>Name in paper</i>	<i>Computation</i>
cfe	(Equity Issues computed from Cash Flow Statement)	Sale of Common and Preferred Stock [data108] – Purchase of Common and Preferred Stock [data115]
cfD	(Debt Issues computed from Cash Flow Statement)	Long Term Issuance [data111] – Long Term Debt Reduction [data114] + Changes in Current Debt [data301]
div2be	Dividends over Book Equity, $\frac{Div}{BE}$	Common Stock Dividends [data21] / Book Equity
div2me	Dividends over Market Equity, $\frac{DIV}{ME}$	Dividends [data21] / Market Equity
dep2a	Depreciation Expense to Assets, $\frac{Dp}{A}$	Depreciation Expense [data14] / Total Assets [data6]
rd2a	R&D to Assets, $\frac{RD}{A}$	Research and Development [data46] / Total Assets [data6]
rdd	Dummy if firm does not report R&D, <i>RDD</i>	RDD=1 if data46 is missing; else RDD=0
loga	Log of Assets, $\log(A)$	$\log(\text{Total Assets [data6]})$

4. Screens/filters

Same screens are applied are applied to IPO sample and to All-firms sample:

- Exclude firms with minimum book value of assets below \$10 million
- Exclude firms without complete data on assets between IPO year and year the firm exits Compustat (All-firms: exclude firms with missing data on assets)
- Exclude individual firm-year outliers for capital structure: drop firm-year observations if $BL > 1$.
- Exclude individual firm-year outliers for market-to-book ratio where market-to-Book > 10 (Table II and Table III regressions)
- For regressions with efwa Market-to-Book, drop firm-year observations with efwa_m2b > 10 (Table III regressions).

5. Replicating Table I, variable definitions

- Definition of “size” variable: for firms with sales under \$500,000, sales are reported as 0 in Compustat. The size variable will be set to missing. How many firms are dropped

due to this problem? (Could this be a source of bias in the results?). Zero sales could be coded as a very small number instead (to 1 cent=\$0.00000001 million, for example) to avoid dropping these firms out of the sample.

- How would you deal with the firms with two IPO dates (repeated GVKEYs)?
- Definition of the IPO year: based on B&W, the IPO year is defined as the first year for which there is market value in Compustat (data25*data199). This is not necessarily the IPO year. How does this impact the results? Also, the regressions in the paper include lagged values. For “true” IPO data, this may drop observations since there should not be market value prior to the IPO year.
- Why eliminate firms with a minimum book value of assets below \$10 million? (Does this have a disproportionate effect on earlier IPOs?)

6. Replicating Tables II and III

B&W do not hold the sample of firms constant in the different regressions. This could be interesting for the IPO-time regressions.

- IPO-time regressions
- “All-firms” regressions – why not use clustering instead of Fama-MacBeth?