

A Data Appendix

Age (firm age) - number of months since the firm first appears on CRSP monthly return file.

An (number of analysts; analyst coverage) - the number of analysts covering the firm (from IBES detail file).

Announcement returns (the dependent variable in Table 8) - cumulative returns in the three days around earnings announcements (the day before the announcement, the day of the announcement, the day after the announcement). The earnings announcement dates are from Compustat. The return in each day is adjusted to size, market-to-book, and momentum by matching the firm to its size, market-to-book, and past performance quintile and deducting the average return of all firms in the intersection of those quintiles from the announcing firm return. The adjusted returns are then log-transformed, added up within the announcement window, and transformed back into simple returns.

Beta (market beta) - from the CAPM regression estimated separately for each firm using monthly returns from the past 36 months.

BLev (book leverage) - total liabilities (lt item from Compustat annual file) divided by total assets (at item).

CapEx - capital expenditures (Compustat item capx) divided by sales (Compustat item sale).

CMA (conservative-minus-aggressive, investment factor) - the arbitrage portfolio that buys firms in top 30% in terms of percentage growth in total assets and shorts firms in bottom 30% in terms of asset growth. The returns to the conservative-minus-aggressive strategy are value-weighted, the strategy is followed separately for small firms (below NYSE market cap median) and large firms. The value-weighted returns of the strategy in small and large firms subsample are then added and divided by two. CMA returns are from the website of Kenneth French.¹

Comp (sales-based complexity) - $1 - HHI$, where HHI is the Herfindahl index computed using segment sales, $HHI = \sum_{i=1}^N s_i^2$. N is the number of segments (from Compustat segment files, segments with the same two-digits SIC code are counted as one segment), s_i is the fraction of total sales generated by segment i .

CongAge (conglomerate age) - the number of consecutive years a firm is reported on Compustat segment files as having business segments with different two-digit SIC codes.

¹http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Conglo (conglomerate dummy) - 1 if the firm is a conglomerate, 0 otherwise. The firm is a conglomerate if it has business segments in more than one two-digit SIC industry.

Corr (cross-segment correlation) - computed using the formula in Hann, Ogneva, and Ozbas (2013) as

$$Corr = \sum_{i=1}^N \sum_{j=1}^N w_i \cdot w_j Corr_{t-10,t-1}(CF_i, CF_j), \quad (A-1)$$

where i and j are segment indices, N is the number of segments in the firm, w_i and w_j are the share of the segment's sales in total sales of the firm, and the correlation term is computed using quarterly data in the past ten years for average cash flows of single-segment firms in the same two-digit SIC code as the segment in question. Cash flows are defined following Sloan (1996) as operating income before depreciation (oibdpq) less the change in current assets (actq) plus the change in current liabilities (lctq) less the change in short-term debt (dlcq) plus the change in cash (cheq). The cash flows are scaled by average total assets (atq) in the past two quarters. All variables are from the Compustat quarterly file.

Credit rating - Standard and Poor's rating (splticrm variable in the Compustat adsprate file).

CVEarn (earnings volatility) - coefficient of variation (standard deviation over the average) of quarterly earnings measured in the past twelve quarters. Earnings are EPS scaled by lagged stock price (epsiq over prccq lagged by one quarter). All variables are from the Compustat quarterly file.

DEF (default spread) - defined as the yield spread between Moody's Baa and Aaa corporate bonds. The yields are obtained from the Federal Reserve Economic Data (FRED) database at <https://fred.stlouisfed.org/>.

DGTW-adjusted returns (the dependent variable in Tables 5-7, 9, and 11) - In each firm-month, the returns adjusted to size, market-to-book, and momentum by matching the firm to its size, market-to-book, and past performance quintile and deducting the average return of all firms in the intersection of those quintiles from the firm return (that is, the return of firm in, e.g., the second size quintile, third market-to-book quintile, and fifth past performance quintile is adjusted by deducting from it the average return to all firms in the second size quintile, third market-to-book quintile, and fifth past performance quintile. Past performance quintiles are formed using cumulative returns in months $t-2$ to $t-12$. All quintiles use NYSE (exchcd=1) breakpoints.

Disp (analyst forecast dispersion) - the standard deviation of all outstanding

earnings-per-share forecasts for the current fiscal year scaled by the absolute value of the outstanding earnings forecast (zero-mean forecasts and forecasts by only one analyst excluded). Earnings forecasts are from the IBES Summary file.

Div (firm-level dividend yield) - dividends (*dv* from Compustat annual file) over the product of shares outstanding (*csho*) and end-of-fiscal-year share price (*prcc.f*).

Diversification discount - the difference between conglomerate's market-to-book and the weighted average of imputed market-to-book ratios of its segments. Segment-level assets (*ias* item on Compustat single-segment files) are used to determine the weights. Imputed market-to-book ratio for a segment is average market-to-book of all single-segment firms with the same two-digit SIC code. Following the diversification discount literature, market-to-book is defined as market capitalization (*prcc.f* times *csho*) divided by total assets (*at*).

DY (dividend yield of the market index) - dividend yield of the CRSP market index, defined as the cumulative difference between its cum-dividend and ex-dividend return in the past 12 months.

Error (analyst forecast error) - the absolute value of the difference between the one-year-ahead consensus forecast and actual earnings divided by actual earnings. All variables are from the IBES Summary file.

Fee (estimated shorting fee) - the estimate of shorting fee from Boehme, Danielsen, and Sorescu (2006):

$$\begin{aligned}
 Fee = & 0.07834 + 0.05438 \cdot VRSI - 0.00664 \cdot VRSI^2 + 0.0003820 \cdot VRSI^3 - 0.5908 \cdot Option + \\
 & + 0.2587 \cdot Option \cdot VRSI - 0.02713 \cdot Option \cdot VRSI^2 + 0.0007583 \cdot Option \cdot VRSI^3
 \end{aligned}
 \tag{A-2}$$

where *VRSI* is the vicile rank of RSI (1 if the firm is in the bottom 5% in terms of RSI, 2 if the firm is between the 5th and 10th percentile, etc.) and *Option* is the dummy for presence of a listed option on the stock - 1 if an option on the stock has non-zero trading volume in this month, 0 otherwise. The trading volume for options is from OptionMetrics.

HiComp/HiSeg/HiRSZ (high complexity dummy) - HiComp/HiRSZ is 1 for conglomerates with Comp/RSZ measure above median, 0 otherwise. The median is determined separately within each year using conglomerates only. HiSeg is 1 for conglomerates with more than two business segments with different two-digit SIC codes, 0 otherwise. HiComp/HiSeg/HiRSZ is always zero for single-segment firms (which report only one business segments or several business segments with the same two-digit SIC code).

HP index (Hadlock and Pierce index) - financial constraints index computed

according to the formula in Hadlock and Pierce (2010):

$$HP = -0.737 \cdot \log(Size) + 0.043 \cdot \log^2(Size) - 0.04 \cdot Age, \quad (A-3)$$

where Age is the number of years the firm has been on Compustat, Size is price times shares outstanding (prcc.f times csho) from Compustat. Following Handlock and Pierce, Size is in millions of constant 2004 dollars, Age is winsorized at 37 years, Size is winsorized at 4.5 billions (i.e., all firms with Age greater or equal to 37 years are assumed to have Age of 37 years).

Intan (intangible assets) - intangible assets (intan item from the annual Compustat file divided by total assets (at item)).

Inv (investment-to-assets) - the change in capital expenditures (capx item from the annual Compustat file) in the previous year divided by total assets (at item from Compustat) in the previous year.

IO (institutional ownership) - the sum of institutional holdings from Thompson Financial 13F database, divided by the shares outstanding from CRSP. All stocks below the 20th NYSE/AMEX size percentile are dropped. If the stock is not dropped, appears on CRSP, but not on Thompson Financial 13Fs, it is assumed to have zero IO.

IVol (idiosyncratic volatility) - the standard deviation of residuals from the Fama-French model, fitted to the daily data for each month (at least 15 valid observations are required).

KZ index (Kaplan and Zingales index) - financial constraints index computed according to the formula in Kaplan and Zingales (1997):

$$KZ_t = -1.002 \cdot (NI_t + Depr_t)/PPE_{t-1} + 0.2826 \cdot (Size_t - ComEq_t - DefTax_t)/TotEq_t + 3.139 \cdot (LTDebt_t + STDebt_t)/PPE_{t-1} - 39.3678 \cdot Div_t/PPE_{t-1} - 1.315 \cdot Cash_t/PPE_{t-1}, \quad (A-4)$$

where NI is income before extraordinary items (ib item from Compustat annual), Depr is depreciation (dp item), PPE is plants, property, and equipment (ppe item), Size is market capitalization at the end of the fiscal year (prcc.f times csho), ComEq is common equity (ceq item), DefTax is deferred taxes (txdb item), TotEq is total equity (teq item), LTDebt is long-term debt (sum of dltt and dd1 items), STDebt is short-term debt (np item), Div is dividends (div item), Cash is cheq item.

Lev (leverage) - long-term debt (dltt) plus short-term debt (dlc) divided by equity value (prcc.f times csho), all items from Compustat annual file.

LoComp/LoSeg/LoRSZ (low complexity dummy) - LoComp/LoRSZ is 1 for conglomerates with Comp/RSZ measure below median, 0 otherwise. The median is determined separately within each year using conglomerates only. LoSeg is 1 for conglomerates with exactly two business segments with different two-digit SIC codes, 0 otherwise. LoComp/LoSeg/LoRSZ is always zero for single-segment firms (which report only one business segments or several business segments with the same two-digit SIC code).

Margin (profit margin) - the ratio of net income before depreciation (Compustat item oibdp) to sales (Compustat item sale).

MB (market-to-book) - equity value (csho times prcc.f) divided by book equity (ceq) plus deferred taxes (+txdb), all items from the Compustat annual file.

Mom (cumulative past return) - cumulative monthly return to the stock between month t-2 and t-12, monthly returns are from CRSP.

Nasdaq (Nasdaq dummy) - 1 if the firm is a Nasdaq firm, 0 otherwise. A firm is classified as a NASDAQ firm if its CRSP events file listing indicator - exched - is equal to 3.

NewConglo1/NewConglo2/NewConglo3 (new conglomerate dummy) - 1 if the firm became a conglomerate in the past one/two/three years, zero otherwise. Single-segment firms always have NewConglo=0.

NSeg (number of segments) - the number of business segments the firm has (from Compustat segment files). Segments with the same two-digit SIC code are counted as one segment.

OverPrice (Stambaugh et al. overpricing measure) - average percentage ranking from ranking firms on 11 priced characteristics (accruals, momentum, ROA, asset growth, gross profitability, investment-to-assets, net stock issuance, cumulative stock issuance, net operating assets, O-score, expected probability of bankruptcy) in a way that ranks the most overpriced firm (highest accruals or lowest momentum) as 100, the most underpriced firm as 1, and firms in between get a corresponding percentage rank (e.g., a firm ranked 50th most overpriced in a sample of 2000 will get a rank of 97.5). The average is taken for all firms with at least five rankings available, otherwise OverPrice is set to missing. Detailed description of the process and the sorting variables is in Stambaugh, Yu, and Yuan (2015).

Price (reciprocal of stock price) - 1 over the stock price from CRSP monthly files.

Prob (probability to be on special) - defined as in D'Avolio (2002) and Ali and

Trombley (2006)

$$Sh = \frac{e^y}{1 + e^y}, \quad (\text{A-5})$$

$$y = -0.46 \cdot \log(\text{Size}) - 2.8 \cdot \text{Inst} + 1.59 \cdot \text{Turn} - 0.09 \cdot \frac{CF}{TA} + 0.86 \cdot \text{IPO} + 0.41 \cdot \text{Glam} \quad (\text{A-6})$$

In the equation above, Size is defined as shares outstanding times the price per share and measured in millions, Inst is IO, Turn is monthly turnover, defined as the trading volume over shares outstanding, CF is cash flow defined as operating income before depreciation (oiadp plus dp) less non-depreciation accruals, which are change in current assets (act) less change in current liabilities (lct) plus change in short-term debt (dlc) less change in cash (che), TA are total assets (at), IPO is the dummy variable equal to 1 if the stock first appeared on CRSP 12 or less months ago, and Glam is the dummy variable equal to 1 for three top market-to-book deciles. All items in the definitions of CF and TA are from the Compustat annual file.

Prof (gross profitability) - total revenue (sale) minus cost of goods sold (cogs) divided by book value of equity (ceq plus txdb), , all items from the Compustat annual file.

R&D (R&D-to-assets) - research-and-development expenditures (xrd item from Compustat) divided by total assets (at item from Compustat) in the previous year.

RetQ1 (return in the past quarter) - cumulative monthly return (from CRSP) in the past quarter.

RetYR1/RetYR2 (return in the past year/two years ago) - cumulative monthly return (from CRSP) in the past year/in the year before that.

Rev (short term reversal) - stock return (from CRSP) in month t-1.

RInst (residual IO) - the residual (ϵ) from the logistic regression of IO on log Size and its square

$$\log\left(\frac{\text{Inst}}{1 - \text{Inst}}\right) = \gamma_0 + \gamma_1 \cdot \log(\text{Size}) + \gamma_2 \cdot \log^2(\text{Size}) + \epsilon \quad (\text{A-7})$$

RMW (robust-minus-weak, profitability factor) - the arbitrage portfolio that buys firms in top 30% in terms of profitability and shorts firms in bottom 30% in terms of profitability. The returns to the robust-minus-weak strategy are value-weighted, the strategy is followed separately for small firms (below NYSE market cap median) and large firms. The value-weighted returns of the strategy in small and large firms subsample are then added and divided by two. RMW returns are from the website of Kenneth French.²

²http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

RSI (residual short interest) – outstanding shorts reported by NYSE and NASDAQ divided by the number of shares outstanding. The data are monthly and reported on the 15th calendar day of each month.

RSZ (RSZ complexity measure) - standard deviation of imputed segment-level market-to-book ratios divided by the weighted average imputed market-to-book ratios of all segments. Segment-level assets (ias item on Compustat single-segment files) are used to determine the weights used to compute the standard deviation and the weighted average. Imputed market-to-book ratio for a segment is average market-to-book of all single-segment firms with the same two-digit SIC code.

S&P500 (S&P 500 dummy) - 1 for firms that are included in S&P500, 0 otherwise. Index membership is from Compustat sec_mth file, S&P 500 have spmim variable equal to 10.

SegInc1/SegInc2/SegInc3 (segment increase dummy) - 1 if the firm experienced an increase in the number of segments in the past two years, zero otherwise. Single-segment firms always have SegInc=0.

Size (market cap) - shares outstanding times price, both from the CRSP monthly returns file.

Skew (total skewness) - skewness of daily stock returns (from CRSP daily file) measured separately for each firm-year.

Spec (number of specialists) - the number of analysts covering the firm who are specialists in the firm's industry. An analyst is considered a specialist in the firm's industry if he/she covers at least five other firms with the same two-digit SIC code in the same quarter. For a conglomerate, an analyst is classified as a specialist based on the industry affiliation of the largest segment.

Special (probability to be on special) - defined as in D'Avolio (2002) and Ali and Trombley (2006)

$$Special = \frac{e^y}{1 + e^y}, \quad (1)$$

$$y = -0.46 \cdot \log(Size) - 2.8 \cdot IO + 1.59 \cdot Turn - 0.09 \cdot \frac{CF}{TA} + 0.86 \cdot IPO + 0.41 \cdot Glam \quad (2)$$

Size is in million dollars, *Turn* is turnover, defined as the trading volume over shares outstanding (from CRSP). *CF* is cash flow defined as Compustat item OIADP plus Compustat item DP) less non-depreciation accruals, which are change in current assets (Compustat item ACT) less change in current liabilities (Compustat item LCT) plus change in short-term debt (Compustat item DLC) less change in cash (Compustat item CHE). *TA*

are total assets (Compustat item AT), *IPO* is the dummy variable equal to 1 if the stock first appeared on CRSP 12 or less months ago, and *Glam* is the dummy variable equal to 1 for three top market-to-book deciles.

TB (Treasury bill rate) - the 30-day T-bill rate from the FRED database at <https://fred.stlouisfed.org/>.

TED (TED spread) - the difference between the three-month LIBOR based in US dollars and the three-month Treasury bill rate as reported by the FRED database at <https://fred.stlouisfed.org/>

TERM (term spread) - the yield spread between the ten-year and the one-year Treasury constant-maturity bond from the FRED database at <https://fred.stlouisfed.org/>.

Turn (turnover) - trading volume divided by shares outstanding (both from CRSP monthly data). The monthly turnover is then averaged in each calendar year with at least 5 valid observations. To make comparisons across exchanges more meaningful, I adjust NASDAQ volume for the double counting following Gao and Ritter (2010): NASDAQ volume is divided by 2 for the period from 1983 to January 2001, by 1.8 for the rest of 2001, by 1.6 for 2002-2003, and is unchanged after that. A firm is classified as a NASDAQ firm if its CRSP events file listing indicator (exchcd) is equal to 3.

TVol (total volatility) - volatility of monthly returns (from CRSP) measured in the past 24 months.

TVolD (daily total volatility) - volatility of daily returns (from CRSP) measured in the past year.

WW index (Whited-Wu index) - financial constraints index computed according to the formula in Whited and Wu (2006):

$$WW = 0.652 - 0.091 \cdot (NI + Depr)/TA - 0.062 \cdot DivPos + 0.021 \cdot LTDebt/TA - 0.044 \cdot \log(TA) + 0.102 \cdot ISG - 0.035 \cdot SG, \quad (\text{A-8})$$

where NI is net income (ni item from Compustat annual), Depr is depreciation (dp item), TA is total assets (at item), DivPos is the dividend payer dummy (1 if either dv or dvc item is positive, 0 otherwise), LTDebt is long-term debt (dltt item), SG is sales growth, i.e., the change in sales (sale item from Compustat) as a fraction of last year sales: $SG_t = \frac{Sales_t - Sales_{t-1}}{Sales_{t-1}}$, ISG is industry sales growth (the average SG for all firms with the same three-digit SIC code).