A Data Appendix

 # An (number of analysts; analyst coverage) – the number of analysts covering the firm (from IBES summary file).

Accruals – change in current assets (Compustat item actq) less change in cash (cheq) less change in current liabilities (lctq) plus change in short-term debt (dlcq) plus change in taxes payable (txpq) less depreciation (dpq). Accruals are scaled by either lagged market cap (item cshprq times item prccq) for the use in the Basu (1997) regression instead of earnings-to-price by the average of current and lagged total assets (item atq) for use in calculating discretionary accruals or cash flows.

Age – the number of months the stock has appeared on CRSP.

Amihud (Amihud illiquidity measure) – the average ratio of absolute return to dollar volume, both from CRSP. The ratio is computed daily and averaged within each firm-year (firms with less than 200 valid return observations in a year and a stock price of less than $5 at the end of the previous year are excluded).

Beta (market beta) – market beta from fitting the CAPM to monthly stock returns between months t and t-36.

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]

C-score – unconditional conservatism measure from Penman and Zhang (2002), defined as

$$C - Score_t = \frac{LIFO_t + R\&DRes_t + AdvRes_t}{NOA_t}, \tag{A-1}$$

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

where LIFO is inventory reserve created by LIFO use (Compustat item lifr), R&D Res is the reserve created by R&D expensing, which assumes that R&D has a useful life of five years and applies sum-of-the-years method of depreciation, so that

$$R\&DRes_t = R\&D_t + \frac{10}{15} \cdot R\&D_{t-1} + \frac{6}{15} \cdot R\&D_{t-2} + \frac{3}{15} \cdot R\&D_{t-3} + \frac{1}{15} \cdot R\&D_{t-4}, \tag{A-2}$$

where R&D is Compustat item xrd; Adv Res is the reserve created by advertising expensing, which assumes that advertising expenses have a useful life of two years and applies sum-of-the-years method of depreciation, so that

$$AdvRes_t = Adv_t + \frac{1}{3} \cdot Adv_{t-1}, \tag{A-3}$$

where Adv is Compustat item xad; and NOA (net operating assets) is common equity (Compustat item ceq) plus long-term (dltt) plus short-term debt (dlc) plus preferred equity (pstk) minus cash and short-term investments (che) minus investment and advances (ivao) minus minority interest (mib).

Cash flows volatility – standard deviation of quarterly cash flows measured in the past 12 quarters. See CVCFO below for definition of cash flows.

Cred (credit rating) – Standard and Poor's rating (splticrm variable in the Compustat adsprate file). The credit rating is coded as 1=AAA, 2=AA+, 3=AA, ... , 21=C, 22=D.

CV DA – coefficient of variation (standard deviation divided by average) of discretionary accruals computed as in Dechow and Dichev (2002), who define discretionary accruals using coefficients from the cross-sectional regression (without intercept):

$$Acc_t = c_1 \cdot \frac{1}{TA_{t-1}} + c_2 \cdot \frac{Sales_t - Sales_{t-1}}{Sales_{t-1}} + c_3 \cdot \frac{PP\&E_t}{TA_{t-1}}, \tag{A-4}$$

where TA is total assets (Compustat item atq), Sales is total revenue (Compustat item saleq), PP&E is plant, property, and equipment (Compustat item ppegtq), and Acc is accruals (see Accruals above). and compute discretionary accruals (DA) as

$$DA_t = Acc_t - c_1 \cdot \frac{1}{TA_{t-1}} + c_2 \cdot \left(\frac{Sales_t - Sales_{t-1}}{Sales_{t-1}} - \frac{Rec_t - Rec_{t-1}}{AT_{t-1}} \right) + c_3 \cdot \frac{PP\&E_t}{TA_{t-1}} \tag{A-5}$$

where Rec is receivables (Compustat item rectq).

CVEarn/CVCFO (earnings/cash flows volatility) – coefficient of variation (standard deviation over the average) of quarterly earnings/cash flows measured in the past 12 quarters. Earnings are EPS (epspiq over prccq lagged by one quarter). Cash flows are operating income before depreciation (oibdpq) less accruals (see Accruals above). The cash flows are scaled by average total assets (atq) in the past two years. All variables are from the Compustat quarterly file.

DC/DR/DR6/DS (dummy for negative cash flows / negative CAPM CAR / negative FF6 CAR / negative sales growth) – 1 if cash flows / CAR / sales growth is negative in the current quarter and 0 otherwise. CFO is defined above under CVCFO. Sales growth is year-on-year change in Compustat sale item. CAR is cumulative abnormal return, cumulated between the day after the previous earnings announcement and the day after the current earnings announcement.

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/.

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

DY (dividend yield) – 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.

EffTick (effective tick size) – measure of effective spread from Holden (2009). On the simple $\$\frac{1}{8}$ grid, frequency of odd $\frac{1}{8}$s prices (prices that end with $\frac{1}{8}$, $\frac{3}{8}$, $\frac{5}{8}$, or $\frac{7}{8}$) measures the probability of the bid-ask spread being equal to $\$\frac{1}{8}$, the frequency of odd $\frac{1}{4}$s prices measures the probability of the bid-ask spread being equal to $\$\frac{1}{4}$, the frequency of the prices that end in $\frac{1}{2}$ measures the probability of the bid-ask spread being $\$\frac{1}{2}$, and the

frequency of whole-dollar prices measures the probability of the spread being \$1. For each firm-month, I estimate the probabilities of the spread as above and compute its expected value by multiplying the probabilities by the respective spread values. I use the $\$\frac{1}{16}$ grid before 2001 (decimalization) and the grid with clustering on dollars, half-dollars, quarters, dimes, nickels, and cents from 2001 on.

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

IO (institutional ownership) – the sum of institutional holdings from Thomson 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 Thomson Financial 13Fs, it is assumed to have zero institutional ownership.

Inv (investment-to-assets) – the annual change in gross PP&E (ppegt item from Compustat) divided by total assets (item at) from the previous year.

IVol (idiosyncratic volatility) – the standard deviation of residuals from the Fama-French (1993) model, fitted to the daily data for each firm-month (at least 15 valid observations are required).

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

MB (market-to-book) – market cap (share price, prcc, times number of shares outstanding, csho) divided by book equity (ceq) plus deferred taxes (txdb), all items from Compustat annual files.

Mom (cumulative past return) – in cross-sectional regressions, cumulative return to the stock between month t-2 and t-12, returns are from CRSP monthly returns file.

MOM (momentum factor) – in time-series regressions, the arbitrage portfolio that buys top 30% of recent winners and shorts bottom 30% of recent losers. Winners and losers are defined using cumulative return to the stock between month t-2 and t-12. The returns to the winners-minus-losers 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. MOM returns are from the website of Kenneth French.[2]

Price – stock price from the CRSP monthly file.

Q-score – conservatism measure from Penman and Zhang (2002), defined as the difference between current C-score (see C-Score above) and the average between lagged C-score and industry-average C-score (industries are defined using two-digit SIC codes).

Rev (reversal) – stock return in the past month, from CRSP monthly files.

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.[3]

Roll (Roll measure) – $Roll_t = 200 \cdot \sqrt{abs(Cov(R_t, R_{t-1}))}$, where R_t are daily stock returns. Roll measure is computed within each firm-year.

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

Spread – the spread implied by the daily high and low prices. Spread is calculated by the formula from Corwin and Schultz (2012):

$$\text{Spread} = \frac{2 \cdot (\exp^{\alpha} - 1)}{1 + \exp^{\alpha}}, \tag{A-6}$$

where

$$\alpha = \frac{\sqrt{\beta} \cdot (\sqrt{2} - 1)}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}, \quad \text{where} \tag{A-7}$$

$$\beta = \log^2\left(\frac{HI_t}{LO_t}\right) + \log^2\left(\frac{HI_{t+1}}{LO_{t+1}}\right) \quad and \quad \gamma = \log^2\left(\frac{\max(HI_t,\ HI_{t+1})}{\min(LO_t,\ LO_{t+1})}\right) \tag{A-8}$$

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

where HI_t (LO_t) is the highest (lowest) price of the stock on day t. Spread is computed within each firm-year.

TB (Treasury bill rate) – the 30-day T-bill rate from 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/.

TED spread – the spread between three-month LIBOR based on US dollars and three-month Treasury bill yield as reported in the FRED database at https://fred.stlouisfed.org/.

VIX – the VIX index, defined as the implied volatility of at-the-money options on S&P 100 (current ticker VXO). VIX is computed by CBOE and obtained from WRDS.

Vol, total volatility – annual volatility of CRSP daily returns, estimated annually (at least 100 non-missing daily returns are required).

Vol DA – standard deviation of discretionary accruals (see CV DA above for the definition of discretionary accruals).

Zero (zero frequency) – the fraction of zero-return days within each firm-year.