# A Data Appendix

The variables are arranged in alphabetical order according to the abbreviated variable name used in the tables.

**Beta (market beta)** - market beta from fitting the CAPM to monthly stock returns between months $t$-1 and $t$-37. Market beta is available from January 1964 to December 2010.

**$\beta_{FVIX}$ (FVIX beta)** - in cross-sectional regressions (Tables 2 and 7), FVIX beta is estimated by running

$$Ret_t - RF_t = \alpha + \beta_{MKT} \cdot (MKT_t - RF_t) + \beta_{FVIX} \cdot FVIX_t, \tag{A-1}$$

separately for each firm-month using the data from months $t$-1 to $t$-36. In time-series regressions (Tables 6 and 8–10), FVIX beta is estimated by fitting the same regression to the returns to the respective portfolio using the full sample. In cross-sectional (time-series) regressions, FVIX beta is available from January 1986 (January 1988) to December 2010.

**$\beta_{PS}$ (Pastor-Stambaugh beta wrt the non-traded factor)** - the loading from the firm-level regression

$$Ret_t - RF_t = \alpha + \beta_{MKT}(MKT_t - RF_t) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{PS} \cdot \overline{\gamma}_t^U. \tag{A-2}$$

The Pastor-Stambaugh (2003) non-traded factor, $\overline{\gamma}_t^U$, is the innovation to the market-wide monthly average of the Pastor-Stambaugh gamma, $\gamma_{PS}$ (available from WRDS). $\beta_{PS}$ is available from January 1965 to December 2010.

**$\beta_{PS-T}$ (Pastor-Stambaugh beta wrt the traded factor)** - the loading from the firm-level regression

$$Ret_t - RF_t = \alpha + \beta_{MKT}(MKT_t - RF_t) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{PS-T}PS_t. \tag{A-3}$$

The Pastor-Stambaugh traded factor, $PS_t$, (also available from WRDS) is the value-weighted return differential between the top and bottom deciles from the sorts on the expected value of $\beta_{PS}$, the Pastor-Stambaugh beta wrt the non-traded factor described above. $\beta_{PS-T}$ is available from January 1969 to December 2010.

***$\beta_{Sad}$* (Sadka beta wrt the non-traded factor)** - the loading from the firm-level regression

$$Ret_t - RF_t = \alpha + \beta_{MKT}(MKT_t - RF_t) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{Sad} \cdot \overline{\lambda}_t^U. \tag{A-4}$$

The Sadka (2006) non-traded factor, $\overline{\lambda}_t^U$, is the innovation to the market-wide average of the variable (information-based) price impact or Kyle's $\lambda$ (available from CRSP). $\beta_{Sad}$ is available from January 1986 to December 2010.

***$\beta_{Sad-T}$* (Sadka beta wrt the traded factor)** - the loading from the firm-level regression

$$Ret_t - RF_t = \alpha + \beta_{MKT}(MKT_t - RF_t) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{Sad-T}Sad_t. \tag{A-5}$$

The Sadka traded factor is the factor-mimicking portfolio that mimics the innovation to the market-wide average of the variable (information-based) price impact or Kyle's $\lambda$ (that is, the Sadka non-traded factor) by regressing the innovation on the excess returns to the six size and book-to-market portfolios from Kenneth French's website:

$$\begin{aligned} \Delta\overline{\lambda}_t^U = \quad & \gamma_0 + \gamma_1 \cdot (SG_t - RF_t) + \gamma_2 \cdot (SN_t - RF_t) + \gamma_3 \cdot (SV_t - RF_t) + \\ & + \gamma_4 \cdot (LG_t - RF_t) + \gamma_5 \cdot (LN_t - RF_t) + \gamma_6 \cdot (LV_t - RF_t), \end{aligned} \tag{A-6}$$

where $S$ ($L$) stands for small (large) firms, and $G$ ($N$, $V$) stands for growth (neutral, value) firms. The Sadka traded factor is the fitted part of this regression less the constant:

$$\begin{aligned} Sad_t = \quad & \hat{\gamma}_1 \cdot (SG_t - RF_t) + \hat{\gamma}_2 \cdot (SN_t - RF_t) + \hat{\gamma}_3 \cdot (SV_t - RF_t) + \\ & + \hat{\gamma}_4 \cdot (LG_t - RF_t) + \hat{\gamma}_5 \cdot (LN_t - RF_t) + \hat{\gamma}_6 \cdot (LV_t - RF_t). \end{aligned} \tag{A-7}$$

$\beta_{Sad-T}$ is available from January 1989 to December 2010.

***$\beta_{Ami}$* (Amihud beta wrt the non-traded factor)** - the loading from the firm-level regression

$$Ret_t - RF_t = \alpha + \beta_{MKT}(MKT_t - RF_t) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{Ami}\overline{Illiq}_t^U. \tag{A-8}$$

The Amihud (2002) non-traded factor, $\overline{Illiq}_t^U$, is the innovation to the monthly market-wide average of the Amihud measure (Illiq). Illiq is computed separately for each firm-month (at least 15 valid return and volume observations within each firm-month are required). The innovation is from the AR(1) model fitted to the average Amihud measure. $\beta_{Ami}$ is available from January 1965 to December 2010.

**$\boldsymbol{\beta_{Ami-T}}$ (Amihud beta wrt the traded factor)** - the loading from the firm-level regression

$$Ret_t - RF_t = \alpha + \beta_{MKT}(MKT_t - RF_t) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{Ami}Ami_t. \tag{A-9}$$

The Amihud traded factor is the factor-mimicking portfolio that mimics the innovation to the market-wide average of the Amihud measure (that is, the Amihud non-traded factor) by regressing the innovation on the excess returns to the six size and book-to-market portfolios, as in (A-6). The Amihud traded factor is the fitted part of this regression less the constant, as in (A-7). $\beta_{Ami-T}$ is available from January 1965 to December 2010.

**Cred (credit rating)** - Standard and Poor's rating (spdr variable in the Compustat quarterly file). The credit rating is coded as 1=AAA, 2=AA+, 3=AA, ... , 21=C, 22=D. Credit rating for a broad cross-section of stocks is available starting in January 1985.

**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 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 years. All variables are from the Compustat quarterly file. Compustat quarterly files cover a broad cross-section of stocks starting in 1974; CVEarn and CVCFO are available from January 1976 to December 2010.

**D, Delay (price delay)** - $D = 1 - \dfrac{R^2_{full}}{R^2_{short}}$, where $R^2_{short}$ is the $R$-square from the firm-level regression of weekly firm returns on weekly market returns

$$R_t - RF_t = \alpha + \beta(MKT_t - RF_t), \tag{A-10}$$

and $R^2_{full}$ is the $R$-square from the same regression with four lags of the market return added:

$$R_t - RF_t = \alpha + \beta_0(MKT_t - RF_t) + \beta_1(MKT_{t-1} - RF_{t-1}) + \beta_2(MKT_{t-2} - RF_{t-2}) + \\ + \beta_3(MKT_{t-3} - RF_{t-3}) + \beta_4(MKT_{t-4} - RF_{t-4}). \tag{A-11}$$

The weekly returns are computed as Wednesday close to Wednesday close. For more details, please refer to Hou and Moskowitz (2006).

**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. Analyst coverage prior to 1980 is incomplete, hence, Disp is available from January 1980 to December 2010.

**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. Error is available from January 1980 to December 2010.

**FVIX beta** - see $\boldsymbol{\beta_{FVIX}}$.

$\boldsymbol{\gamma_{VIX}}$ **(VIX loading)** - in cross-sectional regressions (Tables 2 and 7), VIX loading is estimated by running

$$Ret_t - RF_t = \alpha + \beta_{MKT} \cdot (MKT_t - RF_t) + \gamma_{VIX} \cdot \Delta VIX_t, \tag{A-12}$$

separately for each firm-month using data from months $t$-1 to $t$-36. $\Delta VIX_t$ is the change in VIX in month $t$. $\gamma_{VIX}$ is available from January 1988 to December 2010.

$\boldsymbol{\gamma_{PS}}$ **(Pastor-Stambaugh gamma)** - the firm return sensitivity to the firm previous-day dollar volume times the sign of the previous-day return, from

$$R_{t+1} = \theta + \phi R_t + \gamma_{PS} \cdot sign(R_t) \cdot Vol_t. \tag{A-13}$$

Both the returns and the volume are from CRSP. The dollar volume is scaled by the ratio of current total market value of NYSE and AMEX shares to the total market value of NYSE and AMEX shares in January 1963. $\gamma_{PS}$ is computed only for NYSE (exchcd=1) and AMEX (exchcd=2) shares and is available from January 1964 to December 2010.

**Gibbs (Gibbs measure)** - the slope from the regression $\Delta P_t = a + c\Delta Q_t$, where $P_t$ is the stock price and $Q_t$ is the trade direction indicator. The values of the Gibbs measure come from the website of Joel Hasbrouck and are available from January 1964 to December 2009. For more details, please refer to Hasbrouck (2009).

**Illiq (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 the stock price of less than \$5 at the end of the previous year are excluded). Illiq is available from January 1964 to December 2010.

**IVol (idiosyncratic volatility)** - the standard deviation of residuals from the Fama-French (1993) model, fitted to the daily data for each month (at least 15 valid observations are required). IVol is available from January 1964 to December 2010.

**MB (market-to-book)** - equity value (share price, prcc, times number of shares outstanding, csho) divided by book equity (ceq) plus deferred taxes (txdb), all items from Compustat annual files. MB is available from January 1964 to December 2010.

**Mom (cumulative past return)** - cumulative return to the stock between month $t$-2 and $t$-12.

**NBER (the NBER recession dummy)** - one for the months between NBER-announced peak and trough and zero otherwise.

**O-score** - the probability of bankruptcy measure from Ohlson (1980), computed as

$$O = -1.32 - 0.407 \cdot \ln TA + 6.03 \cdot \frac{TL}{TA} - 1.43 \cdot \frac{WC}{TA} + 0.076 \cdot \frac{CL}{CA} - 1.72 \cdot I(TL > TA)$$
$$-2.37 \cdot \frac{NI}{TA} - 1.83 \cdot \frac{FFO}{TA} + 0.285 \cdot I(NI < 0 \ \& \ NI_{-1} < 0) - 0.521 \cdot \frac{NI - NI_{-1}}{|NI| + |NI_{-1}|}, \tag{A-14}$$

where $TA$ is the book value of total assets (Compustat item at), $TL$ is the book value of total liabilities (lt), $WC$ is working capital (wcap), $CL$ are current liabilities (lct), $CA$ are current assets (act), $NI$ is net income (ni), $NI_{-1}$ is the previous year net income, $FFO$ are funds from operation (pi plus dp), $I(TL > TA)$ is the dummy variable equal to one if the book value of total liabilities exceeds the book value of total assets, and equal to zero otherwise, $I(NI < 0 \ \& \ NI_{-1} < 0)$ is the dummy variable equal to one if the net

income was negative in the two most recent years, and equal to zero otherwise. O-score is available from January 1965 to December 2010.

**Realized (realized market volatility)** - the square root of the average squared daily return to the market portfolio (CRSP value-weighted index) within each given month.

**Rev (reversal)** - stock return in the past month.

**RI (residual IO)** - the residual ($\epsilon$) from the logistic regression of IO on log Size and its square

$$log(\frac{Inst}{1-Inst}) = \gamma_0 + \gamma_1 \cdot log(Size) + \gamma_2 \cdot log^2(Size) + \epsilon. \tag{A-15}$$

IO is the sum of institutional holdings from the 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. For more details, please refer to Nagel (2005).

**Roll (Roll measure)** - $Roll_t = 200 \cdot \sqrt{abs(Cov(R_t, R_{t-1}))}$. The Roll (1984) measure is available from January 1964 to December 2010.

**RNana (residual analyst coverage)** - the residual from the cross-sectional regression of the number of analysts covering the firm (from IBES) on the market cap of the firm.

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

$$Sh = \frac{e^y}{1+e^y}, \tag{A-16}$$

where

$$y = -0.46 \cdot log(Size) - 2.8 \cdot Inst + 1.59 \cdot Turn - 0.09 \cdot \frac{CF}{TA} + 0.86 \cdot IPO + 0.41 \cdot Glam. \tag{A-17}$$

In the equation above, *Size* is defined as shares outstanding times the price per share and measured in millions, *Inst* is institutional ownership, *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 one if the stock first appeared on CRSP in the past 12 months, and *Glam* is the dummy variable equal to one for three top market-to-book deciles. All items in the definitions of *CF* are from the Compustat annual file.

**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):

$$\text{Spread} = \frac{2 \cdot (\exp^{\alpha} - 1)}{1 + \exp^{\alpha}}, \quad \text{where} \tag{A-18}$$

$$\alpha = \frac{\sqrt{\beta} \cdot (\sqrt{2} - 1)}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}}, \quad \text{where} \tag{A-19}$$

$$\beta = \log^2\left(\frac{HI_t}{LO_t}\right) + \log^2\left(\frac{HI_{t+1}}{LO_{t+1}}\right) \quad and \quad \gamma = \log^2\left(\frac{\max(HI_t,\ HI_{t+1})}{\min(LO_t,\ LO_{t+1})}\right) \tag{A-20}$$

where $HI_t$ ($LO_t$) is the highest (lowest) price of the stock on day $t$. Spread is available from January 1964 to December 2010.

**SUE flex** is the slope ($\gamma_2$) from the firm-by-firm regression of earnings announcement returns on SUE squared (controlling for the level of SUE):

$$CAR_t = \gamma_0 + \gamma_1 \cdot SUE_t + \gamma_2 \cdot SUE_t^2 \tag{A-21}$$

The regression uses data from quarters $t$-1 to $t$-20 (at least 12 valid observations are required). Earnings announcement days are from Compustat quarterly file. Cumulative abnormal returns (CAR) are computed in the three days before, during, and after announcement using CAPM. The CAPM beta is estimated using daily returns in the year before the announcement. SUE is the difference between the announced EPS (epspiq over prccq lagged by one quarter) and average EPS in the past eight quarters, scaled by the standard deviation of EPS in the past eight quarters. SUE flex is available from January 1981 to December 2010.

**TARCH (expected market volatility)** -from the TARCH(1,1) model (see Glosten, Jagannathan, and Runkle (1993)) fitted to monthly returns on the CRSP value-weighted index:

$$Ret_t^{CRSP} = \gamma_0 + \gamma_1 \cdot Ret_{t-1}^{CRSP} + \epsilon_t, \quad \sigma_t^2 = c_0 + c_1\sigma_{t-1}^2 + c_2\epsilon_{t-1}^2 + c_3 \cdot I(\epsilon_{t-1} < 0). \quad \text{(A-22)}$$

The regression estimated for the full sample. I take the square root out of the volatility forecast to be consistent with my measure of idiosyncratic volatility.

**TVol Sens** is the sensitivity ($\gamma$) of firm's returns to changes in total firm-specific volatility, from firm-specific regressions of the form

$$Ret_t - RF_t = \alpha + \beta \cdot (MKT_t - RF_t) + \gamma \cdot \Delta Vol_t \quad \text{(A-23)}$$

The regression is estimated using monthly data from months $t$-1 to $t$-60 (at least 24 valid observations are required). Volatility ($Vol_t$) is estimated separately each month by computing the standard deviation of daily returns (at least 15 non-missing returns are required). TVol Sens is available from January 1968 to December 2010.

**VIX loading** - see $\boldsymbol{\gamma_{VIX}}$.

# References


[1] Ali, A., and M. A. Trombley. "Short Sales Constraints and Momentum in Stock Returns." *Journal of Business Finance and Accounting*, 33 (2006), 587–615.

[2] Amihud, Y. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (2002), 31–56.

[3] Corwin, S. A., and P. Schultz. "A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices." *Journal of Finance*, 67 (2012), 719–759.

[4] D'Avolio, G. "The Market for Borrowing Stock." *Journal of Financial Economics*, 66 (2002), 271–306.

[5] Glosten, L. R.; R. Jagannathan; and D. E. Runkle. "On the Relation between the Expected Value and the Volatility of the Nominal Return on Stocks." *Journal of Finance*, 48 (1993), 1779–1801.

[6] Hasbrouck, J. "Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data." *Journal of Finance*, 64 (2009), 1445–1477.

[7] Hou, K., and T. J. Moskowitz. "Market Frictions, Price Delay, and the Cross-Section of Expected Returns." *Review of Financial Studies*, 18 (2005), 981–1020.

[8] Kyle, A. S. "Continuous Auctions and Insider Trading." *Econometrica*, 53 (1985), 1315–1336.

[9] Nagel, S. "Short Sales, Institutional Investors, and the Cross-Section of Stock Returns." *Journal of Financial Economics*, 78 (2005), 277–309.

[10] Ohlson, J. A. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." *Journal of Accounting Research*, 18 (1980), 109–131.

[11] Pastor, L., and R. F. Stambaugh. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111 (2003), 642–685.

[12] Roll, R. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *Journal of Finance*, 39 (1984), 1127–1139.

[13] Sadka, R. "Momentum and Post-Earnings-Announcement Drift Anomalies: The Role of Liquidity Risk." *Journal of Financial Economics*, 80 (2006), 309–349.