

Documentation: Mispricing Measure Data

by

Robert F. Stambaugh, Jianfeng Yu, and Yu Yuan

The research assistance of Jianan Liu and Mengke Zhang is gratefully acknowledged.

Our mispricing measure for a stock is constructed by combining its rankings on 11 anomaly variables computed at the end of each month. The data file contains items PERMNO, YYYYMM, and MISP, where MISP is the value of the mispricing measure for stock PERMNO at the end of month YYYYMM. The first month is 196507, the beginning of the sample period used in Stambaugh, Yu, and Yuan (2015).¹ For each anomaly, we assign a rank to each stock that reflects the sorting on that given anomaly variable, where the highest rank is assigned to the value of the anomaly variable associated with the lowest average abnormal return, as reported in the literature. For example, one documented anomaly is that high asset growth in the previous year is followed by a low return (Cooper, Gulen, and Schill (2008)). We therefore rank firms each month by asset growth, and those with the highest growth receive the highest rank. The higher the rank, the greater the relative degree of overpricing according to the given anomaly variable. A stock’s mispricing measure (MISP), ranging between 0 and 100, is the arithmetic average of its ranking percentile for each of the 11 anomalies. According to this measure, the stocks with the highest values of MISP are the most “overpriced,” and those with the lowest values are the most “underpriced.”

Below we detail the construction of the 11 anomaly variables used to construct a stock’s composite mispricing measure. We exclude stocks with share prices less than \$5, primarily to avoid micro-structure effects, and we use ordinary common shares (CRSP codes 10 and 11). When constructing a stock’s mispricing measure at the end of month t , we require that the stock have non-missing values at the end of that month for at least five anomalies. Also, in order for an anomaly to be included in the mispricing measure at the end of month t , we require at least 30 stocks to have non-missing values for that anomaly. For the anomaly variables requiring Compustat data from annual financial statements, we require at least a four-month gap between the end of month t and the end of the fiscal year. When using quarterly reported earnings, we use the most recent data for which the reporting date provided by Compustat (item RDQ) precedes the end of month t . When using quarterly items reported from the balance sheet, we use those reported for the quarter prior to the quarter used for reported earnings. The latter treatment allows for the fact that a significant number of firms do not include balance-sheet information with earnings announcements and only later release it in 10-Q filings (see Chen, DeFond, and Park (2002)). For anomalies requiring return and market capitalization, we use data recorded for month t and earlier, as reported by CRSP. The measure is a modest modification and update to the measure used

¹There are nearly 1000 stocks that month for which the mispricing measure can be computed. Extending the sample to earlier years results in substantially fewer stocks and encounters well known issues of survival effects in Compustat data.

in Stambaugh, Yu, and Yuan (2015).²

The values computed at the end of month t for each anomaly are constructed as follows:

1. *Net Stock Issues*: The stock issuing market has long been viewed as producing an anomaly arising from sentiment-driven mispricing: smart managers issue shares when sentiment-driven traders push prices to overvalued levels. Ritter (1991) and Loughran and Ritter (1995) show that, in post-issue years, equity issuers underperform matching nonissuers with similar characteristics. Motivated by this evidence, Fama and French (2008) show that net stock issues and subsequent returns are negatively correlated. Following Fama and French (2008), we measure net issuance as the annual log change in split-adjusted shares outstanding. Split-adjusted shares equal shares outstanding (Compustat annual item CSHO) times the adjustment factor (Compustat annual item ADJEX_C). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month t . Stocks with negative net issues are assigned to decile 1, and those with zero net issues are assigned to decile 2. The remaining stocks are assigned to the remaining eight deciles using NYSE breakpoints.
2. *Composite Equity Issues*: Daniel and Titman (2006) find that issuers underperform nonissuers using a measure they denote as composite equity issuance, defined as the growth in the firm's total market value of equity minus (i.e., not attributable to) the stock's rate of return. We compute this measure by subtracting the 12-month cumulative stock return from the 12-month growth in equity market capitalization. We lag the quantity four months, to make its timing more coincident with the above measure of net stock issues.
3. *Accruals*: Sloan (1996) shows that firms with high accruals earn abnormally lower average returns than firms with low accruals, and he suggests that investors overestimate the persistence of the accrual component of earnings when forming earnings expectations. Following Sloan (1996), we measure total accruals as the annual change

²The measure is the same one constructed by Stambaugh and Yuan (2016). The difference from the measure used by Stambaugh, Yu and Yuan (2015) is that the current measure constructs the anomaly variables distress, O-score, and investment-to-assets following the same procedure used previously for the other anomalies that also use annual data. That is, for all anomalies using annual data, construction of the current measure requires a gap of at least 4 months after the end of the fiscal year, following precedent in the accounting literature (e.g., Hirshleifer, Hou, Teoh, and Zhang (2004)). In contrast, Stambaugh, Yu, and Yuan (2015), when constructing the distress, O-score, and investment-to-assets variables, require a gap after the end of the fiscal year of at least 6 months and potentially up to 18 months. The latter timing convention is used, for example, by Fama and French (1993) when using annual book value.

in noncash working capital minus depreciation and amortization expense (Compustat annual item DP), divided by average total assets (item AT) for the previous two fiscal years. Noncash working capital is computed as the change in current assets (item ACT) minus the change in cash and short-term investment (item CHE), minus the change in current liabilities (item DLC), plus the change in debt included in current liabilities (item LCT), plus the change in income taxes payable (item TXP). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month t .

4. *Net Operating Assets:* Hirshleifer, Hou, Teoh, and Zhang (2004) find that net operating assets, defined as the difference on the balance sheet between all operating assets and all operating liabilities, scaled by total assets, is a strong negative predictor of long-run stock returns. The authors suggest that investors with limited attention tend to focus on accounting profitability, neglecting information about cash profitability, in which case net operating assets (equivalently measured as the cumulative difference between operating income and free cash flow) captures such a bias. Following equations (4), (5), and (6) of that study, we measure net operation assets as operating assets minus operating liabilities, divided by lagged total assets (Compustat annual item AT). Operating assets equal total assets (item AT) minus cash and short-term investment (item CHE). Operating liabilities equal total assets minus debt included in current liabilities (item DLC), minus long-term debt (item DLTT), minus common equity (item CEQ), minus minority interests (item MIB), minus preferred stocks (item PSTK). (The last two items are zero if missing.) The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month t .
5. *Asset Growth:* Cooper, Gulen, and Schill (2008) find that companies that grow their total assets more earn lower subsequent returns. They suggest that this phenomenon is due to investors' initial overreaction to changes in future business prospects implied by asset expansions. Asset growth is measured as the growth rate of total assets in the previous fiscal year. Following that study, we measure asset growth as the most recent year-over-year annual growth rate of total assets (Compustat annual item AT). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month t .
6. *Investment-to-Assets:* Titman, Wei, and Xie (2004) and Xing (2008) show that higher past investment predicts abnormally lower future returns. Titman, Wei, and Xie (2004) attribute this anomaly to investors' initial underreaction to overinvestment caused by

managers' empire-building behavior. Here, investment to assets is measured as the annual change in gross property, plant, and equipment, plus the annual change in inventories, scaled by lagged book value of assets. Following the above studies, we compute investment-to-assets as the changes in gross property, plant, and equipment (Compustat annual item PPEGT) plus changes in inventory (item INVT), divided by lagged total assets (item AT). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month t .

7. *Distress*: Financial distress is often invoked to explain otherwise anomalous patterns in the cross-section of stock returns. However, Campbell, Hilscher, and Szilagyi (2008) find that firms with high failure probability have lower rather than higher subsequent returns. The authors suggest that their finding is a challenge to standard models of rational asset pricing. Failure probability is estimated with a dynamic logit model that uses several equity market variables, such as stock price, book-to-market, stock volatility, size relative to the S&P 500, and cumulative excess return relative to the S&P 500. Specifically, using the above study's equations (2) and (3) along with its Table IV (12-month column), we compute the distress anomaly measure—failure probability—as

$$\begin{aligned} \pi = & -20.26 NIMTAAVG + 1.42 TLMTA - 7.13 EXRETAVG + 1.41 SIGMA \\ & -0.045 RSIZE - 2.13 CASHMTA + 0.075 MB - 0.058 PRICE - 9.16, \end{aligned}$$

where

$$\begin{aligned} NIMTAAVG_{t,t-11} &= \frac{1 - \phi^3}{1 - \phi^{12}}(NIMTA_{t,t-2} + \dots + \phi^9 NIMTA_{t-9,t-11}) \\ EXRETAVG_{t,t-11} &= \frac{1 - \phi}{1 - \phi^{12}}(EXRET_t + \dots + \phi^{11} EXRET_{t-11}), \end{aligned}$$

and $\phi = 2^{-1/3}$. $NIMTA$ is net income (Compustat quarterly item NIQ) divided by firm scale, where the latter is computed as the sum of total liabilities (item LTQ) and market equity capitalization (data from CRSP). $EXRET_s$ is the stock's monthly log return in month s minus the log return on the S&P500 index. Missing values for $NIMTA$ and $EXRET$ are replaced by those quantities' cross-sectional means. $TLMTA$ equals total liabilities divided by firm scale. $SIGMA$ is the stock's daily standard deviation for the most recent three months, expressed on an annualized basis. At least five non-zero daily returns are required. $RSIZE$ is the log of the ratio of the stock's market capitalization to that of the S&P500 index. $CASHMTA$ equals cash and short-term investment (item CHEQ) divided by firm scale. MB is the market-to-book ratio. Following Campbell, Hilscher, and Szilagyi (2008), we increase book equity by 10% of

the difference between market equity and book equity. If the resulting value of book equity is negative, then book equity is set to \$1. *PRICE* is the log of the share price, truncated above at \$15. All explanatory variables except *PRICE* are winsorized above and below at the 5% level in the cross section. CRSP based variables, *EXRETAVG*, *SIGMA*, *RSIZE* and *PRICE* are for month t . *NIQ* is for the most recent quarter for which the reporting date provided by Compustat (item RDQ) precedes the end of month t , whereas the items requiring information from the balance sheet (*LTQ*, *CHEQ* and *MB*) are for the prior quarter.

8. *O-score*: This distress measure, from Ohlson (1980), predicts returns in a manner similar to the measure above. It is the probability of bankruptcy estimated in a static model using accounting variables. Following Ohlson (1980), we construct it as:

$$\begin{aligned} O &= -0.407 \textit{SIZE} + 6.03 \textit{TLTA} - 1.43 \textit{WCTA} + 0.076 \textit{CLCA} - 1.72 \textit{OENEG} \\ &= -2.37 \textit{NITA} - 1.83 \textit{FUTL} + 0.285 \textit{INTWO} - 0.521 \textit{CHIN} - 1.32, \end{aligned}$$

where *SIZE* is the log of total assets (Compustat annual item AT), *TLTA* is the book value of debt (item DLC plus item DLTT) divided by total assets, *WCTA* is working capital (item ACT minus item LCT) divided by total assets, *CLCA* is current liabilities (item LCT) divided by current assets (item ACT), *ONEEG* is 1 if total liabilities (item LT) exceed total assets and is zero otherwise, *NITA* is net income (item NI) divided by total assets, *FUTL* is funds provided by operations (item PI) divided by total liabilities, *INTWO* is equal to 1 if net income (item NI) is negative for the last 2 years and zero otherwise, *CHIN* is $(NI_j - NI_{j-1}) / (|NI_j| + |NI_{j-1}|)$, in which NI_j is the income (item NI) for year j , which is the most recent reporting year that ends (according to item DATADATE) at least four months before the end of month t .

9. *Momentum*: The momentum effect, discovered by Jegadeesh and Titman (1993), is one of the most robust anomalies in asset pricing. It refers to the phenomenon whereby high (low) past recent returns forecast high (low) future returns. The momentum ranking at the end of month t uses the cumulative returns from month $t-11$ to month $t-1$. This is the choice of ranking variable used by Carhart (1997) to construct the widely used momentum factor.
10. *Gross Profitability Premium*: Novy-Marx (2013) shows that sorting on the ratio of gross profit to assets creates abnormal benchmark-adjusted returns, with more profitable firms having higher returns than less profitable ones. He argues that gross profit is the cleanest accounting measure of true economic profitability. The farther down the

income statement one goes, the more polluted profitability measures become, and the less related they are to true economic profitability. Following that study, we measure gross profitability as total revenue (Compustat annual item REVT) minus the cost of goods sold (item COGS), divided by current total assets (item AT). The most recent reporting year used is the one that ends (according to item DATADATE) at least four months before the end of month t .

11. *Return on Assets:* Fama and French (2006) find that more profitable firms have higher expected returns than less profitable firms. Chen, Novy-Marx, and Zhang (2010) show that firms with higher past return on assets earn abnormally higher subsequent returns. Return on assets is measured as the ratio of quarterly earnings to last quarter's assets. Wang and Yu (2013) find that the anomaly exists primarily among firms with high arbitrage costs and high information uncertainty, suggesting that mispricing is a culprit. Following Chen, Novy-Marx, and Zhang (2010), we compute return on assets as income before extraordinary items (Compustat quarterly item IBQ) divided by the previous quarter's total assets (item ATQ). Income is for the most recent quarter for which the reporting date provided by Compustat (item RDQ) precedes the end of month t , and assets are for the prior quarter.

REFERENCES

- Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In search of distress risk, *Journal of Finance* 63, 2899–2939.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Chen, Long, Robert Novy-Marx, and Lu Zhang, 2010, An alternative three-factor model, Working paper, CKGSB, University of Rochester, and Ohio State University.
- Chen, Shuping, Mark L. DeFond, and Chul W. Park, 2002, Voluntary disclosure of balance sheet information in quarterly earnings announcements, *Journal of Accounting and Economics* 33, 229–251.
- Cooper, Michael J., Huseyin Gulen, and Michael J. Schill, 2008, Asset growth and the cross-section of stock returns, *Journal of Finance* 63, 1609–1652.
- Daniel, Kent D., and Sheridan Titman, 2006, Market reactions to tangible and intangible information, *Journal of Finance* 61, 1605–1643.
- Fama, Eugene F., and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, Eugene F., and Kenneth French, 2006, Profitability, investment, and average returns, *Journal of Financial Economics* 82, 491–518.
- Fama, Eugene F., and Kenneth French, 2008, Dissecting anomalies, *Journal of Finance* 63, 1653–1678.
- Hirshleifer, David, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang, 2004, Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics* 38, 297–331.
- Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for market efficiency, *Journal of Finance* 48, 65–91.
- Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–51.
- Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.
- Ohlson, James A., 1980, Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* 18, 109–131.
- Ritter, Jay R., 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3–27.
- Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71, 289–315.
- Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.
- Stambaugh, Robert F., and Yu Yuan, 2016, Mispricing Factors, Working Paper, University of Pennsylvania and Shanghai Advanced Institute of Finance.

- Titman, Sheridan, K. C. John Wei, and Feixue Xie, 2004, Capital investments and stock returns, *Journal of Financial and Quantitative Analysis* 39, 677–700.
- Wang, Huijun, and Jianfeng Yu, 2013, Dissecting the profitability premium, Working paper, University of Minnesota.
- Xing, Yuhang, 2008, Interpreting the value effect through the Q-theory: An empirical investigation, *Review of Financial Studies* 21, 1767–1795.