Andrew Sun, Keshav Chawla, Gavin Sun, Anson He
April 26, 2023

CFM 301 Final Project Report – Group 10

Executive Summary

Monthly S&P 1500 stock data between 1980-2019 was used for model development, with a training/validation/test split of 50/25/25 respectively – meaning the model was tested on the last 10 years of data, from 2010-2019. Our main factor selection strategy was to pick a diverse set of factors that covers all major predictive categories (within Gu et al.'s 2020 paper) with sufficient data. For factors relating to recent price trends, we selected % deviation from the mean price of the past 20 days. For factors relating to liquidity, we selected the lagged logarithm of market equity (1-month holding period). For factors relating to risk, we selected idiosyncratic volatility per the CAPM (1-month holding period), Quarterly O-score, and Market beta (5-year holding period). For factors relating to valuation and fundamental signals, we selected return on equity (1-month holding period), trailing twelve-month PE ratio, 3-year investment growth, profitability, and quarterly sales growth (1-month holding period). Additionally, we selected hiring rate, which is an intangible factor as per Hou et al. (2020).

The main objective was to select factors that would maximize the alpha of the corresponding long-short portfolio based on predicted returns. The six statistically significant factors we used in our final portfolio after Fama-Macbeth regression filtering are lagged logarithm of market equity (1-month holding period), % deviation from the mean price of the past 20 days, quarterly sales growth (1-month holding period), hiring rate, 3-year investment growth, and Quarterly O-score. The regression methods we used to test our portfolio are simple linear regression, partial least squares regression, and elastic net regression. From testing with these regression methods, we arrived at an annualized alpha of roughly 18.5%, showing that our portfolio outperforms the market.

Initial Factor Selection

From the 50 provided factors, we selected the following 8 factors.

1. **Market Equity, 1 month (lag_log_size)**

The market equity (A.6.1) or market capitalization of a firm is used as a proxy for the size factor. Market equity is defined as the share price multiplied by shares outstanding, and the logarithm of the previous month's market equity is used for the factor lag_log_size. In the paper, *The relationship between return and market value of common stocks*, Banz demonstrated the "size effect" by showing that common stock of small firms had, on average, higher risk-adjusted returns than common stock of large firms over the 1936-1975 period (Banz, 1981, p. 1-2). In particular, going long on small firms and short on large firms in an equal-weighted portfolio yielded an average annualized excess return of 19.8% with a t-statistic of 2.99 (Banz, 1981, p. 14-16). Hou et al.'s (2020, p. 39) replication of the size factor took the opposite approach (long large firms, short small firms) on NYSE firms with equal weighting, resulting in a monthly -0.46% return with a corresponding t-statistic of -1.83 (Hou et al., 2020, p. 34). These results align with the intuition that small firms outperform large firms on average.

2. **Price Deviation, 20 days (deviation_pct20)**

Price deviation is defined as the percentage deviation from the 20-day mean share price. When the price deviation is high (low), this implies that the stock is overbought (oversold) and will undergo correction in the future, leading to lower stock returns. No supporting related research articles could be found on this factor, but the factor was still considered due to high statistical significance from the Fama-Macbeth cross-sectional regression.

3. Return on Equity, 1 month (roe)

Return on equity (A.4.1) is defined as income before extraordinary items divided by 1-quarter-lagged book equity, which is shareholders' equity plus balance sheet deferred taxes plus investment tax credit minus book value of preferred stock (Hou et al. 2020, p. 73-74). In the paper *Digesting Anomalies: An Investment Approach*, Hou et al. (2015) showed that a high decile minus low decile value-weighted portfolio for ROE on NYSE breakpoints yielded a monthly 0.8% return with a corresponding t-statistic of 3.11 (Hou et al., 2015, p. 669). Furthermore, Hou et al. (2020) showed that high minus low return on equity for NYSE firms with equal weighting resulted in a monthly 1.23% return with a corresponding t-statistic of 6.23 (Hou et al., 2020, p. 29). These results align with the intuition that a firm's stock returns are positively correlated with their return on equity and ultimately their profitability.

4. P/E Ratio, trailing 12 months (pe_ttm)

The price-to-earnings ratio is defined as the inverse of the earnings-to-price ratio (A.2.9), which is defined as income before extraordinary items divided by market equity (share price multiplied by shares outstanding) (Hou et al., 2020, p. 26). Hou et al. (2020) showed that a high minus low portfolio of earnings-to-price ratio for NYSE firms with equal weighting resulted in a monthly 0.65% return with a t-statistic of 4.55 (Hou et al., 2020, p. 23). Since price-to-earnings ratio is the inverse, we can infer that low minus high price-to-earnings ratio would yield similar results and that price-to-earnings negatively predicts returns.

5. Idiosyncratic Volatility by CAPM, 1 month (IV_capm)

Idiosyncratic volatility (A.6.4) is defined as the residual volatility from regressing a stock's excess returns on the value-weighted market excess return under the CAPM model (Hou et al., 2020, p. 99). Hou et al. (2020) showed that a high minus low portfolio of 1-month idiosyncratic volatility for NYSE firms with equal weighting resulted in a monthly –0.08% return with a t-statistic of -0.27 (Hou et al., 2020, p. 40). These results align with the intuition that volatility negatively predicts returns. Furthermore, the insignificant t-statistic aligns with our Fama-Macbeth cross-sectional regression as shown later in Table 1.

6. Profitability, Novy-Marx (profitability)

Novy-Marx profitability is defined as the ratio of a firm's gross profits to its assets. The research paper by Novy-Marx showed that a high minus low portfolio of Novy-Marx profitability for NYSE firms with value-weighted average excess returns under Fama-French three-factor model resulted in a monthly 0.52% return with a t-statistic of 4.49 (Novy-Marx, 2012, p. 52). Given that the t-statistic is above 1.96, we conclude that the predicted value is significant, and conclusions can be drawn from it. These results align with our hypothesis that a firm's stock returns are positively correlated with their stock's Novy-Marx profitability.

7. Market Beta, 5 years (beta_5y)

Market beta (A.6.8) is defined as the covariance between the expected returns on a security and the expected returns on the market divided by the variance of the expected returns on the market. Hou et al. (2020) showed that a high minus low portfolio of market beta estimated over a 5-year horizon for NYSE firms with equal weighting resulted in a monthly –0.07% return with a t-statistic of -0.26 (Hou et al., 2020, p. 40). Thus, we expect 5-year market beta to negatively predict returns. Furthermore, the insignificant t-statistic aligns with our Fama-Macbeth cross-sectional regression as shown later in Table 1.

8. Quarterly Sales Growth, 1 month (sales_g_q)

Quarterly sales growth (A.4.34) is defined as quarterly sales divided by its value four quarters ago (Hou et al., 2020, p. 85). Hou et al. (2020) showed that a high minus low portfolio of 1-month quarterly sales growth for NYSE firms with equal weighting resulted in a monthly 0.56% return with a t-statistic of 4.28 (Hou et al., 2020, p. 33). These results align with the intuition that high sales growth positively predicts returns.

Custom Factor Rationale

In addition to our selected factors from the provided dataset, our group created a multitude of additional custom factors and factors outlined in *Replicating Anomalies* (Hou et al., 2020).

1. Hiring Rate

Utilizing the annual Compustat data from 1976 to 2020 (Compustat - Capital IQ, n.d.-a) for the S&P 1500 universe of stocks, we calculated the hiring rate as the percentage increase in the employee datapoint (EMP). Note that due to the EMP data variable only being accessible on a yearly basis this yearly hiring rate datapoint is utilized for the entire year. This factor follows Belo et al. (2014) as they examined the relationship between a company's hiring and their future stock returns. Belo et al. (2014) research in their paper *Labor Hiring, Investment, and Stock Return Predictability in the Cross Section* showcased that the creation of portfolios of equities which exhibited lower hiring rates generated higher alphas than portfolios exhibiting higher hiring rates (Belo et al., 2014, p. 136). Taking an equal weighting long-short portfolio (going long on low hiring rate firms) Belo et al. (2014, p. 136-137) generated an annual alpha of 10.4% corresponding to 5.78 t-statistic. Hou et al.'s (2020, p. 48) replication of this factor by taking the opposite long/short approach (going long on high hiring rate firms) on NYSE firms with equal weighting resulted in a monthly -0.68% return with a corresponding t-statistic of -5.84 (Hou et al., 2020, p. 34). Similarly, our processing of the data on the training set resulted in a Fama-MacBeth cross-sectional regressions t-statistic of -2.48 following the intuition of the aforementioned papers and our understanding on what we would expect the relationship between a firm's hiring rate in a given year and their future expected stock returns.

2. **Investment Growth 3Yr**

As a result of Anderson & Garcia-Feijóo, (2006), our group sought to use the growth rate of a firm's investments to determine whether this is a predictor of future returns. Following Hou et al.'s (2020, p. 70) replication, we utilized the annual Compustat data item for capital expenditure (CAPX) (Compustat - Capital IQ, n.d.-a) to calculate the percentage increase in CAPX relative to the prior CAPX three years ago. The expected per month return of this three-year growth is -0.47% corresponding to a -5.09 t-statistic (NYSE Firms equally weighted) (Hou et al., 2020, p. 27). This follows the results of Anderson & Garcia-Feijóo (2006, p. 184) who found similar results taking a low three-year CAPX growth rate minus high growth resulting in a monthly return of 0.6% with a t-statistic of 4.71 (note that this return value is based on portfolio sorted by the three-year investment growth and size). In agreement with the statistics mentioned above, our t-statistic corresponding to the three-year investment growth is -2.83.

3. Ohlson's Quarterly O-Score

Ohlson's (1980) O-Score is utilized as a profitability measure which serves as a signalling risk to a firm's potential bankruptcy. The data utilized was obtained from Compustat on a Quarterly basis from 1979 – 2019 (Compustat - Capital IQ, n.d.-b). Our group utilized the model one estimates outlined by Ohlson (1980, p. 121) to create the O-Score on a quarterly basis, as was also done so and outlined in (Hou et al., 2020, p. 82) to replicate the factor. This creation of the factor can be interpreted as "model 1 predicts bankruptcy within one year" (Ohlson, 1980, p. 120). Hou et al.'s (2020, p. 32) replication on the equally weighted NYSE portfolio on a monthly basis of the quarterly O-Score resulted in a monthly average return of −0.14% and the corresponding t-statistic of -0.65. Within our Fama-Macbeth regression, this factor has a corresponding t-statistic of -5.66.

Data Cleaning and Final Factor Selection

As part of data cleaning, stocks with a January share price of below \$5 or market cap of below \$100 million were filtered out, and the remaining stock data was imputed using the median cross-sectional value. Fama-Macbeth cross-sectional regressions were used on the training data to identify the statistically significant factors within the 11 selected portfolio factors. Prior to the cross-sectional regression, the 11 selected factors were first winsorized (using 0.01 and 0.99 quantile cutoffs), then standardized through z-score calculation. The initial Fama-Macbeth regression yielded the following t-statistics:

Factor	T-Statistics
lag_log_size	-7.23
deviation_pct20	-9.80
roe	2.01
pe_ttm	2.25
IV_capm	-0.78
profitability	-0.08
beta_5y	0.08
sales_g_q	4.91
HIRING_RATE	-2.05
INVESTMENT_GROWTH_3Y	-3.29
O_SCORE_Q	-6.95

Table 1. Fama Macbeth Regression T-Statistics of Initial 11 Factors

Given Hou, Xue and Zhang's absolute one test t-statistic cutoff of 1.96 (2017), the above results found that *IV_capm, profitability* and *beta_5y* were statistically insignificant in forecasting month t+1 returns. With these 3 factors removed from the explanatory variates, a secondary Fama-Macbeth regression was performed with the remaining 8 factors. The resulting t-statistics are shown below:

Factor	T-Statistics
lag_log_size	-6.52
deviation_pct20	-8.15
roe	1.56
pe_ttm	1.43
sales_g_q	4.39
HIRING_RATE	-2.54
INVESTMENT_GROWTH_3Y	-2.90
O_SCORE_Q	-5.66

Table 2. Fama Macbeth Regression T-Statistics of Filtered 8 Factors

As such, *roe* and *pe_ttm* were also found to be statistically insignificant, leaving 6 final statistically significant factors left: *lag_log_size, O_SCORE_Q, INVESMENT_GROWTH_3Y, deviation_pct20, sales_g_q and HIRING_RATE*. These final factors encompass all four categories of influential stock-price predictors, and thus represent a comprehensive set of predictor variables.

Methodologies Used

In terms of trading models, we decided to focus on linear factor models, as their out-of-sample average excess return was significantly higher than other machine learning methods, such as random forest and neural networks as seen in the Lecture4_ML.ipynb sample code. As such, in addition to the Fama-Macbeth Simple Linear Regression (SLR) model, Elastic Net (EN) and Partial Least Squares (PLS) regression models are used for predicting month t+1 returns.

Annualized Results

Type	Avg Ret	Avg XRet	Std. Dev	Sharpe Ratio	R^2	CAPM Alpha	FF4 Alpha	FF4 IR
SLR	0.2436	0.2025	0.1927	1.0510	0.0023	0.1751	0.1845	0.9573
EN	0.2458	0.2047	0.1960	1.0444	0.0023	0.1774	0.1919	0.9790
PLS	0.2389	0.1979	0.1980	0.9993	0.0024	0.1696	0.1855	0.9366

Table 3. Annualized Portfolio Statistics of Various Linear Models over the Test Period

From the three different regression methods we used, we note that our long-short portfolio has an average annualized return of roughly 24%, which is roughly 20% over the risk-free rate. Our Sharpe ratio is slightly above 1, meaning our portfolio performs slightly better than a risk-free asset after adjusting for risk. Our R^2 value is relatively low, which indicates that our returns do not correlate much with the excess market return. We outperform the market by roughly 18% every year based on our alphas of roughly 17.5% with CAPM and 18.5% with FF4. We have a low standard deviation of roughly 0.2, which combined with our high alpha results in a high information ratio of roughly 1.

When compared with Berkshire Hathaway (Referencing Lecture 2 Slide 12 & 13), Berkshire Hathaway has an annual excess alpha over the market of 8.76%, while we have an annual excess alpha over the market of roughly 18%, meaning we beat Berkshire Hathaway by roughly 9% annually. Berkshire Hathaway has a standard deviation of 0.33, while we have a standard deviation of roughly 0.2, which means our returns vary less than Berkshire Hathaway and are thus more consistent. Berkshire Hathaway has a Sharpe ratio of 0.61, while we have a Sharpe ratio slightly above 1, so after adjusting for risk we outperform Berkshire Hathaway as well. Berkshire Hathaway has an information ratio of 0.46, while we have an information ratio of roughly 1, which is a lot higher and indicates that our returns fluctuate more with changes in standard deviation.

Retrospective and Conclusion

Looking back on our model development process, we identified several areas that we could've done differently to improve the model. Firstly, we could have sourced alternative data to create new factors, such as utilizing press articles processed through NLP or using customer surveys. Secondly, we could have sourced factors relative to the industries to account for industry wide trends or cyclicality in the market. Thirdly, we could have also experimented with a value-weighted portfolio rather than keeping an equally weighted portfolio, or even taking a mean-variance optimization (though this might potentially increase trading costs). As a next step, we would look at potential alternative sources to source data for our custom factors which were dropped due to a limited number of datapoints. Additionally, we would want to explore creating a script to systematically test different linear combinations of portfolio factors on the training/validation data (which is known already) to identify the best set of factors on the training set to test. Finally, we would want to perform cross-sectional regressions and portfolio balancing on a quarterly basis rather than monthly basis to better account for seasonal trends in certain industries.

In conclusion, we believe our portfolio of factors represents a variety of different groups (profitability, fundaments, technical, and intangibles) with the vast majority of them being statistically significant in our final strategy and generating a strong FF4 yearly alpha ~18.5%.

Appendix

A1. Dropped Custom Factors

In addition to the custom factors mentioned above which were utilized in our trading strategy there are a few additional custom factors created which either due to limited data, insignificant t-statistic or was intuitively correlated with another factor were dropped.

4. **Number of M&A Activity**

   As a proxy to measure if active involvement in M&A activity has a significant impact to returns, we aggregated M&A events from the Thomson/Refinitiv *Mergers and Acquisitions Events* dataset from WRDS (Thomson/Refinitiv SDC, n.d.). To create this factor (noOfMAActivity), our methodology counted the number of M&A activity events in each month by an acquiror's CUSIP. Counting activity for both the acquiror and target CUSIPs were not considered due to the potential to double count share buybacks. Due to the limited set of datapoints within the dataset for the NYSE 1500 universe this resulted in numerous months having zero datapoints at all causing issues generating a Fama-MacBeth cross sectional t-value and ultimate resulted in us dropping this factor from consideration in our trading strategy.

5. **Dividend Amount / Bid-Ask Spread**

   Utilizing the CRSP monthly dataset from WRDS (Center for Research in Security Prices, LLC (CRSP), n.d.), we obtained datapoints for the dividend amount paid, DIVAMT. Our thinking within this process is that investors would place a greater importance on the values of dividends and that the raw dollar amount of the dividend paid would have an influence on the stock's future returns. The intuition behind this is not absent of academic research as

Hartzmark & Solomon (2019) analyzed this phenomenon in their paper *The Dividend Disconnect*. Secondly, the spread datapoint represents "the difference between the closing bid and ask quotes for a security" (Center for Research in Security Prices, LLC (CRSP), 2023b). Our group's logic behind the spread factor was to determine if the spread could be utilized as a proxy to measure an equities' liquidity. Ultimately, due to the limited number of datapoints (sometimes nonexistence for some CUSIPs or a set of months) both of these factors were excluded from our trading strategy.

6. Stock Split

Similar to DIVAMT, our group sought to see if a stock split had a significant effect on an individual firm's returns in the following months. Utilizing the CRSP datapoint "Cumulative Factor to Adjust Shares Outstanding" (Center for Research in Security Prices, LLC (CRSP), 2023a) CFACSHR, from WRDS (Center for Research in Security Prices, LLC (CRSP), n.d.), divided by the previous month's CFACSHR to determine the magnitude of the stock split /reverse split. With an indicator value of "1" representing no split/ reverse split occurred. Due to the limited number of stock splits accounted for (values other than "1") this factor was also excluded from our trading strategy.

A2. Submission File Structure

The submitted files include our Python Jupyter Notebooks, Dataset Files, ML Output Files, and README instructions as follows.

Dataset Files:

1. **provided_dataset_merged_df.sas7bdat**

   The provided dataset utilized from learn which covers the 50 factors for the NYSE 1500 universe of firms.

2. **compustat_annual.xlsx**

   Annual Compustat data from WRDS on the NYSE 1500 firms by CUSIP spanning back to 1976 (Compustat - Capital IQ, n.d.-a) including a variety of firm fundamentals (Current Assents, Debt, Sales, etc.) to calculate Ohlson's O-Score on an annual basis, CAPX to calculate the 3 Year Investment Growth on a yearly basis, and Employee data to create the HIRING_RATE factor.

3. **compustat_quarter.xlsx**

   Quarterly Compustat data from WRDS on the NYSE 1500 firms by CUSIP spanning back to 1979 (Compustat - Capital IQ, n.d.-b) including firm fundamentals (like in the annual dataset file) to calculate Ohlson's O-Score on a quarterly basis.

4. **crsp_monthly.xlsx**

   Monthly CRSP data from WRDS on the NYSE 1500 firms by CUSIP (Center for Research in Security Prices, LLC (CRSP), n.d.) including the datapoints for DIVAMT, cumulative amount to adjust shares (to create the STOCK_SPLIT factor), and the SPREAD between the bid and ask prices. Note that none of these three factors were utilized in our dataset and were dropped (see dropped factors section in the appendix).

5. **ff4_monthly.xlsx**

   Fama-French factor data (MKTRF, SMB, HML, RF, MOM) on a monthly basis from 1980 to the end of 2019 (Fama-French, n.d.) to calculate Fama-French 3 and 4 factor alphas and to calculate returns in excess of the risk-free rate.

6. **M&A_events.xlsx**

   List of a variety of M&A events including share buybacks and repurchases from the Thomson/Refinitiv Mergers and Acquisitions Events dataset from WRDS for the NYSE 1500 firms by CUSIP (Thomson/Refinitiv SDC, n.d.). This factor was utilized to create the number of M&A activity events on a monthly basis factor (which was eventually dropped).

7. **provided_dataset_merged_df.csv**

   Merged dataset with all factors after running the CFM301_FactorCreation.ipynb utilized in the preprocessing file.

8. **merged_df.csv**

   Dataset exported from the CFM301_Preprocessing.ipynb file utilized in the ML regression and portfolio building file.

Python Jupyter Notebooks:

This is also the order in which to run the files after downloading the`merged_df.sas7bdat` from Learn (ML Data and Codes) into the `dataset_files` folder and renaming to `provided_dataset_merged_df.sas7bdat`.

1. **CFM301_FactorCreation.ipynb**

   File the loads all of our various datasets and creates all the custom factors and exports them into one aggregated CSV file.

2. **CFM301_Preprocessing.ipynb**

   Preprocess the output from the factor creation file to filter & cleaning the data, and running the Fama-Macbeth cross sectional regression t-statistics. Note that this file is similar to and uses code provided from the ML lectures (mainly lecture 1) in order to build our model.

3. **CFM301_ML_Regression.ipynb**

   Splitting the portfolio based on the testing, validating and testing datasets (utilizing a 50%, 25%, 25% split respectively) and then generating the portfolio builds on the validation set to validate our portfolio. This also performs the hyperparameter turning for the elastic net regularisations and partial least square number of components. Finally, this generates our out-of-sample performance by running the portfolio build on the testing data (and computing the portfolio statistics). Note that this file is similar to and uses code provided from the ML lectures (mainly lecture 2) in order to build our model.

Works Cited

Anderson, C. W., & Garcia-Feijóo, L. (2006). Empirical Evidence on Capital Investment, Growth Options, and Security Returns. *The Journal of Finance, 61*(1), 171–194. http://www.jstor.org/stable/3699338.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*(1), 3–18. https://doi.org/10.1016/0304-405X(81)90018-0

Belo, F., Lin, X., & Bazdresch, S. (2014). Labor Hiring, Investment, and Stock Return Predictability in the Cross Section. *Journal of Political Economy*, *122*(1), 129–177. https://doi.org/10.1086/674549

Center for Research in Security Prices, LLC (CRSP). (2023a, February 23). *Cumulative Factor to Adjust Shares Outstanding Variable Name = CFACSHR*. Wharton Research Data Services. https://wrds-www.wharton.upenn.edu/data-dictionary/form_metadata/crsp_a_stock_msf_timeseriesinformation/spread/

Center for Research in Security Prices, LLC (CRSP). (2023b, February 23). *Spread Between Bid and Ask Variable Name = SPREAD*. Wharton Research Data Services. https://wrds-www.wharton.upenn.edu/data-dictionary/form_metadata/crsp_a_stock_msf_timeseriesinformation/spread/

Center for Research in Security Prices, LLC (CRSP). (n.d.). *Monthly Stock* [Data set]. Wharton Research Data Services. Retrieved April 13, 2023, from https://wrds-www.wharton.upenn.edu/pages/get-data/center-research-security-prices-crsp/annual-update/stock-security-files/monthly-stock-file/

Compustat - Capital IQ. (n.d.-a). *Daily Updates—Fundamentals Annual* [Data set]. Wharton Research Data Services. Retrieved April 24, 2023, from https://wrds-www.wharton.upenn.edu/pages/get-data/compustat-capital-iq-standard-poors/compustat/north-america-daily/fundamentals-annual/

Compustat - Capital IQ. (n.d.-b). *Daily Updates—Fundamentals Quarterly* [Data set]. Wharton Research Data Services. Retrieved April 13, 2023, from https://wrds-www.wharton.upenn.edu/pages/get-data/compustat-capital-iq-standard-poors/compustat/north-america-daily/fundamentals-quarterly/

Fama-French. (n.d.). *3 Factors Plus Momentum—Monthly Frequency* [Data set]. Wharton Research Data Services. Retrieved April 17, 2023, from https://wrds-www.wharton.upenn.edu/pages/get-data/fama-french-portfolios-and-factors/fama-french-portfolios/factors-monthly-frequency/

Gu, S., Kelly, B., Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies, 33*(5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Hartzmark, S. M., & Solomon, D. H. (2019). The Dividend Disconnect. *The Journal of Finance*, *74*(5), 2153–2199. https://doi.org/10.1111/jofi.12785

Hou, K., Xue, C., & Zhang, L. (2020). Replicating Anomalies. *The Review of Financial Studies*, *33*(5), 2019–2133. https://doi.org/10.1093/rfs/hhy131

Hou, K., Xue, C. and Zhang, L. (2015). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies*, *28*(3), 650-705. https://doi.org/10.1093/rfs/hhu068

Novy-Marx, R. (2012). The Other Side of Value: The Gross Profitability Premium. 52. https://rnm.simon.rochester.edu/research/OSoV.pdf

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131. https://doi.org/10.2307/2490395

Thomson/Refinitiv SDC. (n.d.). *Mergers and Acquisitions Events* [Data set]. Wharton Research Data Services. Retrieved April 13, 2023, from https://wrds-www.wharton.upenn.edu/pages/get-data/thomson-reuters/sdc/mergers-and-acquisitions-events/