https://doi.org/10.1016/j.pacfin.2020.101336



# Does Cash-based Operating Profitability Explain the Accruals Anomaly in China? *

**Qingjie Du**
The Hong Kong Polytechnic University
Email: ddvdavid.du@connect.polyu.hk

**Yang Wang**
The Hong Kong Polytechnic University
Email: yang423q.wang@connect.polyu.hk

**K.C. John Wei**
The Hong Kong Polytechnic University
Email: johnwei@ust.hk

March 30, 2021

## Abstract

We investigate the relations between accruals, operating profitability, cash-based operating profitability, and the cross-section of expected stock returns in U.S. and Chinese markets. By replicating the main results in Ball et al. (2016 JFE), we confirm that cash-based operating profitability subsumes the return predictability of accruals and operating profitability in the U.S. market. Extending to the Chinese market, we similarly find that operating profitability and cash-based profitability can both predict returns but in contrast, operating profitability subsumes cash-based profitability.

*JEL Classification*: G12; M41

*Keywords*: Accruals; Operating profitability; Cash-based operating profitability; Return predictability

* We appreciate the helpful guidance and suggestions from Professor Juhani Linnainmaa and the valuable comments and suggestions from Robert Faff (the editor). All remaining errors are ours. Declarations of interest: none.

**Corresponding author**: Qingjie Du, School of Accounting and Finance, Hong Kong Polytechnic University. Tel: (852)-3400-3371. E-mail: ddvdavid.du@connect.polyu.hk.

**Does Cash-based Operating Profitability Explain the Accruals Anomaly in China?**

**Abstract**

We investigate the relations between accruals, operating profitability, cash-based operating profitability, and the cross-section of expected stock returns in U.S. and Chinese markets. By replicating the main results in Ball et al. (2016 JFE), we confirm that cash-based operating profitability subsumes the return predictability of accruals and operating profitability in the U.S. market. Extending to the Chinese market, we similarly find that operating profitability and cash-based profitability can both predict returns but in contrast, operating profitability subsumes cash-based profitability.

*JEL Classification*: G12; M41

*Keywords*: Accruals; Operating profitability; Cash-based operating profitability; Return predictability

## 1. Introduction

Accrual-based accounting is introduced as an improvement over the cash-only accounting because accruals help to align the timeline matching of a firm's revenues and costs in a fiscal period. Meanwhile, accruals also give firms some flexibility to manipulate financial numbers. Previous studies have documented a robust and negative empirical relation between accruals and expected stock returns (e.g., Sloan, 1996), and a positive relation between gross profitability and expected stock returns (e.g., Novy-Marx, 2013). Because accounting accruals are also part of profitability, it seems puzzling to have both accruals and profitability anomalies at the same time.

Since accruals and cash flows are the two sources of profitability, it is very important to understand the underlying mechanisms that drive accruals and profitability anomalies. Sloan (1996) argues that investors ignore the less persistent nature of accruals, and treat accruals as persistent as cash flows, leading to a temporary mispricing (i.e., overpricing) of accruals. Desai, Rajgopal, and Venkatachalam (2004) construct the ratio of total operating cash flows to stock price as a proxy for the value premium, and find this ratio can explain the accruals anomaly. They thus argue that the accruals anomaly comes from the value premium. Cheng and Thomas (2006) find that abnormal accruals still have incremental return predictability after controlling for operating cash flows to price ratio. They hence argue that accruals are not part of the value premium. In a more recent paper, Ball, Gerakos, Linnainmaa, and Nikolaev (BGLN hereafter, 2016) find that the return predictability of accruals (ACC) and operating profitability (OP, including the accruals component) can be explained by cash-based operating profitability (Cash-based OP). The negative (positive) relation between ACC (OP) and expected returns is due to the negative (positive) correlation between ACC (OP) and Cash-based OP.

Given that all the above studies focus on the U.S. market, it is necessary and interesting to conduct replications outside the U.S. market. According to World Bank, China has the second largest stock market.[1] The Chinese market differs from the U.S. market in many ways, such as market development status, political environment, and investor constituent. Chen et al. (2010) replicate 18 stock anomalies originally discovered in the U.S., find that only five exist in China. Liu, Stambaugh, and Yuan (2019) take into account of the impact of shell firms, and construct the Chinese version three-factor model. They find that their three-factor model subsumes the Fama-French three-factor model and can explain most of the documented anomalies in China.

We first replicate the main results in BGLN. We carefully follow the data cleaning process and variable constructions, and use the same sample period as in BGLN. Through the replication, we obtain very similar results as reported in BGLN, and confirm that accruals and operating profitability have no incremental explanatory power of expected stock returns once cash-based operating profitability is controlled for.

We then conduct the replication for the Chinese market for the sample period from July 1999 to December 2018. We find that OP and Cash-based OP have return predictability in the Chinese market but not ACC in Fama and MacBeth (1973) regressions. Moreover, in contrast to BGLN, OP subsumes Cash-based OP in predicting future returns. However, in portfolio analysis, all three profitability-related measures cannot generate significant hedge portfolio returns. We also find the momentum effect, which is strong in the U.S., does not exist in the China. However, we do find a strong value premium and a strong return reversal effect in China.

[1] According to the World Bank statistics, the world top-five markets based on the 2018 equity value of all listed domestic companies are $30.4 trillion for the U.S., $6.3 trillion for China, $5.3 trillion for Japan, $3.8 trillion for Hong Kong, and $2.4 trillion for France. Please see the link: https://data.worldbank.org/indicator/CM.MKT.LCAP.CD?most_recent_value_desc=true.

## 2. Data, sample construction, and summary statistics

We collect data from various sources. For the U.S., we extract stock data for all common stocks with share codes 10 or 11 listed on the NYSE, AMEX, and NASDAQ from the CRSP. Financial firms (SIC 6000 to 6999) are excluded. Annual accounting data come from Compustat/NA. We require at least six-month gap between a firm's fiscal year-end and portfolio formation date to ensure accounting data are publicly available. Firms must have non-missing market value of equity, book-to-market equity, gross profitability, total assets, current month returns, and returns in the previous one year. All variables are winsorized at the 1% and 99% levels. To mitigate the impact of small-cap stocks, we identify microcap stocks if the market cap falls below the 20% breakpoint of all NYSE stocks. The sample period covers from July 1963 to December 2014, exactly the same as in BGLN.

For the Chinese data, we use the CSMAR database. Only A-share stocks (ticker starting by 60, 30, or 00) from Shanghai Stock Exchange, Shenzhen Stock Exchange, and the board of GEM are included. Following Liu, Stambaugh, and Yuan (2019), we exclude stocks that become public within the past six months. We also require stocks to have at least 120 trading-day data in the past 12 months and at least 10 trading-day data in the previous month.[2] Following Liu et al. (2019), we define shell stocks as those whose market cap falls in the bottom 30%. Liu et al. (2019) argue that shell stocks do not have any fundamental value, and the main market value comes from the reverse mergers of large private firms. The sample period covers from July 1999 to December 2018. All variables are winsorized at the 1% and 99% levels.

Table 1 presents the summary statistics. Panel A shows the U.S. sample distribution. The mean and standard deviation of OP from our replication are 0.128 and 0.160, which differ from

[2] Liu, Stambaugh, and Yuan (2019) require at least 15 trading days. We find some months only have 12 or 13 trading days when they contain the Lunar New Year holiday. We thus relax this filter to 10 trading days.

the original values by only 0.001. For ACC and book-to-market equity (Ln(B/M)), the differences are slightly larger, but the magnitudes are still very close. For correlation matrix in Panel C, only the correlation between OP and Cash-OP shows a little larger difference. Overall, our replication results show very similar distribution properties to the original results in BGLN.[3]

Panel B of Table 1 reports the summary statistics for the Chinese sample. On average, Chinese firms are less profitable than U.S. firms. The mean (median) OP is only 5.4% (4.9%), about 50% smaller compared with U.S. firms. Book-to-market equity (B/M) is also much smaller with the mean (median) of 0.391 (0.350). Chen et al. (2018) report the average B/M of 0.55 during 1993 to 2017, and Fang (2019) reports the median B/M of 0.767 for 1999 to 2017. Our B/M is smaller, because we follow Liu et al. (2019) to use all shares outstanding rather than the tradable shares to calculate market equity. The distribution of ACC is also similar to Chen et al. (2010).

## 3. Empirical results

### 3.1. Fama-MacBeth regressions using the U.S. sample

Table 2 presents our baseline results from Fama-MacBeth (1973) regressions. Panel A report the results based on all-but-microcap stocks and Panel B on microcap stocks only. Columns (2) to (7) present the original results in BGLN.[4] Columns (2a) to (7a) present our replication results. All the coefficients show similar magnitudes and significance levels as the original results. Specifically, the coefficient on OP is positive and significant (coeff = 2.83; *t*-stat = 6.10) in Column (2a), and the coefficient on ACC is negative and significant (coeff = -1.05; *t*-stat = -3.55) in Column (3a). Both ACC and OP have similar signed coefficients when they are

[3] One possible reason for the difference may be due to the data updating in Compustat. Every time when the database is updated, they back-fill and overwrite some data. We check with the original authors and they cannot get 100% of the same results. However, all the interpretations are the same.

[4] Column (1) in the original results uses a different testing sample compared with columns (2) to (7). To save space, we omit to report the results. The results are similar and available upon request.

simultaneously included in the regression (Column (4a)). These replication results show similar magnitudes as in BGLN, and confirm the existence of accruals and operating profitability anomalies in the U.S. market.

Columns (5a) to (7a) further control for cash-based operating profitability (Cash-OP). Cash-OP can positively predict expected stock returns, but ACC and OP lose return predictability once controlling for Cash-OP. Combining the correlation matrix between ACC, OP, and Cash-OP in Panel C of Table 1, we confirm that the negative relation between ACC and expected returns comes from the negative correlation between ACCs and Cash-based OP; and the positive relation between OP and expected returns comes from the positive correlation between OP and Cash-based OP. Besides, all other control variables also have similar coefficients and *t*-statistics as in BGLN. Overall, our replication obtains very close results to those in BGLN.

### 3.2. Fama-MacBeth regressions using the Chinese sample

Table 3 reports our replication results of Fama-MacBeth regressions using the Chinese sample. To mitigate the impact of small stocks, we exclude shell stocks.[5] We find that the coefficient on ACC is insignificant. In contrast, the coefficients on OP (Column (2b)) and Cash-OP (Column (5b)) are both significantly positive, indicating that both have return predictive power in China. However, once both are included in the regression as shown in Column (7b), Cash-OP loses its return predictability, suggesting that OP can subsume the Cash-OP effect in China. This return pattern is substantially different from that reported in Table 2 (Column (7b) in the U.S., where Cash-OP subsumes OP. For control variables, momentum (MOM) is

[5] We also use all stocks to conduct the Fama-MacBeth regressions. The results are similar to those reported in Table 3 and are available upon request.

insignificant. Interesting, similar to the U.S., we find a strong value premium and a strong return reversal effect in China.

### 3.3. Results from portfolio analysis for the U.S. and Chinese markets

To double check the robustness of our results using Fama-MacBeth regressions, we also conduct portfolio analysis. Each month, we rank all stocks into decile portfolios based on a particular firm characteristic. We then construct a zero-cost hedge portfolio that longs the top decile and shorts the bottom decile. The time-series averages of hedge portfolio returns are reported in Table A1 of Appendix B.[6]

Portfolio analysis results are essentially consistent with Fama-MacBeth regression results for the U.S. market except for the small size effect, which is now insignificant. All three profitability-related variables, ACC, Cash-OP, and MOM, continue to predict future returns. In contrast, for the Chinese market, it is a different story. Portfolio analysis shows that none of the ACC, OP, and Cash-OP can predict future returns. However, consistent with Fama-MacBeth regressions, we find a significant value premium, a strong return reversal effect, a marginally significant small size effect, and no momentum effect. Our results are also consistent with previous studies on Chinese market anomalies. Chui et al. (2010) report an insignificant momentum effect in China. Chen et al. (2018) report insignificant operating profitability effects. Chen et al. (2010) and Fang (2019) comprehensively examine Chinese stock anomalies originally discovered in the U.S., and find that many anomalies do not exist in China, including the momentum effect, accruals anomaly, and profitability anomalies. Specifically, similar to our findings, Fang (2019) also documents an insignificant hedge portfolio return of OP using portfolio analysis, but finds a significant Fama-MacBeth regression coefficient.

[6] To save space, we only report the excess hedge returns and omit the alphas using different risk models. The results are similar and available upon request.

### 3.4. Discussions of the replication results

We obtain very similar results as in BGLN using U.S. data, but somewhat different results when we extend to the Chinese market. One caveat is that we use a different sample period for the Chinese replication. The Chinese data are only available since 1993, thus we cannot use the same period as the U.S. sample. Following Liu et al. (2019), we start our sample period from July 1999 because the number of stocks becomes more than 1,000 around 2000 in the Chinese market.

To have a fair comparison, we also replicate the U.S. results based on Fama-MacBeth regressions using the same Chinese sample period. The results are reported in Table A2 of Appendix B. The return patterns are similar to those using the BGLN sample period, except the slope coefficients on ACC become insignificant, which is consistent with the decaying trend of the accruals anomaly in recent years (Green et al., 2011). Importantly, Cash-OP still subsumes the pricing effect of OP (Column (7b)). Therefore, the main result in BGLN is robust to different sample periods.

Based on the same sample period as shown in Table 3 and Table A2, we find some similarity and some distinction between U.S. and Chinese markets. The similarity is that both OP and Cash-OP have return predictability but not ACC in both markets, although the Cash-OP effect is much weaker in China.[7] The distinction is that Cash-OP subsumes OP in the U.S., while it is reverse in China. We conjecture that differences in the stage of market development, investor constituents, culture, or other latent reasons might be potential drivers.

Generally, retail investors are less rational, less patient, and less professional compared to institutions. The majority of investors in the Chinese market are retail investors, who may lack

[7] A recent study by Carpenter et al. (2020) finds that stock prices in China have become as informative as they are in the U.S. about their future profitability.

professional knowledge to apply accounting information in firms' financial statements. They are more likely to be noisy traders, which makes stock prices less informative. The weaker return predictability in China based on portfolio sorts does not imply that Chinese market is more efficient or rational.[8] Instead, it may be a symptom of persistent noisy valuation and persistent mispricing (Chen et al., 2010). Besides, Chui et al. (2010) investigate the individualism and the momentum effect around the world, and conclude the individualism culture is a strong driver of the momentum effect. Chinese investors share a collective culture, while U.S. investors are more individualism, and this culture difference may explain the momentum effect in China.

## 4. Conclusion

Accounting accruals adjust the current period earnings by matching the timeline of revenues and costs, making accounting numbers better reflect a firm's operation. While the negative relation between accruals and expected returns seems puzzling and attracts great academic attention. Ball et al. (2016) propose a cash-based operating profitability measure, and find that this measure can subsume both accruals and profitability anomalies. We can successfully replicate their main results. We then extend their tests to the Chinese market. The results from Fama-MacBeth regressions using the same recent sample period show that both operating profitability and cash-based operating profitability can predict future returns but not accruals in both market. However, cash-based operating profitability subsumes operating profitability in the US, while the reverse is true in China.

[8] Fama and French (2017, p. 442), state, "Finally, the dividend discount model that motivates Eq. (1) (*i.e., the Fama and French (2015) five-factor model*) is useful for suggesting variables related to differences in expected asset returns, but it is silent on economic or behavioral explanations of the differences." The italics are added.

## References


Ball, Ray, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri Nikolaev, 2016, Accruals, cash flows, and operating profitability in the cross section of stock returns, *Journal of Financial Economics* 121, 28-45.

Carpenter, Jennifer N., Fangzhou Lu, and Robert F. Whitelaw, 2020, The real value of China's stock market, *Journal of Financial Economics* forthcoming.

Chen, Te-Feng, Lei Sun, K. C. John Wei, and Feixue Xie, 2018, The profitability effect: Insights from international equity markets, *European Financial Management* 24, 545-580.

Chen, Xuanjuan, Kenneth A. Kim, Tong Yao, and Tong Yu, 2010, On the predictability of Chinese stock returns, *Pacific-Basin Finance Journal* 18, 403-425.

Cheng, C. S. Agnes, and Wayne B. Thomas, 2006, Evidence of the abnormal accrual anomaly incremental to operating cash flows, *The Accounting Review* 81, 1151-1167.

Chui, Andy C.W., Sheridan Titman, and K. C. John Wei, 2010, Individualism and momentum around the world, *Journal of Finance* 65, 361–392.

Desai, Hemang, Shivaram Rajgopal, and Mohan Venkatachalam, 2004, Value-glamour and accruals mispricing: One anomaly or two? *The Accounting Review* 79, 355-385.

Fang, Miao, 2019, Replicating Anomalies in China, Tsinghua University Working Paper.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2017, International tests of a five-factor asset pricing model, *Journal of Financial Economics* 123, 441-463.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Green, Jeremiah, John RM Hand, and Mark T. Soliman, 2011, Going, going, gone? The apparent demise of the accruals anomaly, *Management Science* 57, 797-816.

Liu, Jianan, Robert F. Stambaugh, and Yu Yuan, 2019, Size and value in China, *Journal of Financial Economics* 134, 48-69.

Novy-Marx, Robert, 2013, The other side of value: The gross profitability premium, *Journal of Financial Economics* 108, 1–28.

Sloan, Richard G., 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71, 289–315.

**Table 1. Sample Descriptive Statistics for U.S. Market and Chinese Market**

This table shows the means, standard deviations, and percentiles distribution of the sample used in our analysis. The descriptive statistics are the time-series averages of the corresponding statistics in each month. Panels A, B, and C show the U.S. sample distribution, Chinese sample distribution, and the correlation matrix, respectively. OP is operating profitability, defined as gross profit minus SG&A (excluding R&D expense) deflated by total assets lagged by one year. ACC is accruals, calculated based on Balance Sheet items, defined in the same way as in Sloan (1996). Cash-OP is cash-based operating profitability. Ln(B/M) is the logarithm of book-to-market equity. Ln(ME) is the logarithm of market equity. STR is short term reversal, defined as the past one-month return. MOM is stock momentum, defined as the past one-year cumulative return, skipping the last month. To be included in the sample, firms are required to have non-missing values of the following items: market value of equity, book-to-market equity, gross profit, total assets, current month return, and cumulative return in the past one-year period. Financial firms are excluded. Detailed constructions of variables are described in Appendix A. The sample period for the U.S. sample is from July 1963 to December 2014, and from July 1999 to December 2018 for the Chinese sample. To mitigate the impact of small stocks, the U.S. sample contains only all-but-microcaps stocks, and only all-but-shell stocks for the Chinese sample.

Panel A: Summary statistics for the U.S. market (July 1963 to December 2014)

| | Original results (Panel A of Table 1 in Ball et al., 2016) | | | | | | | Replicated results | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | 1st | 25th | 50th | 75th | 99th | Mean | SD | 1st | 25th | 50th | 75th | 99th |
| OP | 0.129 | 0.159 | -0.389 | 0.077 | 0.138 | 0.203 | 0.468 | 0.128 | 0.160 | -0.394 | 0.076 | 0.137 | 0.202 | 0.466 |
| ACC | -0.029 | 0.114 | -0.342 | -0.069 | -0.028 | 0.016 | 0.235 | -0.021 | 0.164 | -0.474 | -0.073 | -0.025 | 0.030 | 0.439 |
| Cash-OP | 0.117 | 0.175 | -0.433 | 0.056 | 0.126 | 0.195 | 0.491 | 0.110 | 0.184 | -0.485 | 0.049 | 0.123 | 0.193 | 0.492 |
| Ln(B/M) | -0.534 | 0.940 | -3.204 | -1.051 | -0.450 | 0.059 | 1.537 | -0.471 | 0.863 | -2.905 | -0.944 | -0.390 | 0.083 | 1.350 |
| Ln(ME) | 4.577 | 1.958 | 0.685 | 3.161 | 4.448 | 5.883 | 9.407 | 4.609 | 1.974 | 0.683 | 3.168 | 4.488 | 5.947 | 9.442 |
| STR | 0.013 | 0.152 | -0.305 | -0.063 | 0.001 | 0.071 | 0.488 | 0.013 | 0.151 | -0.303 | -0.062 | 0.002 | 0.070 | 0.485 |
| MOM | 0.163 | 1.313 | -0.670 | -0.172 | 0.057 | 0.327 | 2.200 | 0.149 | 0.588 | -0.666 | -0.169 | 0.058 | 0.328 | 2.197 |

Panel B: Summary statistics for the Chinese market (July 1999 to December 2018)

| | Replicated China Sample Summary Statistics | | | | | | |
|---|---|---|---|---|---|---|---|
| | Mean | SD | 1st | 25th | 50th | 75th | 99th |
| OP | 0.054 | 0.064 | -0.128 | 0.023 | 0.049 | 0.082 | 0.231 |
| ACC | 0.002 | 0.179 | -0.285 | -0.059 | -0.007 | 0.051 | 0.347 |
| Cash-OP | 0.022 | 0.171 | -0.326 | -0.029 | 0.028 | 0.084 | 0.310 |
| Ln(B/M) | -1.149 | 0.578 | -2.676 | -1.493 | -1.111 | -0.760 | 0.023 |
| Ln(ME) | 1.686 | 0.743 | 0.820 | 1.130 | 1.494 | 2.043 | 4.072 |
| STR | 0.017 | 0.100 | -0.173 | -0.045 | 0.003 | 0.064 | 0.334 |
| MOM | 0.235 | 0.426 | -0.378 | -0.040 | 0.143 | 0.406 | 1.712 |

Panel C: Correlation matrix

| | Correlation matrix (Original results) | | | Correlation Matrix (Replicated results) | | | Correlation matrix (Replicated Chinese results) | | |
|---|---|---|---|---|---|---|---|---|---|
| | OP | ACC | Cash-OP | OP | ACC | Cash-OP | OP | ACC | Cash-OP |
| OP | | 0.130 | 0.805 | | 0.117 | 0.752 | | 0.167 | 0.398 |
| ACC | 0.163 | | -0.280 | 0.151 | | -0.247 | 0.147 | | -0.588 |
| Cash-OP | 0.845 | -0.252 | | 0.750 | -0.230 | | 0.443 | -0.505 | |

**Table 2. Replication results for the U.S. sample: Fama-MacBeth regressions**

This table reports our replication results of Table 2 in Ball et al. (2016) using the same sample period. The table shows the time-series averages of Fama-MacBeth regression coefficients and the corresponding *t*-statistics as shown in parentheses. Panels A and B report the results using all-but-microcap and microcap stocks, respectively. To save space, we omit the first column in the original table since the first column uses a different sample as other columns. All explanatory variables are winsorized at the 1% and 99% levels. The left half of each panel shows the original results, and the right half shows the replicated results.

Panel A: All-but-microcaps

| | Original results (Panel A of Table 2 in Ball et al., 2016) | | | | | | Replicated results | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (2) | (3) | (4) | (5) | (6) | (7) | (2a) | (3a) | (4a) | (5a) | (6a) | (7a) |
| OP | 2.55 | | 2.55 | | | 0.80 | 2.83 | | 2.83 | | | 0.83 |
| | (7.04) | | (7.09) | | | (1.56) | (6.10) | | (6.21) | | | (1.56) |
| ACC | | -1.41 | -1.58 | | 0.15 | | | -1.05 | -1.02 | | -0.05 | |
| | | (-3.90) | (-4.45) | | (0.34) | | | (-3.55) | (-3.53) | | (-0.18) | |
| Cash-OP | | | | 2.60 | 2.54 | 1.91 | | | | 2.80 | 2.78 | 2.20 |
| | | | | (9.69) | (7.40) | (5.27) | | | | (7.03) | (6.79) | (5.89) |
| Ln(B/M) | 0.36 | 0.21 | 0.33 | 0.33 | 0.32 | 0.33 | 0.39 | 0.20 | 0.35 | 0.34 | 0.34 | 0.36 |
| | (5.80) | (3.28) | (4.66) | (4.76) | (4.53) | (4.73) | (4.16) | (2.52) | (3.97) | (3.81) | (3.79) | (3.96) |
| Ln(ME) | -0.09 | -0.09 | -0.10 | -0.10 | -0.10 | -0.10 | -0.09 | -0.08 | -0.09 | -0.10 | -0.10 | -0.10 |
| | (-2.35) | (-2.24) | (-2.66) | (-2.59) | (-2.69) | (-2.59) | (-2.13) | (-1.93) | (-2.39) | (-2.45) | (-2.50) | (-2.46) |
| STR | -3.23 | -3.34 | -3.30 | -3.27 | -3.32 | -3.28 | -3.83 | -3.75 | -3.88 | -3.84 | -3.88 | -3.88 |
| | (-7.49) | (-7.76) | (-7.72) | (-7.58) | (-7.78) | (-7.66) | (-7.61) | (-7.52) | (-7.68) | (-7.64) | (-7.69) | (-7.68) |
| MOM | 0.95 | 0.86 | 0.91 | 0.92 | 0.91 | 0.92 | 0.72 | 0.77 | 0.72 | 0.71 | 0.71 | 0.71 |
| | (5.30) | (4.81) | (5.12) | (5.11) | (5.08) | (5.15) | (2.86) | (3.06) | (2.87) | (2.80) | (2.83) | (2.81) |
| Adj. $R^2$ | 5.5% | 5.2% | 5.7% | 5.4% | 5.6% | 5.6% | 5.7% | 5.5% | 5.9% | 5.7% | 5.8% | 5.9% |

Panel B: Microcaps

| | Original results (Panel B of Table 2 in Ball et al., 2016) | | | | | | Replicated results | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (2) | (3) | (4) | (5) | (6) | (7) | (2a) | (3a) | (4a) | (5a) | (6a) | (7a) |
| OP | 2.09 | | 2.30 | | | 0.20 | 2.28 | | 2.55 | | | 0.35 |
| | (5.29) | | (5.85) | | | (0.40) | (5.23) | | (6.01) | | | (0.60) |
| ACC | | -1.97 | -2.49 | | -0.75 | | | -1.38 | -1.70 | | -0.61 | |
| | | (-6.30) | (-8.26) | | (-1.74) | | | (-4.67) | (-6.12) | | (-1.90) | |
| Cash-OP | | | | 2.48 | 2.27 | 2.21 | | | | 2.54 | 2.42 | 2.29 |
| | | | | (9.62) | (6.67) | (7.27) | | | | (9.82) | (8.49) | (7.14) |
| Ln(B/M) | 0.41 | 0.41 | 0.39 | 0.40 | 0.39 | 0.39 | 0.39 | 0.37 | 0.37 | 0.36 | 0.36 | 0.36 |
| | (6.25) | (6.33) | (5.89) | (6.33) | (6.06) | (5.87) | (4.39) | (4.26) | (4.17) | (4.21) | (4.17) | (4.07) |
| Ln(ME) | -0.26 | -0.19 | -0.25 | -0.25 | -0.24 | -0.25 | -0.25 | -0.18 | -0.25 | -0.25 | -0.24 | -0.25 |
| | (-4.17) | (-2.92) | (-4.00) | (-3.92) | (-3.91) | (-4.09) | (-4.52) | (-3.01) | (-4.38) | (-4.17) | (-4.16) | (-4.50) |
| STR | -6.12 | -6.02 | -6.17 | -6.11 | -6.16 | -6.18 | -5.93 | -5.84 | -5.96 | -5.92 | -5.94 | -5.97 |
| | (-13.68) | (-13.37) | (-13.83) | (-13.66) | (-13.82) | (-13.84) | (-8.79) | (-8.70) | (-8.83) | (-8.78) | (-8.83) | (-8.83) |
| MOM | 0.90 | 0.91 | 0.85 | 0.88 | 0.85 | 0.86 | 0.78 | 0.80 | 0.76 | 0.76 | 0.75 | 0.76 |
| | (5.02) | (4.98) | (4.74) | (4.89) | (4.75) | (4.79) | (3.81) | (3.90) | (3.69) | (3.72) | (3.68) | (3.70) |
| Adj. $R^2$ | 3.1% | 2.9% | 3.2% | 3.0% | 3.2% | 3.2% | 3.1% | 2.9% | 3.2% | 3.0% | 3.1% | 3.1% |

**Table 3. . Replication results for the Chinese sample: Fama-MacBeth regressions**

This table reports our replication results using all-but-shell stocks for the Chinese market sample. The table shows the time-series averages of Fama-MacBeth regression coefficients and the corresponding *t*-statistics as shown in parentheses. All explanatory variables are trimmed at the 1% and 99% level. The sample period is Jul 1999 to Dec 2018.

| | Replication results for the Chinese market | | | | | |
|---|---|---|---|---|---|---|
| | (2b) | (3b) | (4b) | (5b) | (6b) | (7b) |
| OP | 3.14 | | 3.26 | | | 3.00 |
| | (2.27) | | (2.32) | | | (2.22) |
| ACC | | 0.01 | -0.27 | | 0.60 | |
| | | (0.03) | (-0.80) | | (1.34) | |
| Cash-OP | | | | 0.80 | 1.15 | 0.16 |
| | | | | (1.78) | (2.05) | (0.46) |
| Ln(B/M) | 0.40 | 0.36 | 0.40 | 0.36 | 0.37 | 0.40 |
| | (2.57) | (2.22) | (2.58) | (2.23) | (2.38) | (2.53) |
| Ln(ME) | -0.28 | -0.23 | -0.28 | -0.24 | -0.25 | -0.27 |
| | (-1.56) | (-1.21) | (-1.57) | (-1.28) | (-1.32) | (-1.53) |
| STR | -5.00 | -4.88 | -5.00 | -4.87 | -4.91 | -5.01 |
| | (-6.08) | (-5.79) | (-6.07) | (-5.73) | (-5.88) | (-6.10) |
| MOM | 0.41 | 0.45 | 0.40 | 0.43 | 0.44 | 0.40 |
| | (0.89) | (0.97) | (0.89) | (0.93) | (0.96) | (0.87) |
| Adj. $R^2$ | 8.1% | 7.6% | 8.3% | 7.6% | 7.8% | 8.3% |

**Appendix A: Variable Definition**

*OP*: operating profitability, defined as:

$$OP = (REVT - COGS - XSGA + XRD)/AT,$$

where REVT is total revenue, COGS is cost of goods sold, XSGA is reported sales, general, and administrative expenses, and XRD is total expenditures on research and development (if any). The item is scaled by total assets (AT) in that year. Portfolios sorted by $OP$ are updated annually at the end of June. $OP$ for Chinese firms is constructed as operating income (B001300000) plus the R&D expenses (A001219000), scaled by total assets (A001000000).

*ACC*: accounting accruals, defined as in Sloan (1996):

$$ACC = ((\Delta ACT - \Delta CH) - (\Delta LCT - \Delta DLC - \Delta TXP) - \Delta DP)/AT,$$

where ACT is total current assets, CH is cash and cash-equivalent, LCT is total current liabilities, DLC is debt in current liabilities, TXP is income taxes payables (if any), DP is depreciation and amortization (if any), and AT is total assets. For each variable X, ΔX is the change from the previous fiscal year. If the value is missing, it is set to zero. Portfolios sorted by $ACC$ is rebalanced every year at the end of June. $ACC$ for the Chinese firms is constructed in the same way, using the following items: total current assets (A001100000), cash and equivalent (A001101000), total current liabilities (A002100000), debt in current liabilities (A002101000 and A002125000), income tax payables (A002113000), depreciation and amortization (D000103000, D000104000, and D000105000), and total assets (A001000000).

*Cash-OP*: the cash-based operating profitability, defined as:

$$Cash\_OP = (OP - \Delta RECT - \Delta INVT - \Delta XPP + (\Delta DRC + \Delta DRLT) + \Delta AP + \Delta XACC)/AT,$$

where OP is operating profitability, RECT is accounts receivables, INVT is inventory, XPP is prepaid expenses, DRC + DRLT is total deferred revenue, AP is accounts payables, XACC is total accrued expenses, and AT is total assets. For each variable X, ΔX is the change from the previous fiscal year. If the variable is missing, it is set to zero. Portfolios sorted by $Cash\_OP$ are rebalanced every year at the end of June. $Cash\_OP$ for Chinese firms is constructed in the same way, using the following items: account receivables (A001111000), inventory (A001123000), prepaid expenses (A001112000), total deferred revenue (A002127000 and A002210000), account payables (A002108000), other expenses (A002120000), and total assets (A001000000).

*Ln*(*B/M*): book-to-market equity, defined as the logarithm of a firm's book equity (BE) to market equity (ME) ratio. The market equity (ME) equals SHROUT×PRC, where SHROUT is the number of shares outstanding and PRC is the stock price at the end of calendar year *t*-1. Book equity (BE) is estimated as SEQ, adjusted by preferred stock and deferred tax. If BE is missing, it is supplemented by AT – LT, where AT is total assets and LT is total liabilities. Portfolios based on Ln(B/M) are rebalanced annually at the end of June. BE for Chinese firms is constructed as book value of equity (A003000000) minus the minority interests (A003200000). ME is constructed using total market value (Msmvttl).

*Ln*(*ME*): firm size, defined as the logarithm of a firm's market capitalization at the end of the previous month, calculated as the number of shares outstanding multiplied by share price. This variable is updated monthly.

*STR*: short-term reversal, defined as the return in the last month:

$$STR_t = R_{t-1},$$

where $R_{t-1}$ is the stock return in month t-1. The STR characteristic is updated monthly.

*MOM*: Momentum, defined as the cumulative return over the previous 12 months excluding the last month:

$$MOM_t = \prod_{s=t-12}^{t-2} (1 + R_s) - 1,$$

where $R_s$ is the stock return in month s. The portfolio based on MOM is rebalanced monthly at the end of every month.

**Appendix B: Additional Results**

**Table A1. Excess returns on portfolios sorted by each firm characteristic**

This table reports the time-series averages and *t*-statistics of equal-weighted excess returns in percentage on portfolios sorted by a firm characteristic. The hedge returns are defined as the difference in monthly returns between the top and bottom decile portfolios (H – L). Panels A and B report the results for the U.S. and Chinese markets, respectively. The sample period for the U.S. sample is from July 1963 to December 2014, and from July 1999 to December 2018 for the Chinese sample. To mitigate the impact of small stocks, the U.S. sample contains only all-but-microcaps stocks, and only all-but-shell stocks for the Chinese sample. The *t*-statistics are shown in parentheses.

Panel A: the U.S. market

| | OP | ACC | Cash-OP | Ln(B/M) | Ln(ME) | STR | MOM |
|---|---|---|---|---|---|---|---|
| Bottom 10% | 0.71 | 1.20 | 0.56 | 0.73 | 1.22 | 1.48 | 0.56 |
| | (2.33) | (4.89) | (1.80) | (2.50) | (4.32) | (5.14) | (1.72) |
| Top 10% | 1.24 | 0.73 | 1.36 | 1.38 | 0.90 | 0.57 | 1.64 |
| | (5.33) | (2.50) | (6.21) | (6.68) | (5.11) | (2.24) | (5.76) |
| H – L | 0.53 | -0.47 | 0.80 | 0.66 | -0.32 | -0.91 | 1.08 |
| | (2.78) | (-4.15) | (4.52) | (2.96) | (-1.54) | (-4.71) | (3.66) |

Panel B: the Chinese Market

| | OP | ACC | Cash-OP | Ln(B/M) | Ln(ME) | STR | MOM |
|---|---|---|---|---|---|---|---|
| Bottom 10% | 0.77 | 0.82 | 0.76 | 0.55 | 1.34 | 1.44 | 0.77 |
| | (0.89) | (0.99) | (0.90) | (0.72) | (1.55) | (1.79) | (0.91) |
| Top 10% | 0.98 | 0.85 | 0.93 | 1.24 | 0.66 | 0.05 | 1.09 |
| | (1.36) | (1.03) | (1.22) | (1.48) | (0.85) | (0.06) | (1.47) |
| H – L | 0.21 | 0.02 | 0.17 | 0.69 | -0.69 | -1.39 | 0.32 |
| | (0.79) | (0.17) | (1.02) | (2.05) | (-1.84) | (-4.14) | (0.86) |

**Table A2. Replication results for the U.S. sample using the sample period of July 1999 to December 2018**

This table reports the U.S. replication results using the same sample period as the Chinese sample. Panel A shows the means, standard deviations, and percentiles distribution of the sample. Panel B reports the Fama-MacBeth regression coefficients and the corresponding *t*-statistics using all-but-microcaps stocks. OP is the gross operating profitability, defined as gross profit minus SG&A (excluding R&D expenses), deflated by total assets lagged by one year. ACC is accruals calculated from Balance Sheet items, defined in the same way as in Sloan (1996). Cash-OP is cash-based operating profitability. Ln(B/M) is the logarithm of book-to-market equity. Ln(ME) is the logarithm of market equity. STR is short term reversal, defined as the prior month's return. MOM is stock momentum, defined as the prior year's cumulative return, skipping the last month. To be included in the sample, firms are required to have non-missing values of the following items: market value of equity, book-to-market equity, gross profit, total assets, current month return, and cumulative return in the past one-year period. Financial firms are excluded. Detailed constructions of variables are described in Appendix A.

Panel A: Summary statistics for the U.S. sample from July 1999 to December 2018.

| | Mean | SD | 1st | 25th | 50th | 75th | 99th |
|---|---|---|---|---|---|---|---|
| OP | 0.101 | 0.203 | -0.616 | 0.054 | 0.122 | 0.188 | 0.473 |
| ACC | -0.040 | 0.172 | -0.536 | -0.085 | -0.036 | 0.009 | 0.426 |
| Cash-OP | 0.099 | 0.211 | -0.647 | 0.044 | 0.119 | 0.192 | 0.504 |
| Ln(B/M) | -0.720 | 0.954 | -3.383 | -1.251 | -0.653 | -0.127 | 1.525 |
| Ln(ME) | 5.997 | 2.109 | 1.636 | 4.486 | 5.976 | 7.421 | 11.078 |
| STR | 0.011 | 0.170 | -0.353 | -0.068 | 0.001 | 0.073 | 0.534 |
| MOM | 0.139 | 0.673 | -0.767 | -0.206 | 0.039 | 0.322 | 2.496 |

Panel B: Fama-MacBeth regressions using all-but-micros stocks

| | (2b) | (3b) | (4b) | (5b) | (6b) | (7b) |
|---|---|---|---|---|---|---|
| OP | 2.49 | | 2.53 | | | 1.13 |
| | (2.60) | | (2.71) | | | (1.15) |
| ACC | | -0.10 | 0.03 | | 0.42 | |
| | | (-0.22) | (0.06) | | (0.96) | |
| Cash-OP | | | | 2.51 | 2.61 | 1.48 |
| | | | | (2.90) | (3.04) | (2.54) |
| Ln(B/M) | 0.12 | 0.05 | 0.12 | 0.12 | 0.13 | 0.13 |
| | (0.81) | (0.37) | (0.84) | (0.80) | (0.89) | (0.83) |
| Ln(ME) | -0.10 | -0.07 | -0.10 | -0.11 | -0.10 | -0.11 |
| | (-2.08) | (-1.50) | (-2.04) | (-2.15) | (-2.10) | (-2.13) |
| STR | -1.23 | -1.10 | -1.25 | -1.22 | -1.24 | -1.23 |
| | (-1.61) | (-1.47) | (-1.63) | (-1.60) | (-1.62) | (-1.60) |
| MOM | -0.06 | -0.05 | -0.07 | -0.07 | -0.07 | -0.06 |
| | (-0.12) | (-0.09) | (-0.13) | (-0.14) | (-0.14) | (-0.13) |
| Adj_$R^2$ | 5.8% | 5.6% | 6.0% | 5.7% | 6.0% | 5.9% |

**Appendix C: PBFJ Replication Study Master Pitch**

| | | | | | |
|---|---|---|---|---|---|
| Pitcher Team Names | WEI, KC John, Hong Kong Polytechnic University<br>DU, Qingjie, Hong Kong Polytechnic University<br>WANG, Yang, Hong Kong Polytechnic University | JEL Code | G11;G12;M41 | Date Completed | Feb 28, 2020 |
| FOUR | Four aspects of BIG picture framing | | | | |
| (A) Working Title | Does cash-based operating profitability explain the accruals anomaly in China? | | | | |
| (B) Basic Research Question | Our research question is the same as our working title, which is: does cash-based operating profitability explain the accruals anomaly in China? This RQ is almost the same as that in the original paper, but we change the setting to the Chinese market. | | | | |
| (C) Key paper | Target Replication Paper:<br>Ball, Ray, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri Nikolaev. "**Accruals, cash flows, and operating profitability in the cross section of stock returns,**" Journal of Financial Economics 121, no. 1 (2016): 28-45.<br>Key papers:<br>1, Liu, Jianan, Robert F. Stambaugh, and Yu Yuan. "**Size and value in China**," Journal of Financial Economics 134. No. 1 (2019): 48-69.<br>2, Green, Jeremiah, John R. M. Hand, and Mark T. Soliman. "**Going, going, gone? The apparent demise of the accruals anomaly**," Management Science 57, no. 5 (2011): 797-816. | | | | |
| (D) Motivation/Puzzle | Accrual accounting adjusts the current-period cash flows with the objective of better measuring the firm's current-period performance, making accrual accounting more useful than cash-based accounting in contracting contexts. While previous literature documents two anomalies in empirical asset pricing: accruals anomaly (high accruals, low returns), and profitability anomaly (high profitability, high returns). Given that accruals are a component of a firm's profitability, it seems puzzling to have the two anomalies with opposite directions of prediction. The original paper finds cash-based operating profitability subsumes the return predictability of both accruals and operating profitability, and argues that cash-based operating profitability is a superior profitability measure. We intend to replicate this paper in the Chinese market since there are some unique differences between the two markets, including shorter history, immature market development, different trading rules, different accounting reporting standards, different culture, and delisting rules. Thus, it is worth conducting a replication in the Chinese market. | | | | |
| THREE | Three core aspects of any empirical research project | | | | |
| (E) Idea? | The main idea in the original paper is that only cash-based operating profitability matters in the cross-section of expected stock returns (**Table 2**). After controlling for cash-based operating profitability, accruals and operating profitability become insignificant. The negative return predictability of accruals is due to its negative correlation with the cash-based component of profitability. The accruals component has no incremental contribution to the cross-section of expected stock returns once cash-based operating profitability is controlled for. | | | | |
| (F) Data? | For the U.S. market replication, we will use the CRSP and Compustat/NA, the same as those used in the original paper. For the Chinese market, we will use the CSMAR. All databases are well known commercial data sources, which have | | | | |

| | |
|---|---|
| | the largest coverage in the corresponding market. All databases are panel data, which contain time-series and cross-sectional variations. Stock returns are in both monthly and daily frequency, while accounting data are annual data. |
| (G) Tools | The main research tool is the Fama-MacBeth regression. We will also control a set of well identified control variables in the regression. This testing method is the same as that used in the original paper, and it is a widely accepted standard in the current literature. Portfolio-based analysis is also used to rule out the outlier issue if we have enough space to show the results. |
| TWO | Two key questions |
| (H) What's New? | The essential novelty of the original paper is that they find that cash-based operating profitability subsumes both accruals and the profitability anomalies. The predictability of accruals comes from the negative correlation with the cash component, rather than the misunderstanding of the accruals persistence as proposed by Sloan (1996). |
| (I) So What? | The U.S. is the world's largest equity market. It is important to know how the different markets work. The original paper investigates how accounting profitability is priced in stock returns, and helps investors to understand the market mechanism. Different from the U.S., the Chinse market is relatively young and immature. The market mechanism and institution environments may be different. It is meaningful to conduct such a replication test in the Chinese market. The outcome of this study helps to uncover the market mechanism and deepen our understanding in the largest emerging market. |
| ONE | One bottom line |
| (J) Contributions? | The primary contribution of the original paper is that it finds a superior profitability measure and provides the explanation for both accruals and profitability anomalies. The proposed replication study may also provide some evidence on how market development and institution environments can affect asset pricing, because the Chinese market is relatively young and less developed. |
| (K) Other Considerations | Since Compustat back-fills some data (if any) when it is updated, it is impossible to 100% reproduce the original results. However, all our results show similar return patterns to those in the original paper, suggesting that our replication of the U.S. results are successful. |

```
********************code to replicate the US test************************************;
dm "out;clear;log;clear;";

%let dc = c:\ ;
libname ff "&dc\wrds\ff\sasdata";
libname crspq "&dc\wrds\crsp\sasdata\q_stock";
libname comp ("&dc\wrds\comp\sasdata\d_na","&dc\wrds\comp\sasdata\d_na\company");
libname ddhome "&dc\wrds\ddhome";

%include "&dc/wrds/wrdsmacros/nwords.sas";
%include "&dc/wrds/wrdsmacros/crspmerge.sas";
%include "&dc/wrds/wrdsmacros/winsorize.sas";
%include "&dc/wrds/wrdsmacros/trade_date_windows.sas";
*******************************************************************************;
*==============================================================================
                        start with COMPUSTAT ANNUAL information
==============================================================================;
proc sql; create table funda
        as select             /*header info*/
            c.gvkey,c.conm, datadate,fyear, sic, sich,
            /*firm general variables*/
            AT, CEQ, DLTT,  LT, MIB, PSTK, PSTKR, PSTKL, PSTKRV, SEQ, TXDITC, TXDB, ITCB,
            /*operating profitability from balance sheet*/
            SALE, REVT, COGS, XSGA, XRD,
            /*Cash base operating profability based on balance sheet*/
            RECT, INVT, XPP, DRC, DRLT, AP, XACC,
            /*accruals based on balance sheet*/
            ACT, CH, CHE, LCT, DLC, TXP, DP
      from comp.company as c, comp.funda as f
      where f.gvkey = c.gvkey and f.indfmt='INDL' and f.datafmt='STD' and f.popsrc='D'
and f.consol='C'; quit;

      /*sort and clean up*/
      proc sort data=funda nodupkey; by gvkey datadate; run;


data funda01; set funda; by gvkey datadate; retain fage;
if first.gvkey then fage=1; else fage+1; if sich<0 then sich=sic*1; drop sic;

BE=coalesce(SEQ,(CEQ+coalesce(PSTK,0)),(AT-LT))-
coalesce(PSTKR,PSTKL,PSTK,0)+coalesce(TXDITC,SUM(TXDB,ITCB),0);

*construct the raw operating profitability as in the replicated paper;
opraw =     revt-cogs-sum(xsga,0) + sum(xrd,0);
if at>0 then op_raw = opraw/at;     *gross profitability including accruals;

*balance sheet based accrual and cash flow;
lag_rect=lag(rect); if fage=1 then lag_rect='';
lag_invt=lag(invt); if fage=1 then lag_invt='';
lag_xpp=lag(xpp); if fage=1 then lag_xpp='';
lag_drc=lag(drc); if fage=1 then lag_drc='';
lag_drlt=lag(drlt); if fage=1 then lag_drlt='';
lag_ap=lag(ap); if fage=1 then lag_ap='';
lag_xacc=lag(xacc); if fage=1 then lag_xacc='';

*cash based operating profitability;
if at>0 then opcat=(    opraw
                                  -sum(rect,-lag_rect,0)
                                  -sum(invt,-lag_invt,0)
```

```
                                    -sum(xpp,-lag_xpp,0)
                                    +sum(drc,drlt,-lag_drc,-lag_drlt,0)
                                    +sum(ap,-lag_ap,0)
                                    +sum(xacc,-lag_xacc,0)
                                )/at;
opc=opcat*at;
drop lag_rect lag_invt lag_xpp lag_drc lag_drlt lag_ap lag_xacc;

lag_act=lag(act); if fage=1 then lag_act='';
lag_ch=lag(ch); if fage=1 then lag_ch='';
lag_lct=lag(lct); if fage=1 then lag_lct='';
lag_dlc=lag(dlc); if fage=1 then lag_dlc='';
lag_txp=lag(txp); if fage=1 then lag_txp='';

*cash based accruals; *measure claimed in Ball et al (2016), but it is not correct;
if at>0 then  accat =(  sum(act,-lag_act,0)
                                    - sum(ch,-lag_ch,0)
                                    - ( sum(lct,-lag_lct,0) -sum(dlc,-lag_dlc,0) -
sum(txp,-lag_txp,0) )
                                    - sum(dp,0)
                                )/at;
drop lag_act lag_ch lag_lct lag_dlc lag_txp; run;


*=========================Create merge with CRSP==============================;
*=====================GET CRSP IDENTIFIER============================;
proc sort data=ddhome.ccmxpf_linktable out=lnk;
where LINKTYPE in ("LU", "LC", "LD", "LF", "LN", "LO", "LS", "LX") and
(2018 >= year(LINKDT) or LINKDT = .B) and (1950 <= year(LINKENDDT) or LINKENDDT = .E);
by GVKEY LINKDT; run;

proc sql; create table temp
as select a.lpermno as permno,b.*
from lnk as a, funda01 as b
where a.gvkey=b.gvkey
and (LINKDT <= b.datadate or LINKDT = .B) and (b.datadate <= LINKENDDT or LINKENDDT = .E)
and lpermno ne . and not missing(b.gvkey);
quit;

*=============================================================================
Finalize first Compustat data set
This is most of the annual compustat variables plus a couple components that still need
additional information
=============================================================================;

data funda;
retain permno datadate gvkey conm sich /*identifior*/
fage at be opraw op_raw opc opcat accat /*constructed variables*/
;
set temp; where not missing(permno);
keep gvkey conm permno datadate sich /*identifior*/
fage at be opraw op_raw opc opcat accat/*constructed variables*/
;run;

proc sort data=funda nodupkey; by permno datadate;run;
proc delete data=temp lnk funda01;run;


*=============================================================================
            extract the monthly stock return data
=============================================================================;
```

```
*-----------Included delisted returns in the monthly returns--------------------;
*----------------------screen for only NYSE, AMEX, NASDAQ, and common stock-------------;
proc sort data=crspq.mseall(keep=date permno exchcd shrcd siccd COMNAM) out=msealla
nodupkey; by permno exchcd date; run;

proc sql; create table mseall as
select *,min(date) as exchstdt,max(date) as exchedt
from msealla group by permno,exchcd; quit;

proc sort data=mseall nodupkey;     by permno exchcd; run;
*-----------extract the monthly returns--------------------;
proc sort data=crspq.msf out=msf nodupkey; by permno date; run;

proc sql; create table msf0
as select a.permno, a.date, a.ret, abs(a.prc) as prc, a.shrout, a.vol,
b.exchcd, b.shrcd, b.siccd, b.COMNAM
from msf as a left join mseall as b
on a.permno=b.permno and b.exchstdt<=a.date<= b.exchedt;
run; quit;

proc sort data=msf0 nodupkey; by permno date;run;
proc delete data=msf mseall msealla;run;

*-----------Included delisted returns in the monthly returns--------------------;
proc sql; create table msf1 as select a.*, b.dlret,b.dlstcd, b.DLSTDT
from msf0 as a left join crspq.msedelist as b
on a.permno=b.permno and (intck('month',b.DLSTDT,a.date) between 0 and 0); run; quit;

proc sort data=msf1 nodupkey; by permno date;run;
data msf; set msf1; where  shrcd in (10,11) and exchcd in (1,2,3);
size=abs(prc*shrout);

*Adjust volume for NASDAQ stocks following Anderson and Dyl (2005);
if exchcd = 3 then      vol = (date <= '01jan1997'd)*0.5*vol + (date >
'01jan1997'd)*0.62*vol;
ret_raw=ret;
if missing(dlret) and (dlstcd=500 or (dlstcd>=520 and dlstcd<=584)) and exchcd in (1,2)
then dlret=-.35;
if missing(dlret) and (dlstcd=500 or (dlstcd>=520 and dlstcd<=584)) and exchcd in (3)
then dlret=-.55;
                        *see Johnson and Zhao (2007), Shumway and Warther (1999) etc.;
if not missing(dlret) and dlret<-1 then dlret=-1;

if Nmiss(ret,dlret)<2 then ret=sum(ret,dlret);
if ret<=-1 and missing(ret)=0 then ret=-1;
sic2=floor(siccd/100); run;

proc sort data=msf nodupkey; by permno date; run;


*merge the funda data and the market value data into the msf dataset;
proc sql;   create table msf_funda_ball6
      as select a.*,b.*
      from msf as a left join funda as b
      on a.permno=b.permno and (intck('month',b.datadate,a.date) between 7 and 18);
      quit;

proc sort data=msf_funda_ball6; by permno date datadate;run;
data msf_funda_ball6; set msf_funda_ball6; by permno date datadate; if last.date; run;
data msf_funda; set msf_funda_ball6;run;
```

```
*merge the december MV into the file;
data msf12; set msf; where month(date)=12; keep permno date exchcd size; run;
proc sql;   create table msf_funda01
      as select a.*,b.size as ME12
      from msf_funda as a left join msf12 as b
      on a.permno=b.permno and (intck('month',b.date,a.date) between 7 and 18); quit;

proc sort data=msf_funda01 nodupkey; by permno date;run;

*merge the micro stock information into the file;
*estabilish the portfolio in each July;
data msf6; set msf; where month(date)=6; keep permno date exchcd size; run;

*construct the micro stock and non-mocro stock based on NYSE 20% breakpoint;
data nyse20; set msf6; where exchcd=1;run;
proc sort data=nyse20; by date size;run;
proc univariate data=nyse20 noprint;
var size; by date; output out=nyse20rnk  pctlpts=20 80 pctlpre=nyse; run;

proc sql;   create table msf6a
      as select a.*,b.nyse20  from msf6 as a left join nyse20rnk as b
      on a.date=b.date;       quit;

data msf6;  set msf6a;  if size<nyse20 and size>0 then micro=1;   if size>=nyse20 then
micro=0;    run;
proc delete data=msf6a nyse20 nyse20rnk;run;


proc sql;   create table msf_funda
      as select a.*,b.size as ME6, b.nyse20, b.micro
      from msf_funda01 as a left join msf6 as b
      on a.permno=b.permno and (intck('month',b.date,a.date) between 1 and 12);
      quit;

proc sort data=msf_funda nodupkey; by permno date;run;

proc delete data=msf_funda01 msf6 msf12;run;



*construct addition control variable;
*can use monthly market cap and price, but need to lag because it is currently
contemporaneous with the returns we want to predict;
proc sql;   create table temp01
      as select a.*,b.size/1000 as mep, log(b.size/1000) as lnmep, b.prc as prch, b.ret
as str, b.date as yyyymm
      from msf_funda as a left join msf_funda as b
      on a.permno=b.permno and (intck('month',b.date,a.date) between 1 and 1);
      quit;

proc sort data=temp01 nodupkey; by permno date; run;
data msf_base; set temp01; by permno date;
retain mn;  if first.permno then mn=1; else mn+1;

date12=lag12(date); format date12 date9.;
nmonth=intck('month',date12,date);
mom =  (   (1+lag2(ret))*(1+lag3(ret))*(1+lag4(ret))*(1+lag5(ret))*(1+lag6(ret))*

      (1+lag7(ret))*(1+lag8(ret))*(1+lag9(ret))*(1+lag10(ret))*(1+lag11(ret))*(1+lag12(re
t))   ) - 1;
if mn<13 then mom=.; if nmonth ne 12 then mom='';
```

```
if me12>0 then BtM=1000*BE/ME12;
if BtM>0 then lnbtm=log(BtM);
drop date12 nmonth mn me6 nyse20;
drop DLSTDT dlret dlstcd; run;

proc delete data=temp01;run;

data msf_base01; retain permno date yyyymm ret mep str mom at lnbtm opc opcat micro lnmep
accat opraw op_raw;
set msf_base; if year(date)<=2014 and date>='01JUL1963'd and
nmiss(mep,lnbtm,op_raw,at,ret,str,mom)=0;
drop prc shrout vol; run;
data msf_base01; set msf_base01; if sic2>=70 or sic2<60; run;

*****************************************************************
do some summary statistic
*****************************************************************;
proc sort data=msf_base01; by date permno;run;

proc univariate data=msf_base01 noprint;
var op_raw accat opcat btm lnbtm lnmep str mom; by date;
output out=t1_pa  pctlpts=1 25 50 75 99 pctlpre=op_raw accat opcat btm lnbtm lnmep str
mom; run;

proc means data=t1_pa noprint;
var op_raw1 op_raw25 op_raw50 op_raw75 op_raw99       accat1 accat25 accat50 accat75
accat99
      opcat1 opcat25 opcat50 opcat75 opcat99          btm1 btm25 btm50 btm75 btm99
      lnbtm1 lnbtm25 lnbtm50 lnbtm75 lnbtm99               lnmep1 lnmep25 lnmep50
lnmep75 lnmep99
      str1 str25 str50 str75 str99                        mom1 mom25 mom50 mom75 mom99;
output out=table1_pa  mean=op_raw1 op_raw25 op_raw50 op_raw75 op_raw99  accat1 accat25
accat50 accat75 accat99
      opcat1 opcat25 opcat50 opcat75 opcat99          btm1 btm25 btm50 btm75 btm99
      lnbtm1 lnbtm25 lnbtm50 lnbtm75 lnbtm99               lnmep1 lnmep25 lnmep50
lnmep75 lnmep99
      str1 str25 str50 str75 str99                        mom1 mom25 mom50 mom75 mom99;
run;


proc sort data=msf_base01; by date; run;
proc means data=msf_base01 noprint;
by date;
var op_raw accat opcat btm lnbtm lnmep str mom;
output out=t1_pb n=nstock mean= op_raw accat opcat btm lnbtm lnmep str mom
                              std=op_raw_sd accat_sd opcat_sd btm_sd lnbtm_sd lnmep_sd
str_sd mom_sd;
run;

proc means data=t1_pb noprint;
var op_raw accat opcat btm lnbtm lnmep str mom
      op_raw_sd accat_sd opcat_sd btm_sd lnbtm_sd lnmep_sd str_sd mom_sd;
output out=table1_pb mean= op_raw accat opcat btm lnbtm lnmep str mom
      op_raw_sd accat_sd opcat_sd btm_sd lnbtm_sd lnmep_sd str_sd mom_sd;
run;

*****************************************************************
correlation matrix
*****************************************************************;
```

```
proc sort data=msf_base01; by date;run;
data msf_corr; set msf_base01;
rename op_raw=v10;
rename accat=v11;
rename opcat=v12; run;

proc corr data=msf_corr outp=acc_pcr outs=acc_scr noprint; by date; var    v10-v12; run;

%let bitcordd=acc_pcr;
data bit_c; set &bitcordd; if _type_='CORR'; run;

proc sort data=bit_c; by _name_;run;
proc means data=bit_c noprint; by _name_; var  v10-v12; output out=bit_cormat(drop=_type_
_freq_); run;

data table1_pc; set bit_cormat; where _STAT_='MEAN';
if _name_='v10' then label='op_gross';
if _name_='v11' then label='accat';
if _name_='v12' then label='opcat'; run;


%let bitcordd=acc_scr;
data bit_c; set &bitcordd; if _type_='CORR'; run;

proc sort data=bit_c; by _name_;run;
proc means data=bit_c noprint; by _name_; var  v10-v12; output out=bit_cormat(drop=_type_
_freq_); run;

data table1_pd; set bit_cormat; where _STAT_='MEAN';
if _name_='v10' then label='op_raw';
if _name_='v11' then label='accat';
if _name_='v12' then label='opcat'; run;

*****************************************************************
summary statistic ends here
*****************************************************************;


* Call and Run the table1 SAS Macro ;
%include 'C:\wrds\macro\winsor.sas';
%winsor(dsetin=msf_base01, dsetout=msf_trim, byvar=date, vars=op_raw opcat lnbtm lnmep
str mom, type=delete, pctl=1 99);

data msf_fmb; set msf_trim;
retp=ret*100; ewew=1; vwvw=mep;
keep permno date retp ret ewew vwvw
       mep str mom at lnbtm opc opcat micro lnmep accat opraw op_raw;
attrib _all_ lable=''; run;

proc sort data=msf_fmb; by date;run;

*set 1, ewew result using large stocks;
%let y = retp; *define dep var here;
%let wtt=ewew;  *ewew or vwvw;
%let micro = 0; *indicator of micro stock or not;

proc reg data=msf_fmb outest=FB noprint;
where micro = &micro ;  by date;
 *ORIGNAL setting;
      model &y = op_raw lnbtm lnmep str mom/adjrsq;
      model &y = accat lnbtm lnmep str mom/adjrsq;
```

```
      model &y = op_raw accat lnbtm lnmep str mom/adjrsq;
      model &y = opcat lnbtm lnmep str mom/adjrsq;
      model &y = accat opcat lnbtm lnmep str mom/adjrsq;
      model &y = op_raw opcat lnbtm lnmep str mom/adjrsq;
      weight &wtt;
quit;

** 2. drop irrelevant estimates;
proc sort data=FB; by _model_ date; run;
data FB2; set FB; drop  &y _TYPE_  _DEPVAR_  _RMSE_ _IN_  _P_  _EDF_;  rename
_model_=model; run;
proc transpose data=FB2 out=FBny name=name prefix=coef; by model date; run;
data FBny; set FBny; retain code;  by model date; code=code+1; if first.date then
code=1;run;
proc sort data=FBny; by model code name;run;

** 3. Newey-West t-stat for the time-series average of coefficients;
%let lag=12;*lags for Newey-West t-stat;

proc model data=FBny;  by model code name;
 parms a; exogenous coef1 ;
 instruments / intonly;
 coef1 =a;
 fit coef1 / gmm kernel=(bart, %eval(&lag+1), 0);
 ods output parameterestimates=param1  fitstatistics=fitresult
 OutputStatistics=residual;
quit;

** 4. output into column table;
data param1; set param1;
 tvalue2=put(tvalue,7.3); if probt<0.1 then p='*  ';
 if probt<0.05 then p='** '; if probt<0.01 then p='***';
 T=compress('('||tvalue2||')'); PARAM=compress(put(estimate,7.3)||p);
 run;

data param1a; set param1; keep model code name coef _name_;  _name_='PARAM'; coef=PARAM;
run;
data param1b; set param1; keep model code name coef _name_; _name_='T'; coef=T; run;
data param2; set param1a param1b;run;
proc sort data=param2; by  code _name_ model;run;
proc transpose data=param2 out=param3;  by code name _name_; id model; var coef; run;
data param3; set param3; if _name_='T' then do; code=. ;name=.;end;run;

** 5. find the range of periods and obs used;
proc sort data=fb out=fb3; by _model_; run;
data fb3; set fb3; keep _model_ date num; num = _edf_+_p_;  rename _model_=model; run;
proc sql; create table num(where=(model='MODEL1')) as select
 model, min(date) as start, max(date) as end, count(date) as range,
 sum(num) as obs from fb3 group by model;quit;
proc transpose data=num out=num; by model; var start -- obs; run;
data num; set num; rename _name_=name; MODEL1=put(col1, 7.0);  drop model col1; run;
data param3; set param3 num; run;

** 6. output as excel file;
proc export   data=param3  dbms=xlsx
outfile="c:\project107_acc_pbfj\result\T2_PA_FMB_micro_&micro._&wtt..xlsx"
  replace; run;


************ US market code ends here ****************************************;
```

```
********************code to replicate the Chinese test*********************************;
*****************************************************************************************;
dm "out;clear;log;clear;";
%let dc = c:\ ;
libname ccrsp "&dc\wrds\csmar\sasdata\ccrsp";
libname ccomp ("&dc\wrds\csmar\sasdata\ccomp");
%include "&dc/wrds/wrdsmacros/nwords.sas";
%include "&dc/wrds/wrdsmacros/crspmerge.sas";
%include "&dc/wrds/wrdsmacros/winsorize.sas";
%include "&dc/wrds/wrdsmacros/trade_date_windows.sas";
*****************************************************************************************;

*=======================================================================================
                        start with CSMAR Balance sheet information
=======================================================================================;
data aa; set ccomp.combas; where typrep in ("A") and (day(datadate) ne 1) and
month(datadate)=12; gvkey     =     Stkcd;
AT          =     A001000000; SEQ         =     A003000000;
MIN_INT     =     A003200000; XRD         =     A001219000;
*items to adjust the operating profitability;
RECT   =     A001111000; *account receivables;
INVT   =     A001123000; *inventory;
XPP    =     A001112000; *Net Prepayments;
DRC    =     A002127000; *Deferred Income Current Liabilities;
DRLT   =     A002210000; *Deferred Income Non-current Liabilities;
AP     =     A002108000; *account payable;
XACC   =     A002120000; *Other Payables;
*items to calculate the accrual measure;
ACT         =     A001100000; *total current asset;
CH          =     A001101000; *cash amount;
LCT         =     A002100000; *total current liability;
DLC         =     sum(A002101000,A002125000); *debt in current liability;
TXP         =     A002113000; *Taxes Payable;

keep gvkey datadate at seq min_int XRD RECT INVT XPP DRC DRLT AP XACC ACT CH LCT DLC TXP;
run;
proc sort data=aa nodupkey; by gvkey datadate;run;

data bb; set ccomp.comins; where typrep in ("A"); gvkey     =     Stkcd;
Operat_p=   B001300000; *operating income;
keep gvkey datadate Operat_p; run;
proc sort data=bb nodupkey; by gvkey datadate;run;

data dd; set ccomp.comscfi; where typrep in ("A"); gvkey    =     Stkcd;
DP          =     sum(D000103000,D000104000,D000105000);
keep gvkey datadate DP; run;
proc sort data=dd nodupkey; by gvkey datadate;run;

proc sql; create table temp as select c.Conme_en, c.Indcd, c.Indnme_en, a.*, operat_p
      from AA as A, BB as B, ccrsp.company as c
      where A.gvkey = B.gvkey=c.stkcd and A.datadate = b.datadate; run; quit;
proc sort data=temp nodupkey; by gvkey datadate; run;

proc sql; create table funda as select    a.*, d.dp
      from temp as A left join dd as d on A.gvkey = d.gvkey and A.datadate = d.datadate;
      run; quit;
proc sort data=funda nodupkey; by gvkey datadate; run;

data funda01;           set funda; by gvkey datadate; where indcd ne "0001";
retain fage; if first.gvkey then fage=1; else fage+1;
```

```
BE=SEQ-coalesce(MIN_INT,0);

*construct the raw operating profitability as in the replicated paper;
opraw =   operat_p + sum(xrd,0);
if at>0 then op_raw = opraw/at;     *gross profitability including accruals;

*total accruals, balance sheet based accrual and cash flow;
lag_rect=lag(rect); if fage=1 then lag_rect='';
lag_invt=lag(invt); if fage=1 then lag_invt='';
lag_xpp=lag(xpp); if fage=1 then lag_xpp='';
lag_drc=lag(drc); if fage=1 then lag_drc='';
lag_drlt=lag(drlt); if fage=1 then lag_drlt='';
lag_ap=lag(ap); if fage=1 then lag_ap='';
lag_xacc=lag(xacc); if fage=1 then lag_xacc='';

*cash based operating profitability;
if at>0 then opcat=(    opraw
                                    -sum(rect,-lag_rect,0)
                                    -sum(invt,-lag_invt,0)
                                    -sum(xpp,-lag_xpp,0)
                                    +sum(drc,drlt,-lag_drc,-lag_drlt,0)
                                    +sum(ap,-lag_ap,0)
                                    +sum(xacc,-lag_xacc,0)
                               )/at;
opc=opcat*at;
drop lag_rect lag_invt lag_xpp lag_drc lag_drlt lag_ap lag_xacc;

*balance sheet accruals; *measure claimed in Ball et al (2016), but it is not correct;
lag_act=lag(act); if fage=1 then lag_act='';
lag_ch=lag(ch); if fage=1 then lag_ch='';
lag_lct=lag(lct); if fage=1 then lag_lct='';
lag_dlc=lag(dlc); if fage=1 then lag_dlc='';
lag_txp=lag(txp); if fage=1 then lag_txp='';
if at>0 then  accat     =(    sum(act,-lag_act,0)
                                    - sum(ch,-lag_ch,0)
                                    - ( sum(lct,-lag_lct,0) -sum(dlc,-lag_dlc,0) -
sum(txp,-lag_txp,0) )
                                    - sum(dp,0)
                               )/at;
drop lag_act lag_ch lag_lct lag_dlc lag_txp; run;

*=============================================================================================
Finalize first Compustat data set
==============================================================================================;
data cfunda; retain permno datadate gvkey fage at be opraw op_raw opc opcat accat;
set funda01; permno=gvkey;
at=at/1000000000; *change to billions;
be=be/1000000000;
opraw=opraw/1000000000;
opc=opc/1000000000;
keep gvkey permno datadate fage at be opraw op_raw opc opcat accat; run;

proc sort data=cfunda nodupkey; by permno datadate;run;

*----------------------screen for only A share stock, and common stock-------------;
**** step 1: extract all the A-share stock;
data company; set ccrsp.company; permno=stkcd; yr=substr(Listdt,1,4)*1;
mn=substr(Listdt,6,2)*1; dy=substr(Listdt,9,2)*1;
IPOdate=mdy(mn,dy,yr); format IPOdate YYMMDDN8.; drop yr mn dy;  run;

data cmsf1; set ccrsp.cmsf; where Markettype in (1,4,16);
```

```
permno=stkcd; date=datedd; ret=Mretwd; prc=Mclsprc; size=Msmvttl/1000000;
ME=Msmvttl/1000000; MET=Msmvosd/1000000;
format date YYMMDDN8.; drop stkcd datedd Msmvttl Msmvosd;
drop Opndt Capchgdt Mretnd Mretwd Mclsprc Mopnprc Clsdt Mnshrtrd Mnvaltrd;
run;
*Market Type: 1=SSE A share, 2= SSE B share, 4= SZSE A share, 8= SZSE B share, 16= GEM;

proc sql; create table cmsf as select a.permno, a.date, b.Conme_en, b.IPOdate, a.*
from cmsf1 as a left join company as b on a.permno=b.permno; quit;
proc delete data=cmsf1 company; run;
proc sort data=cmsf nodupkey; by permno date;run;

proc sql; create table cmsf1 as select a.*,
                        b.ret as str, b.Ndaytrd as trade10, b.prc as PRCH, b.Me as MEP,
log(b.me) as lnmep, b.met as metp
from cmsf as a left join cmsf as b
on a.permno=b.permno and  intck('month',b.date,a.date) = 1; quit;
proc sort data=cmsf1; by permno date; run;

data cmsf_base; set cmsf1; by permno date;
retain count;     if first.permno then count=1; else count+1;
date12=lag12(date); format date12 date9.;
nmonth=intck('month',date12,date);
*construct the mom factor;
mom =  (   (1+lag2(ret))*(1+lag3(ret))*(1+lag4(ret))*(1+lag5(ret))*(1+lag6(ret))*

      (1+lag7(ret))*(1+lag8(ret))*(1+lag9(ret))*(1+lag10(ret))*(1+lag11(ret))*(1+lag12(re
t))   ) - 1;
if count<13 then mom=.; if nmonth ne 12 then mom='';
drop date12 nmonth count; run;
proc delete data=cmsf1; run;

**** Three main filters ************************************************************;
*filter 1: identify the IPO stocks;
data cmsf2; set cmsf_base; ipom=intck('month',IPOdate,date)+1; run;

*filter 2: identify stocks has less than 120 trading days in past 12 months and less than
10 trading days in last month;
proc sql; create table cmsf3 as select a.*, b.date as date12p, b.Ndaytrd as Ndaytrd12p
from cmsf2 as a left join cmsf2 as b on a.permno=b.permno and
intck('month',b.date,a.date) between 1 and 12; quit;

proc sort data=cmsf3; by permno date;run;
proc means data=cmsf3 noprint; by permno date; var Ndaytrd12p; output out=trade120 n=nm
sum=trade120;run;
proc delete data=cmsf3;run;

proc sql; create table cmsf3 as select a.*, b.trade120, b.nm as nm12
from cmsf2 as a left join trade120 as b on a.permno=b.permno and  b.date=a.date; quit;
proc delete data=cmsf2 trade120;run;

data cmsf4; retain permno Conme_en date Trdmnt IPOdate ipom nm12 trade120 trade10;
set cmsf3; if ipom>=7 and (trade120>=120 and trade10>=10);run;
proc delete data=cmsf3;run;

************ Identify the shell stocks*******************;
proc sort data=cmsf4; by date yr mn mep;run;
proc univariate data=cmsf4 noprint; where mep>0;
var mep; by date yr mn; output out=cm30  pctlpts=10 20 30 40 50 60 70 80 90 pctlpre=cm;
run;
```

```
proc sql; create table cmsf5 as select a.*, b.cm30 from cmsf4 as a left join cm30 as b on
a.yr=b.yr and a.mn=b.mn; quit;

**check the percentage of market cap**;
data cmsf6; set cmsf5; if mep>=cm30 then shell=0; else shell=1;run;
proc delete data=cmsf5 cmsf4 cm30;run;

*merge the funda data and the market value data into the msf dataset;
proc sql;   create table cmsf_funda as select a.*,b.*
      from cmsf6 as a left join cfunda as b
      on a.permno=b.permno and (intck('month',b.datadate,a.date) between 7 and 18); quit;

proc sort data=cmsf_funda; by permno date datadate;run;
data cmsf_funda; set cmsf_funda; by permno date datadate; if last.date; run;

*merge the december MV into the file;
data cmsf12; set cmsf; where month(date)=12; keep permno date size; run;
proc sql;   create table cmsf_funda01     as select a.*,b.size as ME12
      from cmsf_funda as a left join cmsf12 as b
      on a.permno=b.permno and (intck('month',b.date,a.date) between 7 and 18);     quit;
proc sort data=cmsf_funda01 nodupkey; by permno date;run;

*construct addition control variable;
data cmsf_base; set cmsf_funda01; by permno date; if me12>0 then BtM=BE/ME12; if BtM>0
then lnbtm=log(BtM);run;

************************do some summary statistic***************;
data msf_base01; retain permno date ret mep str mom at lnbtm opc opcat shell lnmep accat
opraw op_raw;
set cmsf_base; if year(date)<=2018 and date>='01JUL1999'd and
nmiss(mep,lnbtm,op_raw,at,ret,str,mom)=0; run;
data msf_base01; set msf_base01; if shell=0; run; *exclude the shell stocks;
proc sort data=msf_base01; by date permno;run;

proc univariate data=msf_base01 noprint;
var op_raw accat opcat btm lnbtm lnmep str mom; by date;
output out=t1_pa  pctlpts=1 25 50 75 99 pctlpre=op_raw accat opcat btm lnbtm lnmep str
mom; run;

proc means data=t1_pa noprint;
var op_raw1 op_raw25 op_raw50 op_raw75 op_raw99       accat1 accat25 accat50 accat75
accat99
      opcat1 opcat25 opcat50 opcat75 opcat99          btm1 btm25 btm50 btm75 btm99
      lnbtm1 lnbtm25 lnbtm50 lnbtm75 lnbtm99                lnmep1 lnmep25 lnmep50
lnmep75 lnmep99
      str1 str25 str50 str75 str99                          mom1 mom25 mom50 mom75 mom99;
output out=table1_pa  mean=op_raw1 op_raw25 op_raw50 op_raw75 op_raw99  accat1 accat25
accat50 accat75 accat99
      opcat1 opcat25 opcat50 opcat75 opcat99          btm1 btm25 btm50 btm75 btm99
      lnbtm1 lnbtm25 lnbtm50 lnbtm75 lnbtm99                lnmep1 lnmep25 lnmep50
lnmep75 lnmep99
      str1 str25 str50 str75 str99                          mom1 mom25 mom50 mom75 mom99;
run;

proc sort data=msf_base01; by date; run;
proc means data=msf_base01 noprint;
by date; var op_raw accat opcat btm lnbtm lnmep str mom;
output out=t1_pb n=nstock mean= op_raw accat opcat btm lnbtm lnmep str mom
                                std=op_raw_sd accat_sd opcat_sd btm_sd lnbtm_sd lnmep_sd
str_sd mom_sd; run;
```

```
proc means data=t1_pb noprint;
var op_raw accat opcat btm lnbtm lnmep str mom
      op_raw_sd accat_sd opcat_sd btm_sd lnbtm_sd lnmep_sd str_sd mom_sd;
output out=table1_pb mean= op_raw accat opcat btm lnbtm lnmep str mom
      op_raw_sd accat_sd opcat_sd btm_sd lnbtm_sd lnmep_sd str_sd mom_sd; run;
*************************************************************
correlation matrix
*************************************************************;

proc sort data=msf_base01; by date;run;
data msf_corr; set msf_base01;
rename op_raw=v10;
rename accat=v11;
rename opcat=v12; run;

proc corr data=msf_corr outp=acc_pcr outs=acc_scr noprint; by date; var    v10-v12; run;

%let bitcordd=acc_pcr;
data bit_c; set &bitcordd; if _type_='CORR'; run;

proc sort data=bit_c; by _name_;run;
proc means data=bit_c noprint; by _name_; var  v10-v12; output out=bit_cormat(drop=_type_
_freq_); run;

data table1_pc; set bit_cormat; where _STAT_='MEAN';
if _name_='v10' then label='op_gross';
if _name_='v11' then label='accat';
if _name_='v12' then label='opcat'; run;

%let bitcordd=acc_scr;
data bit_c; set &bitcordd; if _type_='CORR'; run;

proc sort data=bit_c; by _name_;run;
proc means data=bit_c noprint; by _name_; var  v10-v12; output out=bit_cormat(drop=_type_
_freq_); run;

data table1_pd ; set bit_cormat; where _STAT_='MEAN';
if _name_='v10' then label='op_raw';
if _name_='v11' then label='accat';
if _name_='v12' then label='opcat'; run;
*summary statistic ends here*********************************************************;

*Fama and MacBeth (1973) Regression;
data cmsf_base01; retain permno date ret mep str mom at lnbtm opc opcat micro lnmep accat
opraw op_raw;
set cmsf_base; if year(date)<=2018 and date>='01JUL1999'd and
nmiss(mep,lnbtm,op_raw,at,ret,str,mom)=0;  run;
data msf_base01; set cmsf_base01; if shell>=0; run;
* Call and Run the table1 SAS Macro ;
%include 'C:\wrds\macro\winsor.sas';
%winsor(dsetin=msf_base01, dsetout=cmsf_trim, byvar=date, vars=op_raw opcat lnbtm lnmep
str mom, type=delete, pctl=1 99);

data msf_fmb; set cmsf_trim;
retp=ret*100; ewew=1; vwvw=mep; micro=shell;
keep permno date retp ret ewew vwvw
       mep str mom at lnbtm opc opcat micro shell lnmep accat opraw op_raw;
attrib _all_ lable=''; run;

proc sort data=msf_fmb; by date;run;
```

```
*set 1, ewew result using large stocks;
%let y = retp; *define dep var here;
%let wtt=ewew;  *ewew or vwvw;
%let micro = 0; *indicator of micro stock or not;

proc reg data=msf_fmb outest=FB noprint;
where micro = &micro ;  by date;
 *ORIGNAL setting;
      model &y = op_raw lnbtm lnmep str mom/adjrsq;
      model &y = accat lnbtm lnmep str mom/adjrsq;
      model &y = op_raw accat lnbtm lnmep str mom/adjrsq;
      model &y = opcat lnbtm lnmep str mom/adjrsq;
      model &y = accat opcat lnbtm lnmep str mom/adjrsq;
      model &y = op_raw opcat lnbtm lnmep str mom/adjrsq;
      weight &wtt; quit;

** 2. drop irrelevant estimates;
proc sort data=FB; by _model_ date; run;
data FB2; set FB; drop  &y _TYPE_  _DEPVAR_  _RMSE_ _IN_  _P_  _EDF_;  rename
_model_=model; run;
proc transpose data=FB2 out=FBny name=name prefix=coef; by model date; run;
data FBny; set FBny; retain code;  by model date; code=code+1; if first.date then
code=1;run;
proc sort data=FBny; by model code name;run;

** 3. Newey-West t-stat for the time-series average of coefficients;
%let lag=12;*lags for Newey-West t-stat;
proc model data=FBny;  by model code name;
 parms a; exogenous coef1 ;
 instruments / intonly;
 coef1 =a;
 fit coef1 / gmm kernel=(bart, %eval(&lag+1), 0);
 ods output parameterestimates=param1  fitstatistics=fitresult
 OutputStatistics=residual; quit;

** 4. output into column table;
data param1; set param1;
 tvalue2=put(tvalue,7.2); if probt<0.1 then p='*  ';
 if probt<0.05 then p='** '; if probt<0.01 then p='***';
 T=compress('('||tvalue2||')'); PARAM=compress(put(estimate,7.2)||p); run;

data param1a; set param1; keep model code name coef _name_;  _name_='PARAM'; coef=PARAM;
run;
data param1b; set param1; keep model code name coef _name_; _name_='T'; coef=T; run;
data param2; set param1a param1b;run;
proc sort data=param2; by  code _name_ model;run;
proc transpose data=param2 out=param3;  by code name _name_; id model; var coef; run;
data param3; set param3; if _name_='T' then do; code=. ;name=.;end;run;

** 5. find the range of periods and obs used;
proc sort data=fb out=fb3; by _model_; run;
data fb3; set fb3; keep _model_ date num; num = _edf_+_p_;  rename _model_=model; run;
proc sql; create table num(where=(model='MODEL1')) as select
 model, min(date) as start, max(date) as end, count(date) as range,
 sum(num) as obs from fb3 group by model;quit;
proc transpose data=num out=num; by model; var start -- obs; run;
data num; set num; rename _name_=name; MODEL1=put(col1, 7.0);  drop model col1; run;
data param3; set param3 num; run;
** 6. output as excel file;
proc export   data=param3  dbms=xlsx
outfile="c:\project107_acc_pbfj\result\T2_csmar_FMB_shell_&micro._&wtt..xlsx"
```

```
replace; run;
```