Three Essays on Asset Pricing with Firm Characteristics

Jiaxin Peng

Doctor of Philosophy

University of York

Economics and Related Studies

June 2022

Abstract

This thesis examines the relationship between expected stock returns and various types of firm characteristics.

Chapter 1 decomposes the return on equity by DuPont analysis into five ratios: Tax Burden, Interest Burden, Margin, Turnover and Leverage. I utilise the portfolio analysis and Fama-MacBeth regressions to evaluate the predictive performance on future stock returns. The results show that all popular asset pricing models fail to explain the long-short portfolio returns sorted by leverage. Interestingly, tax burden and turnover are the only two variables that can capture the expected stock returns in Fama-MacBeth regressions. Additionally, turnover is found to have a significant impact when including operating profitability, but not with other profitability proxies.

Chapter 2 segments the proxies of value effects (book-to-market ratio, retained earnings-to-market ratio, and contributed capital-to-market ratio) into past five-year variables and compares the performance at five-year, one-year, and one-month lags. The results show that the retained earnings-to-market ratio is only significant with no lags, while the contributed capital growth is the most significant factor in the regressions with past information. Prior stock returns less than one year can have greater predictive power on future stock returns than the other. Additionally, updating the data monthly can have incremental improvements on the significance of price-scaled variables.

Chapter 3 modifies Kelly et al. (2019) Instrumented Principal Component Analysis (IPCA) to a static version using a large-dimensional firm characteristic matrix to construct managed portfolio returns. I evaluate the relationship between characteristic-adjusted managed portfolios from IPCA and 126 macroeconomic variables using the three-pass method by Giglio and Xiu (2021). The outcomes reveal two significant macro factors associated with *Price Index* and the *Market* factor, whereas mimicking portfolio analysis shows that most of significant macroeconomic variables are related to *Industry Production* and *Housing*. The result also indicates that rank transformation with a 0.5 shift provides better estimation.

Contents

List of Figures

List of Tables

Acknowledgements

This thesis could not be completed without the support of many people. First of all, I would like to express my deepest gratitude to my supervisor, Dr. Adam Golinski, who has provided me with invaluable suggestions and feedback during my four-year studies. I would also want to thank my TAP members, Professor Peter Smith and Professor Takashi Yamagata, who give important comments on my way to progress. I would like to thank everyone I have met at the university, I have learnt a lot about how to find balance between study and life.

I wish to extend my sincere gratitude to my entire family. I would like to thank my wife, Hang Gao, for taking care of my daily life and helping me to relax when I felt anxious. Without her dedication, I would not have been able to begin my journey at the University of York. I would also want to thank my parents and sister, Naishun Peng, Jun Li and Mingyu Peng, who deliver support and encouragement from thousands of miles away. Furthermore, I would like to thank my mother-in-law, Ye Tian, who came to the UK and assisted us in caring for my daughter over the last six months before submission.

Lastly, I would really want to thank and dedicate my thesis to my daughter Xiyue Peng, who has taught me the essence of life and who I strive to become.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this University, or any elsewhere. All sources are acknowledged as references.

Chapter 1

DuPont decomposition with asset pricing model

1.1 Introduction

Fama and French (2006) assume and confirm the relation between the expected profitability and expected stock returns via the modified Dividend Discount Model (DDM) at a given book-to-market ratio and expected investment. Their findings indicate that simple lagged earnings are a good predictor of expected profitability. Then, Fama and French (2015) expand on the concept by using operating profitability as a proxy for the profitability factor to construct the Fama-French five-factor model (FF5). Theoretically, expected profitability can have a significant positive impact on expected stock returns. However, they argue that it is not clear that predictive power of profitability is due to rational risk compensation or an irrational mispricing reaction. Profitability can vary depending on the book-to-market ratio (BTM) and investment, according to valuation theory.

Besides, Hou et al. (2015) (HXZ) introduce the profitability factor through the investment-based asset pricing, Tobin's q-factor theory. They explain how their five-factor model (HXZQ) with profitability and investment factors can have significant explanatory power in measuring the expected stock returns and digesting the anomaly variables, which cannot be captured by Fama and French (1993) three-factor model (FF3) and Carhart (1997) four-factor model (FFC4). They also highlight that the positive relationship between expected stock returns and profitability under given investments and the discount rate is due to the mispricing reaction. According to Campbell (2017), the reason for the increase in expected stock returns may differ in the q-theory. However, given other variables, the relationship between expected stock returns and profitability is clear.

The significance of return on equity (ROE) in empirical analysis motivates this chapter. The movement of excess stock returns are attributed to cash flow or discount rate volatility, according to Campbell and Shiller (1988). Vuolteenaho (2002) estabilishes a model between BTM, earnings and expected stock returns. He finds ROE is a good proxy for cash-flow based fundamentals, and firm-level stock returns are driven by cash-flow surprises. HXZ also introduce ROE as the proxy for profitability factor in HXZQ.

However, Novy-Marx (2013) points out that the bottom line income can be contaminated by indirect expenses such as the sale force costs, which are unrelated to true economic activities. This chapter is motivated to decompose the ROE and examine the individual contribution of each component on predicting stock returns from the standpoint of a clean measure of profitability.

The most common methodology related to ROE is DuPont analysis, which decomposes the ROE into several parts to investigate a specific fundamental factor's contribution to firm's profitability and operating status. Scholars focus mainly on *profit margin*, *asset turnover* and *leverage*. For example, Novy-Marx

(2013), and Ball et al. (2015) apply the DuPont decomposition to gross profitability (GP) and operating profitability (OP) to determine whether the components can have the ability to explain the expected stock returns. These individual ratios may have significant explanatory power in predicting the future stock returns, and they can be interpreted by different asset pricing channels. However, they only consider the profit deflated by assets rather than the total equity, which means that their variable cannot account the information such interest rates or taxes. According to their findings, the decomposed parts' performance cannot outperform the general profitability factors. This chapter expands the DuPont decomposition from two parts to five ratios: *tax burden, interest burden, profit margin, asset turnover* and *financial leverage*.[1]

From the previous studies, the components of ROE can have the ability to explain the expected stock returns. They can be interpreted by different asset pricing channels. Most studies focus on the individual rate, such as tax rate or interest rate, which means that DuPont disaggregated ratios will have negative effects on the expected stock returns compared to the individual variable.

In the risk premium channel, Chaudhry (2021) and Kim et al. (2011) find that the aggressive tax planning (less tax payment) can increase the firm's idiosyncratic risk or crach risk. According to the negative correlation between idosyncratic volatility and expected stock returns (Ang et al., 2006), the investors will have lower expected stock returns if the firm has a higher *tax burden*. Based on the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965), investors can earn higher returns by investing in the fixed income securities as interest rates rise, resulting in a less risky portfolio. This can lead to a drop in stock demand and a decrease in Stock prices. Expected stock returns will also fall in relative terms. As a result, stock returns should have a positive correla-

[1] *Tax burden* and *interest burden* are the inverse of the tax or interest rate. For instance, if the tax rate is high, the *tax burden* will be low because of $\frac{\text{Net Income}}{\text{Pretax Income}}$.

tion with the *interest burden*. Furthermore, Eriotis et al. (2002) and Muradoğlu and Sivaprasad (2012) argue that the negative correlation between *financial leverage* and expected stock returns can be interpreted as a compensation for the risk of financial distress and the loss of investment opportunity.

Through mispricing channel, Fairfield and Yohn (2001) mention that *profit margin* and *asset turnover* can be used as proxies of profitability to predict the expected stock returns. Patin et al. (2020) discover that the *asset turnover* can measure the firm's operating efficiency and adjust the future value estimation to increase/decrease the stock price. Haugen and Baker (1996) measure the *profit margin* as a proxy for future growth potential, which means that the stock price will go up if the firm has a reasonable perspective expectation.

The main target of this chapter is to answer to what extent the components of ROE can provide incremental explanatory power in capturing the expected stock returns. I do some empirical analysis and have some contributions compared to previous studies. I examine the significance of five ratios via two methodologies, portfolio sorting analysis proposed by Fama and French (1993) and Fama and MacBeth (1973) two-step cross-sectional regressions (FM regression).

First, this chapter find the significant *financial leverage*, where Novy-Marx (2013) ignore it as the noise of profitability. I construct the portfolios univariate sorted by the DuPont decomposed ratios and examine whether the asset pricing models can capture the equal-weighted (EW) and value-weighted (VW) average excess returns on the difference between large percentile and small percentile sorted portfolios. The traditional FF3 and FFC4 show different results from FF5 and HXZQ that the former two models cannot explain the tax burden and interest burden anomaly. While Fama-French five-factor model (FF5) and Hou et al. (2015) q-theory model (HXZQ) leave significant unexplained parts when pricing the portfolios sorted by margin and turnover. However, leverage is the "anomaly" that all underlying asset pricing models cannot fully capture.

I also have a contribution that income based components, such as *tax burden* and *asset turnover* can work well on the empirical analysis for firm-level All-but-Microcaps data, which is consistent with Vuolteenaho (2002). In the second part, I test the significance of five ratios via FM regressions. Ratios can barely maintain the capacity to predict the future stock returns when including all data samples or only Microcap firms whose market equity is under the 20th percentile of the New York Stock Exchange (NYSE). If the Microcaps are booted out of the regression samples, I find that *tax burden* and *asset turnover* have significant explanatory power in predicting the future stock returns in univariate regression with control variables.[2] When turning attention from single-ratio regressions to multi-ratio regressions, ratios can interact with each other since the correlation between some variables is strong, which results in reduced significance, especially encapsulating all decomposed ratios in one regression.

Finally, I do the analysis a step further to examine the DuPont decomposed ratios with robust proxies of profitability, built on Novy-Marx (2013), Ball et al. (2015) and Ball et al. (2016). These three studies derive the GP, OP and cashed-based operating profitability (CBOP) related to firms' profit, and these proxies of profitability have a tremendous effect on expected stock returns. From the outcomes, the performance is consistent with the previous part, which shows that five ratios cannot survive compared with GP or CBOP. Only variables related to earnings can have a marginal explanatory power in predicting the future stock returns for firm-level analysis.

In the regression with OP, turnover can significantly predict the future stock

[2] I follow Novy-Marx (2013) to include four control variables in all FM regressions. They are log BTM (*btm*), size factor (logarithm of market equity (ME), *size*), prior one-month return ($r_{1,1}$) and prior one-year return skip last month ($r_{12,2}$, *momentum*).

returns, while other ratios will lose the explanatory power, absorbed by the robust proxy of profitability. The suprised performance may attribute to the structure of profitability since *asset turnover* has the same denominator as OP even the sale is not a preferred indicator of firms' operating status. Ball et al. (2015) find that using total assets is a better way to deflate the profit than BE or ME. Also, *asset turnover* have the least correlation with OP compared with other two measures of profitability, which means that only a small proportion of the information of *asset turnover* is absorbed by OP.

Apart from the literature mentioned above, the previous studies related to this chapter can be divided into three areas. The first area is focusing on profitability. Ball and Brown (1968) introduce the positive relationship between the stock returns and earnings, which can be known as a proxy of profitability. Sloan (1996) examines the relationship between the stock prices and future earnings, which is highly dependent on the size of accruals and cash flows. The high cash flow and low accruals can generate positive excess stock returns. Haugen and Baker (1996) find a positive relationship between the stock returns and profitability under a given book-to-market level. Fama and French (2008b) evidence that abnormal high stock returns can be explained by high profitability, while the relationship between unprofitable firms and low stock returns is not robust.

Novy-Marx (2013) proposes a proxy of profitability from Fama and French (2006). He uses the gross profit deflated by total assets to capture the expected stock returns with control variables. He claims that gross profitability has a similar power as earnings to predict the expected stock returns. Ball et al. (2015) develop a new measure, operating profitability (OP), based on Novy-Marx (2013), which can better deliver the past and current operating information related to expected stock returns compared to gross profitability. This alternative profitability has better performance than gross profitability in capturing the expected stock returns. Ball et al. (2016) exclude the accruals, which measure the non-cash earnings in the financial statement from the operating profit. They assert that cash-based operating profitability (CBOP) can easily outperform gross and operating profit in predicting stock returns. They also find evidence that this variable predicts the future stock returns even ten years ahead.

There are also some other scholars who study the DuPont analysis. Previous studies related to DuPont decomposition focus on *margin, turnover* and *leverage*, which are highly correlated with firms' financial operating activities. Allozi and Obeidat (2016) estimate the relationship between *profit margin* and stock returns from the Amman Stock Exchange, and they find that the *gross profit margin* has a significant relationship with stock returns. Eisfeldt and Papanikolaou (2013) test the stock returns by using the operating expenses, which is the difference between operating profit and gross profit. There is evidence that the operating expenses will affect the stock return dramatically. Beccalli et al. (2006) and Liadaki and Gaganis (2010) also find that bank operating efficiency can have postivie correlation with stock returns. Soliman (2008) applies DuPont analysis on return on net operating assets, and finds that change in asset turnover can have explanatory power on predicting the profitability facto which in turn affects the stock returns. Chang et al. (2014) test the performance of the DuPont components in the healthcare industry, and they find that the *margin* and *asset turnover* have different performance than all other industries.

The third area is related to the individual ratio related to DuPont analysis. In the past few decades, researchers concentrate on the last three parts of DuPont Analysis (margin, turnover and leverage). Besides the above three ratios I have mentioned, there is no study containing the tax burden and interest burden with other decomposed ratios. Although there are some studies related to the tax rate or interest rate, which can also provide evidence for the tax burden and interest burden of the DuPont decomposition in this chapter.

Lang and Shackelford (2000) find that a reduced tax rate will raise the performance of stock value, and also increase the investors' investment. It means that the capital gains tax rate has a negative relationship with expected stock returns. However, Dhaliwal et al. (2005) find a positive relationship between firm-level tax rates and the required stock returns. Then Dhaliwal et al. (2007) extend the previous study and find that the stock returns positively comove with the tax rate of investor-level when they apply the Tax Act 2003 in the US. Sialm (2009) examines the personal taxation of securities and the valuation of stocks. He finds that the firm, which has a higher tax burden, will compensate the investors who require paying taxes. The action of compensation will lead to higher pre-tax returns, and firms have a significant tax burden if they require to pay out earnings as dividends with high dividend taxation. Sikes and Verrecchia (2012) and Hail et al. (2017) suggest a negative relationship between stock returns and tax rates in three different circumstances: high systematic risk, high market risk premium and a low risk-free rate of return of the market. They also test the return with three tax rate events in the US and find that tax rate changes can have opposite implications than policymakers imagine.

Since the 1980s, many studies focus on the relationship between stock returns and interest rates. Fama (1981) uses the inflation rate as the proxy for short-term interest rate, and he suggests that the stock returns should have a negative relation with the short-term interest rate by using the proxy method. In contrast, the discount rate model can reflect the negative relationship between stock returns and long-term interest rates. The increasing interest rate, similar to the discount rate, will decrease the present values of stock and returns. Some other scholars recently found a negative relationship between these two variables. Spyrou (2001) find that inflation and stock returns are negatively related, but only up to 1995, after which the link becomes insignificant. Alam and Uddin (2009) applies the empirical regression for 18 different developed and developing countries with monthly data, and they give the evidence to support that neither stock market follows a random walk. The interest rate negatively correlates with stock prices, and changes in interest rate also negatively correlate with stock returns.

Besides, Swaminathan and Weintrop (1991) documents that revenue is a proxy of incremental earnings to predict the stock returns. Novy-Marx (2013) proposes asset turnover as an indicator of operating efficiency. He tests the asset turnover and margin with the profitability factor, gross profit to assets. He concludes that asset turnover and stock returns have a positive correlation, but the correlation is insignificant. The result does not show an incremental predictive power of turnover with the gross profitability in the regression. Patin et al. (2020) reveal postive correlation between *asset turnover* and the expected stock returns by Generalised Method of Moments (GMM) and Dynamic Ordinary Least Square (DOLS).

Modigliani and Miller (1958) introduce that the *financial leverage* can be a proxy of capital structure to measure the stock returns. Bhandari (1988) notes that the debt-to-equity ratio has a positive relationship with expected stock returns. He also illustrates that the ratio is not sensitive to the beta of the market portfolio factor. Dimitrov and Jain (2008) assume that the changes in leverage relate to the stock returns. They show that financial leverage has a significantly negative effect on the same period stock returns, while this negative relation also exists between financial leverage information and future stock returns. Sivaprasad and Muradoglu (2009) show a negative correlation between financial leverage and expected stock returns except the utility industry. George and Hwang (2010) also find a negative relation between stock returns and financial leverage. They mention that the financial distress risk will be priced in their model, and the leverage ratio will catch the change of firms' reaction to the distress risk. The expected stock return will be affected by this negative relation. However, Allozi and Obeidat (2016) tests eight financial ratios derived from profitability and leverage measures. The leverage does not have a significant relationship with stock returns, while the factor derived from the profitability measure can predict the stock returns significantly.

The rest of this chapter is organised as follows. Section 1.2 describes the model, including the construction of DuPont decomposition. Section 1.3 explains the data used in this chapter and how to adjust the outliers. Section 1.4.1 displays the summary statistics for the decomposed ratios and the proxies of profitability. Section 1.4.2 provides results for the portfolio sorts to analyse whether the asset pricing models can measure the excess returns to the portfolios sorted by DuPont ratios and the single-ratio FM regressions. Then, 1.4.3 explain the results of FM regressions with different decomposed ratios. Section 1.4.4 compares the performance of DuPont decomposed ratios with GP, OP and CBOP. Then section 1.5 concludes the main finding and interprets the possible further study.

1.2 Model Construction

The fundamental idea of this chapter starts from the ROE or net income (NI). Following the DuPont analysis, the ROE can be decomposed as follows:

$$\begin{aligned} ROE = \frac{Net\ Income}{Equity} =& \frac{Net\ Income}{Pretax\ Income} \times \frac{Pretax\ Income}{EBIT} \times \frac{EBIT}{Revenue} \\ & \times \frac{Revenue}{Total\ Asset} \times \frac{Total\ Asset}{Total\ Equity} \end{aligned} \tag{1.1}$$

where five parts can represent different financial status of the specific firms. These five ratios can be denoted with different meanings:

- *Net Income/Pretax Income* is the Tax Burden;
- *Pretax Income/EBIT* is the Interest Burden;
- *EBIT/Revenue* is the Margin;
- *Revenue/Assets* is the Turnover;
- *Assets/Equities* is the Leverage.

Novy-Marx (2013), Ball et al. (2015) and Ball et al. (2016) apply the DuPont decomposition to construct different explanatory variables to predict the expected return. They only use two parts of DuPont decomposition:

$$\frac{Gross\ Profit}{Assets} = \frac{Sales}{Assets} \times \frac{Gross\ Profit}{Sales} \tag{1.2}$$

They names the first part of right-hand side as ***asset turnover*** (AT) and the second part ***gross margin*** (GM). The second part is slightly different from conventional DuPont decomposition, but it can be still known as the derivative. Then, they run the regressions of expected stock returns on the proxy of profitability with the GM and AT to examine the significance of their new factors.

Following Novy-Marx (2013), I establish the regression with five ratios as the individual explanatory variables to measure the expected stock returns. The cross-sectional model can be constructed as:

$$r_{i,t+1} = \alpha_{i,t} + \beta_{1,t} tax_{i,t} + \beta_{2,t} int_{i,t} + \beta_{3,t} mar_{i,t} + \beta_{4,t} turn_{i,t} + \beta_{5,t} leve_{i,t} + \epsilon_{i,t} \tag{1.3}$$

where $r_{i,t+1}$ is the one-month ahead stock return on firm i at the beginning of period *t+1*, $\beta_{i,t}$ represents the coresponding coefficient between stock returns and the specific DuPont decomposed independent variables, $tax_{i,t}$, $int_{i,t}$, etc, in the beginning of the period t.

Additionally, I also include four control variables related to the systematic risk, which have been used for the estimation by Novy-Marx (2013), Ball et al. (2015) and Ball et al. (2016). They are the natural logarithm of book-to-market ratio ($log(BE/ME)$, BTM), natural logarithm of market equity ($log(ME)$, size), prior one-month return ($r_{1,1}$) and prior annual return skip the last one month ($r_{12,2}$).

After adding the control variables, the estimation regression can be written as:

$$\begin{aligned} r_{i,t+1} = {} & \alpha_{i,t} + \beta_{1,t} tax_{i,t} + \beta_{2,t} int_{i,t} + \beta_{3,t} mar_{i,t} + \beta_{4,t} turn_{i,t} + \beta_{5,t} leve_{i,t} \\ & + \beta_{6,t} btm_{i,t} + \beta_{7,t} size_{i,t} + \beta_{8,t} r_{1,1} + \beta_{9,t} r_{12,2} + \epsilon_{i,t} \end{aligned} \tag{1.4}$$

Based on the above equation, I apply Fama and MacBeth (1973) two-step cross-sectional regression to check whether individual ratios can have significant explanatory power in predicting the expected stock returns. Also, I test the joint performance between these ratios in one regression.

Besides, I sort the individual firms' returns to build the portfolios to investigate whether the factor model can capture the sorted portfolio returns without anomaly via a different aspect.[3]

[3]The full details are described in following section.

1.3 Data

1.3.1 Data collection

This chapter obtains the samples from two databases based on the different properties of the data items. The first group are annual accounting-based information from Compustat, for example, net income (NI), common shareholder's equity (CEQ) and total assets (AT). Another group includes price-related variables from the Centre for Research in Security Prices (CRSP), such as stock returns (RET) and total market capitalisation (TCAP), which update month-by-month. All variables cover the period from January 1962 to December 2020. Because of the calculation of past information, the total sample size will only cover from July 1964 to June 2019 to keep the availability of variables. In addition, this chapter requires matching between CRSP and Compustat to ensure the consistency of the two databases.

Following previous research, the data collection process has several restrictions. First, Novy-Marx (2013) mentions that the majority of firms are assumed to publish their fiscal year t statement at the end of December of the calendar year t, while some firms report the fiscal year t statements at the end of June in the calendar year t. This chapter also resorts to the same measure that annual accounting data has at least six-month lags with monthly returns, and all firms have 12 months of available returns for a year. Therefore, the annual accounting data will be collected at the end of June of each calendar year t for fiscal year t and available from July in year t to June in year $t+1$ matching the monthly updated CRSP data items.

Second, Fama and French (1993), Novy-Marx (2013), Ball et al. (2015) and Ball et al. (2016) argue that the financial firms are unstable and the structure of their statement is different with other industries. Financial firms' information will bias the regression results. So firms with the Standard Industrial Classification code (SIC) between 6000 and 6799 are ignored from the collection.[4]

Finally, all accounting-based variables should be available for at least one year without missing data. All firms satisfy the condition that they have available monthly returns, one-lag month returns and one-year return momentum.

In addition, I follow Bali et al. (2016) and Novy-Marx (2013) to build the control variables. BTM is constructed at the end of December each year when I divide the shareholder's total equity (TEQ) of fiscal year t by the market equity at the end of December in calendar year t. The data will be available for the next July in year $t+1$ to June in year $t+2$. Ball et al. (2020) use the CEQ as the reported book equity to perform the empirical analysis. I use the common shareholder's equity (CEQ) to substitute if TEQ is missing. *size* is computed via the market capitalisation (CAP) at the end of June for each calendar year and remains unchanged in the next 12 months from July in the same year to June of the next year by Fama and French (1992). Following Bali et al. (2016), if the regression establishes at the beginning of the period $t+1$, the left-hand side stock returns will cover the whole month from the beginning of the period $t+1$ till the end of the period $t+1$, the prior one-month return will cover the whole month t. I compute the momentum at the beginning of month $t+1$ to cover the prior 12-month stock returns, which cover the compounded stock returns from the beginning of $t-11$ to the end of $t-1$.

[4]In this chapter, I use the latest SIC rather than the historic SIC to filter the firms that belonged to the financial sector.

1.3.2 Data adjustment

This chapter adjusts the data to eliminate the missing information and outliers in three steps.

The first step is to diminish missing stock returns. CRSP gives the description for some extreme stock returns, which is *-66, -77, -88* or *-99*.[5] If CRSP records the above values, the stock returns will be adjusted to missing, which will be deleted from the data sample in this chapter. Moreover, Shumway (1997) points out that original monthly returns without delisting will cause bias in the regression analysis. I add the delisting return for those firms delisted from the market in the database and follow Ball et al. (2015) to filter the delisting information. If delisting returns are available, they will directly replace the last monthly return. If the delisting return is missing, it will be marked as *-0.3(30%)* with specific delisting code (*500, 520, 551:573, 574, 580, 584*). Otherwise, the delisting return should be set as *-1(-100%)*. Furthermore, all explanatory variables, following Novy-Marx (2013), will be trimmed at *1%* and *99%* to decrease the bias from outliers since the outliers of accounting-based information will be extremely large or small.

Subsequently, this chapter adjusts some variables with specific calculations. The first one is the book equity. In DuPont decomposition, the original denominator of leverage is the shareholder's equity. Ball et al. (2015) point out that the book value of equity can deliver the information more accurately. Moreover, Novy-Marx (2013), and Bali et al. (2016) mention the process of calculating the book value of equity, which should be calculated based on the shareholder's equity (SEQ from Compustat), subtract tax effects as well, and add deferred taxes (TXDB) and investment tax credit (ITCB). Finally, remove the influence of

[5]Most time, the extreme values will be found at the beginning published month or the delisting month.

the value of preferred stocks. Book value of preferred stocks (BVPS) equals the redemption value (PSTKRV), the liquidating value (PSTKL) or the par value (PSTK) if it is available. If not, it will be zero. If ITCB is missing, I calculate the book value by the sum of common equity (CEQ) plus PSTKV by following Heath and Mace (2020). The missing data of SEQ and TXDB can cause a failure to compute the book value of equity. If so, I use the difference between the total asset and total liability to represent the book equity. If the firm has non-positive book equity, it will be removed from the sample.

I also compare the performance of decomposed ratios with the gross profitability and the operating profitability generated by Ball et al. (2015). Novy-Marx (2013) defines the gross profitability, which is simply the gross profit (GP) divided by the one-year lagged total asset (ATLAG1). Ball et al. (2015) derive the operating profitability that is equal to the gross profit minus sell, general, and administrative expenses (XSGA), excluding research and development expenses (XRD), and deflated by the same, the one-year lagged total asset. However, the availability of XRD in Compustat is significantly low (about *39%*), which is attributable to the indistinguishability of XRD from other operating expenses. The unobserved XRD will lead to unavailable operating profitability. Deleting the unavailable data directly from the data sample will decrease the number of observations significantly. One solution is to adjust the missing XRD to zero (Koh and Reeb (2015)), suggesting that the missing XRD should be set as zero. Also, missing XSGA define as zero following the rule.

1.4 Empirical Analysis

In this section, I show the results of the portfolio sorting analysis and FM regressions to examine the performance of the DuPont decomposed ratios and whether they can have the capacity to capture the expected stock returns.

1.4.1 Summary statistics

Table 1.1 displays the summary statistics of all decomposed ratios and four control variables. I collect the cross-sectional information for each period and then take an average to find the summary statistics. The summary statistics cover July 1964 to June 2019, with the average number of observations around 2644. Intuitively, the average values of *profit margin, ROE and BTM* are negative. At the same time, the other variables all have a positive mean value.[6] The median margin and ROE are both positive, 0.065 and 0.085, while the minimum values are -2462 and -328, respectively, and first percentile values are about -21.9 and -2.865. The negative average value is caused by the extreme outliers, which also interpret the enormous skewness value for margin and ROE (-26.34 and -19.68). Some firms have small, positive revenues and a large amount of operating cost, which cause relative significant earnings divided by the small revenues to have considerably extreme values. The average tax burden is about 72.9%, which means that about 27% of gross income should be paid for the tax expenses.[7] The average interest burden is around 93.7%, which only takes a minority of the income for most firms. Besides, the asset turnover is the most stable in five decomposed ratios with slight standard deviation, skewness and kurtosis. However, five decomposed ratios have a large kurtosis except for asset turnover, which means the sample distribution cannot follow the normal distribution.

[6]BTM is negative due to the natural logarithm calculation, but the average book-to-market ratio is still positive and approximately 0.61.

[7]I also calculate another version of the tax burden (taxmii) consisting of net income (NI) and noncontrolling interest (MII), which is mentioned by Welc (2017). He argues that it can give more precise net profit results for a firm. The effect increases about 1.7% of the tax burden on average.

Table 1.1: Summary statistics of DuPont ratios decomposition with control variables

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Explanatory Variables															
tax	0.729	3.674	3.132	1263.308	-83.180	-0.357	0.446	0.572	0.626	0.830	1.120	2.304	122.762	2644.880	35.745
taxmii	0.746	3.778	2.467	1172.976	-78.830	-0.274	0.475	0.575	0.632	0.842	1.116	2.260	138.496	2381.442	29.289
int	0.937	12.550	-1.093	1158.839	-365.161	-3.203	0.092	0.775	0.945	1.049	1.669	4.916	387.973	2644.880	124.048
mar	-2.074	51.838	-26.342	1247.875	-2462.496	-21.893	-1.131	-0.001	0.065	0.124	0.264	0.430	82.050	2644.880	624.138
turn	1.292	0.944	3.963	71.201	0.009	0.074	0.246	0.737	1.138	1.601	2.857	4.748	14.043	2644.880	0.000
leve	3.583	47.447	31.023	1308.054	0.997	1.071	1.157	1.455	1.877	2.571	5.521	16.391	2510.500	2644.880	0.000
roe	-0.166	6.858	-19.680	1121.694	-328.891	-2.865	-0.671	-0.016	0.085	0.146	0.279	0.558	42.418	2644.880	685.141
be	4.201	1.903	0.235	3.025	-2.641	0.280	1.247	2.865	4.120	5.428	7.526	8.893	10.898	2644.880	37.598
gm	0.475	2.066	9.291	455.074	-3.088	-0.334	0.030	0.228	0.389	0.599	1.065	1.601	105.682	2644.880	102.409
opm	0.163	0.535	3.020	508.010	-10.403	-0.402	-0.096	0.088	0.158	0.238	0.430	0.686	14.725	2644.880	302.323
Control variables															
$r_{1,1}$	0.013	0.149	2.765	47.681	-0.596	-0.302	-0.182	-0.062	0.002	0.071	0.236	0.479	2.087	2644.880	1247.652
$r_{12,2}$	0.147	0.579	3.764	52.074	-0.859	-0.671	-0.487	-0.170	0.059	0.329	1.046	2.165	8.974	2644.880	1194.512
size	4.774	1.959	0.329	2.855	-0.281	0.930	1.773	3.350	4.656	6.075	8.187	9.697	11.581	2644.880	11.971
btm	-0.516	0.864	-0.685	5.518	-6.030	-2.980	-1.996	-0.998	-0.437	0.044	0.727	1.347	2.984	2644.880	1871.479
r	0.013	0.150	2.830	50.454	-0.615	-0.305	-0.183	-0.062	0.002	0.071	0.237	0.481	2.138	2644.880	1244.552

Note: This table presents statistics for variables used in this chapter. I include five DuPont decomposed ratios and four control variables. The five explanatory variables are collected and constructed from Compustat. They are: tax burden (Net Income/Pretax Income), interest burden (Pretax Income/EBIT), margin (EBIT/Revenue), turnover (Revenue/Total Assets) and leverage (Total Assets/Total Equity). Four control variables used in our analysis are defined as follows: natural logarithm of book-to-market ratio (*btm*, $log(BE/ME)$), natural logarithm of market equity (*size*,$log(ME)$), prior one year skip last month ($r_{12,2}$), prior one month ($r_{1,1}$) performance and current monthly return (r). The sample period starts from July 1964 and ends in June 2019.

For most ratios, minimum and maximum values are significantly different from the value of 1% and 99%. I winsorize all explanatory variables, including control variables at 1% and 99%.[8]

Table 1.2 shows the summary statistics after winsorization. The mean value of the margin increases significantly from -2.074 to -0.377, and the average ROE increase to near zero (-0.023). Instead, tax burden with or without noncontrolling interest and interest burden slightly decrease, and financial leverage sharply drops from 3.583 to around 2.463. The average cross-sectional mean value of turnover keeps stable at about 1.27, benefiting from the small skewness and kurtosis. The winsorisation does not significantly affect two reference proxies of profitability, GP and OP, and control variables.

In Table 1.3, the above-diagonal entries show the average cross-sectional Pearson correlation, and the below-diagonal entries display the average cross-sectional Spearman's ranking correlation between five decomposed ratios and four control variables. In the right upper triangle of Pearson coefficients, five ratios merely have significant correlations while only the tax burden is negatively correlated with margin and turnover, -15.84% and -13.37%. The coefficients of Spearman's correlation in the lower triangle will decrease to -33.92% and -23.66% for margin and turnover, respectively. Notably, the interest burden is highly correlated with leverage (coefficient $= -33.45\%$) while the coefficient of this case is only -8.97% of Pearson correlation. The significant difference between the two types of correlation calculation can be interpreted as that interest burden and leverage do not have a standard normal distribution, which is consistent with summary statistics results.

Besides, ROE has a negative Pearson's correlation with both tax burden (-26.8%) and leverage (-29.78%), while the correlation between margin and ROE

[8] In this chapter, I match the outliers below 1% and 99% with the value of 1 and 99 percentiles rather trimming the outliers.

Table 1.2: Summary statistics of DuPont ratios decomposition with control variables after winsorization

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Explanatory Variables															
tax	0.704	0.309	1.696	11.662	-0.357	-0.349	0.446	0.572	0.626	0.830	1.120	2.294	2.304	2644.880	30.324
taxmii	0.713	0.300	1.752	11.352	-0.274	-0.267	0.475	0.575	0.632	0.842	1.116	2.249	2.260	2381.442	22.117
int	0.910	0.807	-0.153	18.631	-3.203	-3.170	0.092	0.775	0.945	1.049	1.669	4.885	4.916	2644.880	124.012
mar	-0.377	2.593	-4.425	37.969	-21.893	-21.621	-1.131	-0.001	0.065	0.124	0.264	0.429	0.430	2644.880	624.138
turn	1.277	0.832	1.590	6.805	0.074	0.075	0.246	0.737	1.138	1.601	2.857	4.733	4.748	2644.880	0.000
leve	2.463	2.131	4.137	24.144	1.071	1.071	1.157	1.455	1.877	2.571	5.521	16.296	16.391	2644.880	0.000
roe	-0.023	0.443	-3.439	21.158	-2.865	-2.848	-0.671	-0.016	0.085	0.146	0.279	0.556	0.558	2644.880	685.141
be	4.202	1.867	0.236	2.668	0.280	0.284	1.247	2.865	4.120	5.428	7.526	8.887	8.893	2644.880	31.248
gm	0.442	0.330	0.822	4.839	-0.334	-0.333	0.030	0.228	0.389	0.599	1.065	1.599	1.601	2644.880	100.030
opm	0.162	0.164	-0.029	5.884	-0.402	-0.400	-0.096	0.088	0.158	0.238	0.430	0.685	0.686	2644.880	302.323
Control variables															
$r_{1,1}$	0.011	0.129	0.674	5.107	-0.302	-0.301	-0.182	-0.062	0.002	0.071	0.236	0.477	0.479	2644.880	1247.652
$r_{12,2}$	0.136	0.488	1.313	6.178	-0.671	-0.671	-0.487	-0.170	0.059	0.329	1.046	2.158	2.165	2644.880	1194.512
size	4.772	1.933	0.291	2.617	0.930	0.934	1.773	3.350	4.656	6.075	8.187	9.688	9.697	2644.880	8.415
btm	-0.512	0.825	-0.466	3.411	-2.980	-2.974	-1.996	-0.998	-0.437	0.044	0.727	1.344	1.347	2644.880	1871.479
r	0.013	0.150	2.830	50.454	-0.615	-0.305	-0.183	-0.062	0.002	0.071	0.237	0.481	2.138	2644.880	1244.552

Note: This table presents statistics for variables used in this chapter. I include five DuPont decomposed ratios and four control variables. The five explanatory variables are collected and constructed from Compustat. They are: tax burden (Net Income/Pretax Income), interest burden (Pretax Income/EBIT), margin (EBIT/Revenue), turnover (Revenue/Total Assets) and leverage (Total Assets/Total Equity). Four control variables used in our analysis are defined as follows: natural logarithm of book-to-market ratio (btm, $log(BE/ME)$), natural logarithm of market equity ($size$,$log(ME)$), prior one year skip last month ($r_{12,2}$), prior one month ($r_{1,1}$) performance and current monthly return (r). The sample period starts from July 1964 and ends in June 2019.

Table 1.3: Correlation between DuPont decomposition ratios and control variables

Variables	tax	taxmii	int	mar	turn	leve	roe	*size*	*btm*	$r_{1,1}$	$r_{12,2}$
tax	100.00	96.10	-1.89	-15.84	-13.37	10.13	-26.80	-15.74	-5.08	-0.86	-2.55
taxmii	96.01	100.00	-1.73	-16.09	-14.52	11.58	-27.43	-15.61	-5.24	-0.89	-2.72
int	4.05	3.64	100.00	-0.02	0.73	-8.97	-3.54	1.52	-5.28	-0.14	-0.14
mar	-33.92	-33.62	-2.62	100.00	4.91	-5.09	43.36	21.21	-0.56	1.67	5.53
turn	-23.66	-25.49	2.47	-13.13	100.00	2.30	13.02	-12.43	5.28	1.26	4.42
leve	2.12	4.53	-33.45	-12.32	9.96	100.00	-29.78	-6.79	-20.55	-0.68	-0.35
roe	-34.19	-34.84	2.52	74.06	25.24	-2.82	100.00	27.86	3.30	2.12	7.24
size	-17.61	-16.60	-0.58	45.10	-13.99	0.19	38.71	100.00	-27.90	0.73	9.51
btm	-4.69	-4.41	-15.67	-20.90	4.79	-0.13	-31.84	-29.82	100.00	2.40	1.23
$r_{1,1}$	-3.18	-3.22	-0.82	4.54	1.89	-0.54	4.56	5.09	1.89	100.00	1.59
$r_{12,2}$	-7.45	-7.55	-1.93	11.94	5.55	0.02	13.39	16.09	2.04	3.17	100.00

Note: This table presents Pearson (Panel 1) and Spearman rank (Panel 2) correlations between the variables used in our analysis. I include five DuPont decomposed ratios and four control variables. The five explanatory variables are collected and constructed from Compustat. They are: tax burden (Net Income/Pretax Income), interest burden (Pretax Income/EBIT), margin (EBIT/Revenue), turnover (Revenue/Total Assets) and leverage (Total Assets/Total Equity). Four control variables used in our analysis are defined as follows: natural logarithm of book-to-market ratio (btm, $log(BE/ME)$), natural logarithm of market equity ($size$,$log(ME)$), prior one year skip last month ($r_{12,2}$) and prior one month ($r_{1,1}$) performance. I also compare the proxies of profitability: gross profitability and operating profitability. The sample period starts from July 1964 and ends in June 2019.

(43.36%) is the strongest across the five ratios, but the Spearmen's ranking correlation of ROE and leverage is only -2.82%, which means that the sequential of leverage is not significant like other ratios.

In addition, *size* is significantly correlated with tax burden (-15%), margin (21%) and turnover (-12%) except the interest burden and leverage. The strong correlation indicates that these ratios can differ significantly between large and small firms. However, leverage linearly comoves with BTM differently (Pearson coefficient -20.55%). Apart from the above, decomposed ratios do not correlate robustly with past stock returns.

1.4.2 Single-variable analysis

This part analyses whether individual DuPont decomposed ratios can capture the expected stock returns with control variables via Fama-MacBeth regressions or the asset pricing factor model can measure the univariate sorting portfolios without unexplained intercept.

1.4.2.1 Portfolio analysis

I first construct the portfolios, univariate sorted by five decomposed ratios and ROE, respectively. The portfolios are sorted from the first quintile (*P1*, 0%-20%, column (1) or (7)) to fifth quintile (*P5*, 80%-100%, column (5) or (11)) and also the excess portfolio returns between the large quintile and small quintile ($P5 - P1$, column (6) or (12)).[9] Columns (1) to (6) represent the equal-weighted (EW) sorted portfolio returns. In contrast, columns (7) to (12) are the summary information of the value-weighted (VW) portfolios sorted by the underlying decomposed DuPont ratio or ROE. The first-row panel reports the excess returns of the quintile portfolios. The list of firms is created at the end of June every year and is rebalanced in the same period next year. Table 1.4 is constructed with all available firms without filtering at the small size firms, which may affect the result significantly.

The critical point of portfolio analysis is to examine whether the asset pricing model can measure the excess portfolio returns without unexplained parts left in the intercept. So I apply four asset pricing models, Fama-French three-factor model (FF3), Carhart four-factor model (FFC4), Fama-French five-factor model (FF5) and Hou et al. (2015) q-theory five-factor model (XHZQ) to test whether the DuPont decomposition can be denoted as the anomaly. The panels in each table repeat the regression estimates of factor loadings:

- FF3: mkt (market factor), smb (small-minus-big), hml (high-minus-low);

[9] Instead of constructing the long-short portfolio with the considering of positive and negative relations, like SMB in FF3 or FF5, I build the long-short portfolio simply based on the difference between the largest and smallest quintiles.

- FFC4: mkt (market factor), smb (small-minus-big), hml (high-minus-low), mom (high-minus-low);
- FF5: mkt (market factor), smb (small-minus-big), hml (high-minus-low), rmw (robust-minus-weak), cma (conservative-minus-aggressive);
- HXZQ: mkt (market factor), me (size factor, high-minus-low), ia (investment factor, high-minus-low), roe (profitability factor, high-minus-low), eg (expected growth factor, high-minus-low).

$$r_{i,t}^{e} = \alpha_i + \beta_{l,i} f_{l,t} + \epsilon_{i,t},\ l = 1, ..., 5 \tag{1.5}$$

where $r_{i,t}^{e}$ is the excess return of stock i at time t, $f_{l,t}$ is the factor return at time t, $\beta_{l,i}$ is the factor loading of stock i on factor l, α_i is included in all factor model to examine the unexplained proportion of the portfolio excess returns.

In Table 1.4 Panel A, the average excess returns of EW quintile portfolios sorted by tax burden do not significantly differ from *P1* to *P5*. In contrast, the average excess return on VW portfolios decreases with the quintile level's increase from 0.64 to 0.46. The downward trend of returns with tax burden is consistent with Chaudhry (2021), indicating the negative correlation between stock returns and tax burden. For anomaly test, FF3 and FF4 can have a good explanation on both EW and VW high-minus-low portfolios, while the intercept is marginally significant under VW FF3. After adding investment and profitability factors, a becomes highly significant in the regressions on FF5 and HXZQ for EW case, indicating that FF5 and HXZQ cannot digest the information included in the tax burden. However, FF5 and HXZQ can still explain tax burden-based long-short portfolio around 45% and 29%, respectively.

Table 1.4 Panel B and Table 1.4 Panel C report the results of interest burden and profit margin. They do not have a significant monotonic trend of quintile portfolio excess returns. For long-short portfolios, FF3 and FF4 can have a sig-

Table 1.4: Summary information of Portfolio analysis single sorted

Panel A: Sorted by tax burden

Model	Coefficient	P1	P2	P3	P4	P5	P5_P1	P1	P2	P3	P4	P5	P5_P1
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Excess return	avg rt	0.88	0.88	0.84	0.82	0.93	0.05	0.64	0.66	0.59	0.49	0.46	-0.18
	t-stats	(3.39)	(3.61)	(3.42)	(2.61)	(2.79)	(0.34)	(3.16)	(3.43)	(3.1)	(2.2)	(1.72)	(-1.23)
FF3	α	0.07	0.10	0.08	0.04	0.12	0.05	0.02	0.09	0.05	-0.10	-0.22	-0.25
		(1.08)	(1.56)	(1.38)	(0.32)	(0.89)	(0.35)	(0.35)	(1.45)	(1.26)	(-1.42)	(-1.85)	(-1.83)
	mkt	1.01	0.99	1.01	1.04	1.05	0.04	1.04	0.99	1.01	1.11	1.20	0.16
		(39.52)	(40.25)	(44.02)	(27.11)	(27.33)	(1.03)	(55.37)	(50.3)	(76.25)	(45.2)	(22.39)	(2.84)
	smb	0.83	0.74	0.73	1.07	1.11	0.28	0.09	0.06	-0.02	0.06	0.21	0.12
		(8.79)	(8.29)	(9.77)	(16.53)	(16.01)	(3.19)	(1.78)	(0.99)	(-0.66)	(1.73)	(3.2)	(2.2)
	hml	0.34	0.33	0.26	0.04	0.11	-0.23	0.11	0.07	-0.03	-0.09	-0.04	-0.15
		(5.4)	(5.24)	(4.79)	(0.56)	(1.17)	(-2.18)	(2.58)	(1.37)	(-1.01)	(-1.78)	(-0.31)	(-1.07)
	adj-R^2	0.92	0.93	0.94	0.82	0.81	0.16	0.92	0.92	0.95	0.90	0.78	0.10
FF4	α	0.22	0.24	0.22	0.26	0.33	0.11	0.07	0.18	0.15	0.05	0.03	-0.04
		(3.03)	(3.7)	(3.9)	(1.72)	(2.08)	(0.8)	(1.14)	(2.91)	(3.76)	(0.78)	(0.26)	(-0.35)
	mkt	0.98	0.96	0.98	0.99	1.00	0.02	1.03	0.97	0.98	1.07	1.15	0.11
		(44.69)	(47.92)	(57.49)	(29.49)	(28.01)	(0.68)	(55.89)	(59.38)	(87.96)	(55.83)	(28.52)	(2.47)
	smb	0.83	0.73	0.72	1.06	1.11	0.28	0.08	0.06	-0.03	0.06	0.20	0.12
		(10.54)	(9.87)	(12.46)	(14.57)	(17.36)	(2.95)	(1.92)	(1.12)	(-1.1)	(1.74)	(3.54)	(1.83)
	hml	0.28	0.27	0.20	-0.05	0.02	-0.26	0.09	0.03	-0.08	-0.16	-0.15	-0.24
		(5.45)	(5.41)	(4.62)	(-0.57)	(0.2)	(-2.19)	(2.23)	(0.71)	(-3.12)	(-3.33)	(-1.33)	(-1.76)
	mom	-0.17	-0.15	-0.17	-0.25	-0.24	-0.08	-0.06	-0.10	-0.12	-0.17	-0.29	-0.24
		(-4.12)	(-4.13)	(-5.65)	(-3.41)	(-2.76)	(-0.85)	(-2.27)	(-3.59)	(-6.24)	(-4.69)	(-4.28)	(-3.24)
	adj-R^2	0.94	0.95	0.96	0.84	0.83	0.17	0.92	0.93	0.96	0.92	0.81	0.18
FF5	α	0.03	0.02	0.05	0.33	0.39	0.36	-0.05	-0.03	0.00	0.06	0.12	0.17
		(0.36)	(0.21)	(0.77)	(2.53)	(2.65)	(3.28)	(-0.79)	(-0.48)	(-0.04)	(0.89)	(0.95)	(1.32)
	mkt	1.01	1.00	1.01	0.99	1.00	-0.01	1.05	1.01	1.02	1.08	1.13	0.08
		(42.17)	(46.93)	(45.94)	(29.1)	(23.27)	(-0.5)	(62.07)	(56.31)	(74.42)	(53.49)	(29.67)	(2.19)
	smb	0.89	0.82	0.77	0.86	0.92	0.03	0.14	0.15	0.01	-0.06	-0.02	-0.16
		(16.35)	(19.22)	(16.93)	(14.95)	(12.48)	(0.56)	(4.53)	(5.68)	(0.41)	(-1.47)	(-0.19)	(-2.12)
	hml	0.22	0.21	0.15	-0.08	-0.01	-0.24	0.08	0.04	-0.06	-0.07	0.06	-0.02
		(3.53)	(3.79)	(2.63)	(-1.03)	(-0.13)	(-4.02)	(1.86)	(1.1)	(-2.04)	(-1.42)	(0.57)	(-0.24)
	rmw	0.16	0.26	0.11	-0.76	-0.71	-0.87	0.18	0.32	0.11	-0.42	-0.82	-1.00
		(1.75)	(3.31)	(1.44)	(-9.61)	(-5.68)	(-14.69)	(3.79)	(6.08)	(2.46)	(-4.28)	(-4.17)	(-5.98)
	cma	-0.02	0.01	0.00	-0.02	-0.03	-0.01	0.02	0.02	0.06	-0.04	-0.23	-0.25
		(-0.28)	(0.16)	(-0.01)	(-0.16)	(-0.19)	(-0.07)	(0.27)	(0.43)	(1.13)	(-0.48)	(-1.72)	(-2.08)
	adj-R^2	0.94	0.96	0.95	0.86	0.85	0.47	0.93	0.94	0.95	0.93	0.85	0.43
HXZQ	α	0.27	0.23	0.26	0.60	0.67	0.40	0.14	0.10	0.09	0.17	0.38	0.24
		(2.87)	(2.68)	(3.57)	(3.33)	(3.24)	(2.37)	(2.13)	(1.53)	(1.75)	(2.04)	(2.6)	(1.48)
	mkt	97.09	96.00	97.33	97.52	98.39	1.31	102.43	97.99	100.16	107.99	112.76	10.33
		(31.37)	(34.29)	(40.7)	(29.9)	(23.55)	(0.37)	(52.37)	(48.81)	(80.11)	(68.34)	(43.06)	(3.57)
	me	75.69	69.78	65.89	80.52	82.93	7.25	5.44	7.36	-3.69	-5.80	-1.39	-6.83
		(7.6)	(7.74)	(8.84)	(14.32)	(10.82)	(0.81)	(1.1)	(1.47)	(-1.42)	(-1.7)	(-0.16)	(-0.8)
	ia	15.41	18.36	10.06	-20.78	-17.14	-32.55	10.10	9.22	-2.34	-12.43	-15.90	-25.99
		(1.5)	(1.95)	(1.25)	(-1.58)	(-1.02)	(-2.11)	(1.88)	(1.55)	(-0.54)	(-1.78)	(-1.25)	(-1.88)
	roe	-19.58	-8.75	-18.55	-77.37	-80.19	-60.61	-2.83	11.67	-1.79	-31.93	-57.72	-54.89
		(-2.36)	(-1.08)	(-2.68)	(-6.93)	(-5.79)	(-4.63)	(-0.57)	(1.97)	(-0.41)	(-5.53)	(-6.04)	(-5.22)
	eg	-10.17	-10.33	-8.38	-5.62	-1.92	8.25	-12.00	-11.94	-3.30	-4.69	-19.89	-7.89
		(-1.62)	(-1.64)	(-1.52)	(-0.65)	(-0.22)	(0.83)	(-2.69)	(-2.45)	(-0.89)	(-0.83)	(-1.6)	(-0.59)
	adj-R^2	0.92	0.93	0.95	0.87	0.85	0.29	0.92	0.92	0.95	0.92	0.83	0.24

Panel B: Sorted by interest burden

Model	Coefficient	P1	P2	P3	P4	P5	P5_P1	P1	P2	P3	P4	P5	P5_P1
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Excess return	avg rt	0.94	0.90	0.95	0.83	0.85	-0.09	0.60	0.62	0.60	0.65	0.50	-0.11
	t-stats	(3.12)	(3.52)	(3.32)	(3.04)	(2.86)	(-1.03)	(2.46)	(3.12)	(3.08)	(3.3)	(2.16)	(-0.92)
FF3	α	0.04	0.12	0.19	0.08	0.04	0.00	-0.19	0.00	0.04	0.15	-0.09	0.10
		(0.45)	(2.01)	(2.25)	(1.15)	(0.45)	(0)	(-2.5)	(0)	(0.81)	(2.97)	(-1.19)	(0.93)
	mkt	1.05	1.00	1.04	1.03	1.03	-0.02	1.17	1.01	1.05	1.03	1.10	-0.07
		(33.07)	(48.71)	(37.29)	(40.25)	(37.27)	(-0.94)	(46.31)	(50.85)	(73.24)	(73.1)	(38.29)	(-2.18)
	smb	1.01	0.81	0.93	0.90	1.02	0.02	0.28	0.07	-0.01	0.02	0.19	-0.09
		(9.99)	(11.92)	(19.5)	(14.34)	(16.11)	(0.26)	(4.31)	(1.16)	(-0.12)	(0.87)	(4.76)	(-1.42)
	hml	0.46	0.28	0.06	0.07	0.20	-0.26	0.32	0.19	-0.01	-0.22	-0.16	-0.48
		(6.23)	(5.76)	(1.06)	(1.36)	(3.14)	(-5.29)	(5.71)	(4.09)	(-0.27)	(-7.9)	(-2.48)	(-6.36)
	adj-R^2	0.89	0.94	0.90	0.91	0.89	0.15	0.91	0.90	0.94	0.95	0.89	0.24
FF4	α	0.21	0.24	0.39	0.25	0.21	0.01	-0.07	0.07	0.13	0.21	0.06	0.13
		(2.11)	(3.8)	(3.64)	(2.7)	(2.09)	(0.11)	(-0.96)	(1.01)	(2.9)	(3.87)	(0.76)	(1.29)
	mkt	1.01	0.98	0.99	0.99	0.99	-0.02	1.14	0.99	1.02	1.02	1.06	-0.08
		(34.78)	(54.77)	(42.95)	(44.82)	(39.63)	(-1.14)	(50.27)	(54.64)	(79.94)	(73.06)	(51.16)	(-2.92)
	smb	1.00	0.80	0.92	0.89	1.02	0.02	0.28	0.07	-0.01	0.02	0.18	-0.10
		(11.99)	(14.8)	(23.2)	(18.4)	(20.67)	(0.25)	(5.11)	(1.28)	(-0.25)	(0.93)	(4.76)	(-1.37)
	hml	0.38	0.23	-0.03	-0.01	0.12	-0.27	0.27	0.16	-0.05	-0.24	-0.22	-0.49
		(6.04)	(5.46)	(-0.59)	(-0.11)	(1.93)	(-5.59)	(6.19)	(4.1)	(-1.7)	(-9.88)	(-3.42)	(-6.13)
	mom	-0.19	-0.14	-0.23	-0.20	-0.20	-0.01	-0.13	-0.08	-0.11	-0.07	-0.17	-0.04
		(-3.77)	(-4.83)	(-4.43)	(-4.32)	(-4.53)	(-0.25)	(-3.2)	(-2.31)	(-5.27)	(-4.03)	(-3.5)	(-0.56)
	adj-R^2	0.90	0.95	0.93	0.93	0.91	0.15	0.92	0.91	0.95	0.96	0.91	0.24
FF5	α	0.04	0.11	0.33	0.20	0.18	0.14	-0.24	-0.11	-0.01	0.17	0.08	0.32
		(0.35)	(1.59)	(3.16)	(2.22)	(1.75)	(1.9)	(-2.94)	(-1.61)	(-0.24)	(3.36)	(0.99)	(3.31)
	mkt	1.04	1.00	1.01	1.01	1.00	-0.04	1.19	1.03	1.05	1.02	1.06	-0.12
		(30.82)	(47.67)	(37.03)	(34.29)	(32.67)	(-2.44)	(45.36)	(46.57)	(78.31)	(73.9)	(51.84)	(-4.29)
	smb	1.04	0.83	0.85	0.83	0.93	-0.10	0.31	0.14	0.04	0.03	0.07	-0.24
		(15.75)	(19.37)	(18.62)	(15.74)	(15.68)	(-2.69)	(6.95)	(3.7)	(1.77)	(1.23)	(1.49)	(-5.81)
	hml	0.31	0.15	-0.05	-0.03	0.06	-0.25	0.21	0.11	-0.01	-0.18	-0.14	-0.34
		(4.03)	(2.93)	(-0.67)	(-0.44)	(0.84)	(-7.49)	(3.17)	(2.64)	(-0.27)	(-5.87)	(-2.57)	(-6.07)
	rmw	0.05	0.05	-0.33	-0.27	-0.35	-0.40	0.07	0.23	0.15	0.01	-0.42	-0.49
		(0.42)	(0.7)	(-4.4)	(-2.97)	(-3.75)	(-10.68)	(0.74)	(3.67)	(2.73)	(0.29)	(-4.39)	(-7.27)
	cma	-0.01	0.02	-0.05	-0.05	-0.01	0.01	0.16	0.14	-0.01	-0.09	-0.08	-0.24
		(-0.13)	(0.36)	(-0.49)	(-0.49)	(-0.06)	(0.16)	(2.46)	(2.48)	(-0.18)	(-1.93)	(-0.87)	(-2.7)
	adj-R^2	0.90	0.95	0.92	0.92	0.90	0.34	0.91	0.92	0.95	0.95	0.91	0.38
HXZQ	α	0.37	0.33	0.54	0.42	0.46	0.09	-0.03	0.05	0.09	0.15	0.24	0.27
		(2.93)	(4.34)	(3.79)	(3.24)	(3.54)	(0.95)	(-0.37)	(0.68)	(1.68)	(2.1)	(2.27)	(2.2)
	mkt	99.27	96.49	98.94	98.72	96.98	-2.30	114.92	99.82	103.65	103.36	106.15	-8.77
		(25.16)	(41.16)	(37.92)	(36.47)	(33.14)	(-1)	(38.67)	(44.86)	(73.6)	(66.99)	(54.7)	(-2.31)
	me	87.18	72.06	76.60	73.55	83.11	-4.07	22.63	7.39	-2.21	1.70	4.26	-18.37
		(7.96)	(10.23)	(18.66)	(12.64)	(12.09)	(-0.65)	(3)	(1.2)	(-0.61)	(0.76)	(0.91)	(-2.13)
	ia	26.16	11.78	-19.87	-16.10	-1.63	-27.78	36.47	26.06	-2.88	-30.90	-25.42	-61.89
		(2.08)	(1.56)	(-1.96)	(-1.59)	(-0.15)	(-3.65)	(4.45)	(3.75)	(-0.62)	(-7.82)	(-2.74)	(-4.8)
	roe	-31.79	-21.11	-47.47	-42.37	-50.71	-18.92	-15.50	6.02	1.69	0.37	-32.70	-17.20
		(-3.04)	(-3.59)	(-5.27)	(-4.92)	(-5.97)	(-2.68)	(-2.05)	(1.05)	(0.38)	(0.1)	(-4.4)	(-1.61)
	eg	-17.02	-10.80	-2.81	-5.39	-11.28	5.75	-13.80	-14.79	-5.86	6.04	-8.35	5.45
		(-2.15)	(-2.08)	(-0.52)	(-0.85)	(-1.68)	(0.83)	(-2.16)	(-2.79)	(-1.74)	(1.4)	(-1.05)	(0.55)
	adj-R^2	0.88	0.94	0.92	0.92	0.91	0.10	0.91	0.90	0.95	0.95	0.91	0.19

Panel C: Sorted by margin

Model	Coefficient	P1	P2	P3	P4	P5	P5_P1	P1	P2	P3	P4	P5	P5_P1
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Excess return	avg rt	0.91	1.04	0.93	0.79	0.80	-0.11	0.45	0.69	0.65	0.59	0.60	0.14
	t-stats	(2.3)	(3.4)	(3.55)	(3.24)	(3.5)	(-0.47)	(1.27)	(2.87)	(3.01)	(2.98)	(3.26)	(0.59)
FF3	α	0.04	0.18	0.12	0.03	0.12	0.08	-0.32	-0.01	-0.03	-0.03	0.11	0.43
		(0.23)	(1.93)	(1.77)	(0.42)	(1.83)	(0.4)	(-1.68)	(-0.17)	(-0.59)	(-0.45)	(2.4)	(2.14)
	mkt	1.07	1.03	1.02	1.02	1.00	-0.07	1.26	1.16	1.11	1.06	0.98	-0.29
		(19.7)	(31.68)	(45.1)	(41.26)	(46.3)	(-1.18)	(19.28)	(32.95)	(65.46)	(65.35)	(59.9)	(-4.24)
	smb	1.48	1.07	0.85	0.70	0.55	-0.93	0.76	0.34	0.16	0.05	-0.11	-0.87
		(17.08)	(14.91)	(10.31)	(7.71)	(8.95)	(-8.75)	(10.83)	(4.62)	(3.32)	(1.03)	(-3.78)	(-11.77)
	hml	0.05	0.33	0.34	0.25	0.11	0.06	-0.19	0.03	0.17	0.09	-0.09	0.10
		(0.35)	(4.75)	(5.8)	(3.83)	(2.14)	(0.41)	(-1.47)	(0.39)	(3.95)	(1.78)	(-2.89)	(0.72)
	adj-R^2	0.76	0.88	0.93	0.94	0.93	0.33	0.78	0.87	0.93	0.93	0.95	0.43
FF4	α	0.33	0.39	0.25	0.13	0.21	-0.13	-0.05	0.20	0.06	0.05	0.16	0.21
		(1.62)	(3.55)	(3.74)	(2.05)	(3.18)	(-0.64)	(-0.28)	(2.48)	(1.02)	(0.93)	(3.14)	(1.11)
	mkt	1.00	0.99	0.99	1.00	0.98	-0.02	1.20	1.11	1.08	1.04	0.96	-0.24
		(22.49)	(35.23)	(48.77)	(47.46)	(51.92)	(-0.41)	(24.36)	(40.06)	(72.65)	(65.5)	(63.26)	(-4.34)
	smb	1.47	1.06	0.84	0.69	0.55	-0.92	0.76	0.33	0.16	0.05	-0.11	-0.86
		(15.83)	(20.63)	(12.5)	(8.71)	(10.34)	(-7.59)	(11.14)	(6.57)	(3.97)	(1.14)	(-4.38)	(-11)
	hml	-0.08	0.24	0.28	0.20	0.07	0.15	-0.31	-0.06	0.13	0.06	-0.11	0.20
		(-0.6)	(3.59)	(6.12)	(3.73)	(1.6)	(0.93)	(-2.38)	(-0.85)	(3.28)	(1.18)	(-3.84)	(1.38)
	mom	-0.34	-0.25	-0.16	-0.12	-0.10	0.24	-0.31	-0.25	-0.11	-0.09	-0.06	0.26
		(-3.16)	(-4.46)	(-3.76)	(-3.09)	(-2.99)	(2)	(-3.58)	(-4.85)	(-3.57)	(-3.25)	(-2.41)	(2.73)
	adj-R^2	0.79	0.91	0.95	0.95	0.94	0.37	0.81	0.90	0.93	0.94	0.95	0.47
FF5	α	0.46	0.28	0.07	-0.07	0.09	-0.37	0.14	0.11	-0.12	-0.15	0.08	-0.05
		(2.56)	(2.37)	(0.93)	(-1.03)	(1.41)	(-2.14)	(0.99)	(0.9)	(-2.17)	(-2.69)	(1.7)	(-0.37)
	mkt	0.99	1.01	1.02	1.03	1.00	0.00	1.19	1.14	1.13	1.09	0.97	-0.21
		(20.04)	(27.63)	(44.45)	(51.31)	(47.56)	(0.05)	(23.61)	(31.98)	(63.87)	(72.54)	(58.35)	(-3.9)
	smb	1.17	1.00	0.90	0.79	0.60	-0.57	0.43	0.25	0.21	0.13	-0.07	-0.50
		(12.56)	(16.26)	(21.46)	(19.24)	(18.66)	(-6.82)	(4.08)	(3.68)	(6.82)	(4.24)	(-3.49)	(-4.65)
	hml	-0.12	0.18	0.22	0.14	0.05	0.18	-0.24	0.00	0.09	0.00	-0.04	0.20
		(-1.01)	(2.36)	(4.12)	(2.6)	(1.16)	(1.62)	(-1.6)	(0.04)	(2.11)	(0)	(-1.41)	(1.24)
	rmw	-1.12	-0.26	0.15	0.29	0.13	1.25	-1.22	-0.33	0.18	0.25	0.12	1.34
		(-8.79)	(-2.21)	(1.56)	(3.87)	(2.27)	(11.78)	(-6.58)	(-1.89)	(3.76)	(4.85)	(2.93)	(8.28)
	cma	-0.02	0.00	-0.01	0.00	-0.06	-0.04	-0.05	-0.02	0.12	0.17	-0.07	-0.02
		(-0.11)	(0.02)	(-0.09)	(-0.05)	(-1.28)	(-0.23)	(-0.27)	(-0.18)	(2.7)	(3.07)	(-1.39)	(-0.11)
	adj-R^2	0.82	0.89	0.95	0.96	0.95	0.55	0.87	0.89	0.94	0.95	0.95	0.69
HXZQ	α	0.83	0.60	0.28	0.12	0.27	-0.56	0.44	0.42	0.06	-0.05	0.12	-0.32
		(3.3)	(3.91)	(2.75)	(1.66)	(3.95)	(-2.29)	(2.37)	(2.79)	(0.97)	(-0.78)	(2.29)	(-1.66)
	mkt	98.46	97.56	98.68	99.25	96.79	-1.67	119.26	110.66	108.94	106.28	97.21	-22.05
		(20.34)	(28.25)	(33.92)	(36.18)	(39.97)	(-0.33)	(23.76)	(38.51)	(56.14)	(59.19)	(56.13)	(-3.85)
	me	106.71	86.74	77.93	67.21	50.88	-55.83	39.03	17.71	15.05	6.91	-10.34	-49.37
		(12.63)	(10.6)	(8.98)	(7.57)	(9.04)	(-5.42)	(4.39)	(2.19)	(3.24)	(1.61)	(-4.57)	(-5.29)
	ia	-31.11	11.36	16.56	8.96	-3.73	27.38	-36.37	-4.56	21.45	15.77	-10.80	25.57
		(-1.59)	(0.83)	(1.5)	(1.02)	(-0.62)	(1.29)	(-2.45)	(-0.37)	(3.98)	(2.89)	(-2.75)	(1.61)
	roe	-116.35	-52.85	-16.67	-2.28	-3.37	112.98	-98.77	-34.40	1.39	7.88	5.67	104.43
		(-7.35)	(-4.94)	(-1.88)	(-0.32)	(-0.6)	(6.84)	(-7.77)	(-3.72)	(0.23)	(1.42)	(1.23)	(7.59)
	eg	-2.25	-11.74	-9.19	-10.58	-13.99	-11.74	-10.59	-22.11	-15.92	-7.83	-2.51	8.08
		(-0.19)	(-1.74)	(-1.47)	(-1.84)	(-3)	(-0.9)	(-0.83)	(-2.41)	(-3.14)	(-1.9)	(-0.71)	(0.6)
	adj-R^2	0.82	0.90	0.93	0.94	0.94	0.51	0.84	0.90	0.93	0.94	0.95	0.60

Panel D: Sorted by asset turnover

Model	Coefficient	P1	P2	P3	P4	P5	P5_P1	P1	P2	P3	P4	P5	P5_P1
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Excess return	avg rt	0.53	0.80	0.97	0.98	1.08	0.55	0.35	0.58	0.68	0.64	0.76	0.42
	t-stats	(1.85)	(2.94)	(3.53)	(3.54)	(3.79)	(3.38)	(1.67)	(2.93)	(3.31)	(3.21)	(3.54)	(2.71)
FF3	α	-0.21	0.03	0.17	0.16	0.26	0.46	-0.21	0.02	0.08	0.04	0.14	0.35
		(-1.52)	(0.37)	(2.54)	(2.34)	(2.93)	(2.91)	(-2.28)	(0.42)	(1.56)	(0.6)	(1.58)	(2.38)
	mkt	1.01	1.06	1.04	1.01	0.98	-0.03	1.00	1.05	1.06	1.02	1.02	0.03
		(25.97)	(43.98)	(46.33)	(37.39)	(33.53)	(-0.57)	(26.23)	(66.31)	(68.84)	(49.68)	(32.66)	(0.45)
	smb	0.81	0.87	0.91	0.94	0.97	0.16	-0.05	-0.01	0.06	0.13	0.24	0.29
		(13.06)	(15.37)	(13.82)	(12.19)	(9.83)	(1.28)	(-1.49)	(-0.3)	(1.4)	(2.78)	(2.69)	(3.14)
	hml	0.11	0.12	0.20	0.31	0.34	0.23	0.07	-0.05	0.02	0.08	0.08	0.01
		(1.63)	(2.53)	(3.92)	(5.25)	(5.08)	(2.83)	(0.86)	(-1.52)	(0.35)	(1.93)	(1.03)	(0.06)
	adj-R^2	0.79	0.92	0.93	0.92	0.89	0.05	0.80	0.94	0.94	0.92	0.86	0.07
FF4	α	-0.01	0.20	0.32	0.32	0.42	0.43	-0.06	0.12	0.17	0.10	0.25	0.31
		(-0.05)	(2.16)	(4.21)	(3.87)	(4.47)	(3.06)	(-0.73)	(2.04)	(3.12)	(1.65)	(2.98)	(2.29)
	mkt	0.96	1.02	1.00	0.97	0.95	-0.02	0.96	1.03	1.04	1.00	1.00	0.03
		(30.7)	(46.52)	(48.59)	(42.31)	(36.61)	(-0.48)	(31.8)	(65.71)	(66.91)	(52.16)	(34.84)	(0.66)
	smb	0.80	0.86	0.90	0.93	0.96	0.16	-0.05	-0.01	0.06	0.13	0.23	0.29
		(11.18)	(20.7)	(18.21)	(15.58)	(11.99)	(1.24)	(-1.54)	(-0.49)	(1.68)	(3.13)	(3.09)	(3)
	hml	0.03	0.05	0.13	0.24	0.27	0.24	0.00	-0.09	-0.02	0.05	0.03	0.02
		(0.37)	(1.08)	(3.01)	(4.83)	(4.96)	(3.14)	(0.05)	(-3.19)	(-0.45)	(1.35)	(0.43)	(0.18)
	mom	-0.23	-0.19	-0.18	-0.18	-0.19	0.04	-0.17	-0.11	-0.10	-0.07	-0.12	0.04
		(-3.57)	(-5.1)	(-5.13)	(-3.95)	(-4.02)	(0.54)	(-3.33)	(-4.42)	(-3.68)	(-2.44)	(-3.39)	(0.58)
	adj-R^2	0.81	0.93	0.95	0.94	0.91	0.05	0.82	0.95	0.95	0.92	0.87	0.08
FF5	α	0.08	0.12	0.21	0.16	0.23	0.15	0.03	0.01	-0.02	-0.09	0.02	-0.01
		(0.62)	(1.35)	(2.7)	(1.87)	(2.1)	(1.18)	(0.27)	(0.18)	(-0.36)	(-1.6)	(0.22)	(-0.05)
	mkt	0.96	1.04	1.02	1.00	0.98	0.02	0.95	1.06	1.08	1.04	1.04	0.09
		(30.27)	(37.91)	(42.13)	(36.58)	(32.41)	(0.77)	(30.34)	(72.35)	(75.75)	(56.87)	(39.85)	(2.24)
	smb	0.61	0.80	0.90	0.96	1.00	0.40	-0.19	-0.03	0.14	0.22	0.34	0.53
		(10.29)	(15.76)	(18.61)	(19.94)	(17.17)	(6.39)	(-4.15)	(-1.07)	(6.89)	(7.7)	(7.8)	(10.1)
	hml	0.03	-0.03	0.09	0.18	0.21	0.18	0.19	-0.12	-0.03	0.03	0.04	-0.15
		(0.41)	(-0.42)	(1.42)	(2.96)	(3.01)	(2.75)	(3)	(-3.23)	(-0.89)	(1.07)	(0.56)	(-1.91)
	rmw	-0.76	-0.27	-0.07	0.04	0.12	0.87	-0.52	-0.06	0.25	0.32	0.36	0.88
		(-9.4)	(-3.16)	(-0.98)	(0.39)	(1.06)	(9.16)	(-5)	(-0.93)	(4.87)	(10.11)	(4.83)	(9.68)
	cma	-0.02	0.08	-0.03	-0.02	-0.02	0.00	-0.22	0.16	0.06	0.05	-0.02	0.19
		(-0.21)	(0.84)	(-0.39)	(-0.23)	(-0.2)	(0.02)	(-2.17)	(3.1)	(1.22)	(1)	(-0.32)	(1.91)
	adj-R^2	0.84	0.93	0.94	0.93	0.90	0.35	0.84	0.94	0.95	0.94	0.89	0.37
HXZQ	α	0.41	0.36	0.42	0.37	0.45	0.05	0.25	0.11	0.07	0.01	0.21	-0.04
		(2.87)	(3.19)	(4.01)	(3.24)	(3.3)	(0.31)	(2.17)	(1.56)	(1.13)	(0.16)	(2.23)	(-0.28)
	mkt	93.46	101.53	99.42	97.16	94.82	1.36	93.82	105.12	105.94	101.20	99.93	6.11
		(30.16)	(37.48)	(39.46)	(31.02)	(26.64)	(0.32)	(32.55)	(70.76)	(69.18)	(43.94)	(33.51)	(1.39)
	me	55.10	70.47	79.50	83.92	86.87	31.77	-19.44	-6.74	6.69	15.52	25.91	45.34
		(11.09)	(13.27)	(12.2)	(10.19)	(8.49)	(2.79)	(-4.25)	(-2.53)	(1.81)	(3.54)	(3.45)	(5.77)
	ia	-6.01	-5.90	-5.42	9.06	15.67	21.68	4.87	-0.80	-0.54	6.96	5.14	0.27
		(-0.68)	(-0.71)	(-0.62)	(0.81)	(1.16)	(1.77)	(0.59)	(-0.16)	(-0.09)	(1.52)	(0.69)	(0.02)
	roe	-74.41	-45.39	-33.23	-28.82	-21.10	53.31	-34.85	-13.63	4.02	9.88	19.70	54.55
		(-8.57)	(-6.24)	(-4.76)	(-3.13)	(-1.99)	(4.78)	(-5.44)	(-2.82)	(0.83)	(1.79)	(2.42)	(6.14)
	eg	-14.80	-5.03	-2.86	-4.68	-9.98	4.82	-22.86	-1.23	-1.65	-5.99	-22.77	0.08
		(-1.78)	(-0.95)	(-0.6)	(-0.8)	(-1.29)	(0.43)	(-2.9)	(-0.25)	(-0.33)	(-1.17)	(-3.21)	(0.01)
	adj-R^2	0.85	0.93	0.94	0.92	0.89	0.19	0.84	0.94	0.94	0.92	0.87	0.24

Panel E: Sorted by leverage

Model	Coefficient	P1	P2	P3	P4	P5	P5_P1	P1	P2	P3	P4	P5	P5_P1
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Excess return	avg rt	0.93	0.93	0.90	0.84	0.76	-0.17	0.59	0.59	0.59	0.59	0.57	-0.02
	t-stats	(3.24)	(3.43)	(3.39)	(3.26)	(2.53)	(-1.41)	(2.39)	(2.76)	(3.19)	(3.25)	(2.54)	(-0.09)
FF3	α	0.25	0.18	0.10	0.03	-0.14	-0.39	0.10	0.04	0.03	0.01	-0.14	-0.24
		(2.48)	(2.47)	(1.48)	(0.44)	(-1.45)	(-3.66)	(1)	(0.51)	(0.51)	(0.17)	(-2.08)	(-1.71)
	mkt	0.96	1.01	1.03	1.02	1.09	0.13	1.07	1.08	1.00	0.97	1.11	0.04
		(32.37)	(41.04)	(47.39)	(44.22)	(34.29)	(4)	(30.07)	(43.58)	(75.09)	(65.17)	(49.72)	(0.97)
	smb	1.02	0.91	0.86	0.79	0.94	-0.08	0.27	0.12	0.01	-0.01	0.12	-0.15
		(22.63)	(16.59)	(14.87)	(10.23)	(10.47)	(-0.94)	(5.6)	(3.29)	(0.31)	(-0.3)	(1.9)	(-1.97)
	hml	-0.10	0.11	0.26	0.36	0.44	0.53	-0.50	-0.17	0.06	0.19	0.28	0.78
		(-1.68)	(2.22)	(5.75)	(6.43)	(6.11)	(9.19)	(-7.29)	(-3.73)	(1.48)	(4.05)	(5.38)	(8.44)
	adj-R^2	0.88	0.92	0.93	0.92	0.88	0.35	0.87	0.91	0.94	0.91	0.91	0.39
FF4	α	0.43	0.35	0.25	0.19	0.06	-0.37	0.22	0.16	0.10	0.08	-0.04	-0.26
		(3.7)	(3.83)	(3.12)	(2.67)	(0.56)	(-3.5)	(2.18)	(2.1)	(2.1)	(1.39)	(-0.57)	(-1.89)
	mkt	0.92	0.97	0.99	0.98	1.05	0.13	1.04	1.05	0.99	0.95	1.09	0.05
		(36.25)	(43.42)	(51.73)	(50.58)	(39.28)	(4.78)	(37.08)	(58.4)	(74.67)	(58.51)	(54.56)	(1.24)
	smb	1.01	0.91	0.85	0.78	0.93	-0.08	0.27	0.12	0.01	-0.01	0.12	-0.15
		(22.01)	(21.57)	(20.31)	(13.13)	(13.26)	(-0.97)	(5.36)	(2.87)	(0.31)	(-0.42)	(2.19)	(-1.9)
	hml	-0.17	0.04	0.20	0.29	0.35	0.52	-0.55	-0.22	0.02	0.16	0.23	0.78
		(-2.76)	(0.79)	(5.17)	(6.37)	(5.65)	(9.45)	(-8.05)	(-4.9)	(0.7)	(3.4)	(5.21)	(8.11)
	mom	-0.20	-0.19	-0.18	-0.18	-0.23	-0.02	-0.14	-0.14	-0.09	-0.08	-0.12	0.03
		(-3.58)	(-4.53)	(-4.56)	(-5.16)	(-4.56)	(-0.42)	(-2.66)	(-3.47)	(-4.11)	(-2.51)	(-3.68)	(0.35)
	adj-R^2	0.90	0.94	0.95	0.94	0.90	0.35	0.88	0.92	0.95	0.92	0.92	0.39
FF5	α	0.47	0.29	0.14	0.01	-0.09	-0.56	0.37	0.17	-0.02	-0.13	-0.19	-0.56
		(4.73)	(3.32)	(1.69)	(0.13)	(-0.8)	(-5.42)	(3.75)	(2.26)	(-0.37)	(-2.27)	(-2.61)	(-3.87)
	mkt	0.91	0.98	1.02	1.02	1.07	0.16	1.01	1.05	1.01	1.00	1.11	0.11
		(34.67)	(39.03)	(42.25)	(41.98)	(30.75)	(6.57)	(38.82)	(57.12)	(70.02)	(66.69)	(48.86)	(3.35)
	smb	0.87	0.84	0.84	0.81	0.93	0.06	0.11	0.03	0.05	0.06	0.17	0.06
		(18.43)	(17.01)	(20.7)	(16.8)	(14.38)	(1.43)	(2.07)	(0.89)	(1.65)	(2.66)	(4.26)	(1.14)
	hml	-0.21	0.00	0.14	0.23	0.31	0.51	-0.40	-0.14	0.04	0.09	0.24	0.64
		(-3.17)	(-0.07)	(2.64)	(3.56)	(3.59)	(9.1)	(-6.58)	(-3.29)	(0.86)	(2.77)	(4)	(9.39)
	rmw	-0.56	-0.26	-0.10	0.05	-0.09	0.47	-0.57	-0.30	0.10	0.25	0.12	0.69
		(-7.3)	(-3.56)	(-1.32)	(0.59)	(-0.81)	(5.49)	(-7.09)	(-4.46)	(1.92)	(5.4)	(1.75)	(8.33)
	cma	-0.05	-0.02	0.00	0.04	-0.01	0.04	-0.28	-0.09	0.04	0.22	0.03	0.30
		(-0.47)	(-0.27)	(0.02)	(0.53)	(-0.06)	(0.5)	(-2.95)	(-1.16)	(0.72)	(4.55)	(0.48)	(2.95)
	adj-R^2	0.91	0.93	0.94	0.93	0.89	0.48	0.90	0.92	0.95	0.93	0.92	0.52
HXZQ	α	0.68	0.47	0.39	0.27	0.23	-0.45	0.33	0.25	0.13	-0.05	-0.06	-0.39
		(4.4)	(3.92)	(3.5)	(3.01)	(1.75)	(-3.32)	(2.21)	(2.35)	(2.38)	(-0.7)	(-0.78)	(-1.97)
	mkt	90.55	96.85	98.44	97.46	103.36	12.81	104.67	105.31	98.93	97.63	108.73	4.06
		(29.75)	(38.65)	(39.24)	(35.39)	(28.18)	(3.15)	(38.51)	(54.26)	(74.75)	(60.4)	(38.96)	(0.86)
	me	81.16	76.04	72.94	69.40	78.73	-2.43	14.40	2.17	-1.84	1.11	11.89	-2.51
		(18.38)	(14.41)	(12.32)	(8.2)	(7.66)	(-0.23)	(2.29)	(0.55)	(-0.48)	(0.26)	(1.59)	(-0.21)
	ia	-38.42	-12.72	9.77	21.31	25.23	63.65	-74.42	-27.18	6.47	29.74	25.48	99.90
		(-3.42)	(-1.35)	(1.08)	(2.26)	(2.12)	(5.73)	(-7.89)	(-3.63)	(1.3)	(4.39)	(3.45)	(6.84)
	roe	-58.13	-43.54	-31.72	-26.13	-44.41	13.72	-36.73	-22.26	-2.36	4.10	-9.08	27.65
		(-6.33)	(-5.76)	(-4.28)	(-3.41)	(-4.29)	(1.28)	(-3.95)	(-3.35)	(-0.53)	(0.81)	(-1.21)	(1.89)
	eg	-1.92	0.16	-11.94	-11.43	-12.36	-10.45	13.64	-2.05	-10.85	-3.64	-4.99	-18.63
		(-0.3)	(0.03)	(-2.38)	(-1.97)	(-1.59)	(-1.18)	(1.22)	(-0.31)	(-2.93)	(-0.75)	(-0.87)	(-1.25)
	adj-R^2	0.91	0.93	0.94	0.92	0.89	0.19	0.86	0.92	0.95	0.91	0.91	0.25

Panel F: Sorted by return on equity

Model	Coefficient	P1 (1)	P2 (2)	P3 (3)	P4 (4)	P5 (5)	P5_P1 (6)	P1 (7)	P2 (8)	P3 (9)	P4 (10)	P5 (11)	P5_P1 (12)
Excess return	avg rt	0.91	0.93	0.84	0.85	0.84	-0.08	0.50	0.50	0.55	0.61	0.63	0.13
	t-stats	(2.27)	(3.22)	(3.66)	(3.69)	(3.31)	(-0.34)	(1.45)	(2.04)	(2.94)	(3.15)	(3.2)	(0.57)
FF3	α	0.00	0.11	0.10	0.12	0.10	0.10	-0.32	-0.22	-0.04	0.02	0.12	0.44
		(0.02)	(1.14)	(1.74)	(2.15)	(1.6)	(0.49)	(-1.87)	(-2.3)	(-0.69)	(0.47)	(2.41)	(2.28)
	mkt	1.10	1.02	0.95	0.98	1.06	-0.04	1.30	1.17	0.99	1.02	1.02	-0.28
		(20.75)	(31.24)	(50.64)	(49.46)	(41.67)	(-0.68)	(21.44)	(27.19)	(52.38)	(53.84)	(62.68)	(-4.15)
	smb	1.48	0.96	0.70	0.67	0.69	-0.79	0.68	0.23	0.02	0.04	-0.05	-0.73
		(16.61)	(16.62)	(11.99)	(9.62)	(7.57)	(-6.31)	(8.86)	(4.77)	(0.32)	(1.31)	(-1.02)	(-10.89)
	hml	0.13	0.28	0.31	0.25	0.11	-0.02	-0.05	0.13	0.16	0.06	-0.12	-0.08
		(0.99)	(4.32)	(7.28)	(4.72)	(1.76)	(-0.11)	(-0.4)	(1.43)	(3.97)	(1.73)	(-2.44)	(-0.53)
	adj-R^2	0.75	0.88	0.94	0.95	0.94	0.25	0.78	0.84	0.90	0.93	0.94	0.33
FF4	α	0.29	0.33	0.22	0.21	0.21	-0.09	-0.06	0.01	0.05	0.11	0.17	0.23
		(1.39)	(2.98)	(3.67)	(3.94)	(3.28)	(-0.42)	(-0.37)	(0.13)	(0.92)	(2.14)	(3.17)	(1.19)
	mkt	1.03	0.97	0.92	0.96	1.03	0.00	1.24	1.12	0.97	1.00	1.01	-0.23
		(22.94)	(34.76)	(56.38)	(59.86)	(48.55)	(0.03)	(25.15)	(35.72)	(50.72)	(67.91)	(64.89)	(-4.12)
	smb	1.47	0.95	0.70	0.67	0.68	-0.79	0.67	0.22	0.01	0.04	-0.05	-0.72
		(15.51)	(23.54)	(15.25)	(11.17)	(8.62)	(-5.52)	(10.17)	(5.84)	(0.3)	(1.42)	(-1.17)	(-9.93)
	hml	0.00	0.18	0.26	0.21	0.06	0.06	-0.16	0.02	0.12	0.02	-0.14	0.02
		(0)	(2.93)	(7.91)	(4.79)	(1.23)	(0.38)	(-1.33)	(0.3)	(3.72)	(0.62)	(-2.96)	(0.12)
	mom	-0.33	-0.26	-0.14	-0.11	-0.12	0.21	-0.30	-0.27	-0.11	-0.10	-0.06	0.25
		(-3.26)	(-4.84)	(-4.68)	(-3.63)	(-3.16)	(1.82)	(-3.5)	(-5.29)	(-3.73)	(-4.2)	(-1.9)	(2.41)
	adj-R^2	0.78	0.91	0.95	0.96	0.95	0.28	0.81	0.88	0.91	0.94	0.95	0.37
FF5	α	0.39	0.27	0.08	0.06	0.02	-0.37	0.02	0.01	-0.07	0.01	0.00	-0.02
		(2.09)	(2.42)	(1.17)	(1.14)	(0.35)	(-2.21)	(0.1)	(0.07)	(-1.11)	(0.17)	(-0.04)	(-0.11)
	mkt	1.03	0.99	0.95	0.98	1.05	0.03	1.25	1.14	1.01	1.02	1.03	-0.21
		(20.74)	(28.8)	(47.17)	(58.46)	(54.34)	(0.64)	(22.75)	(34.04)	(49.57)	(49.95)	(80.98)	(-3.8)
	smb	1.20	0.85	0.71	0.74	0.79	-0.41	0.42	0.06	0.00	0.06	0.06	-0.36
		(13.01)	(14.85)	(18.4)	(22.95)	(19.93)	(-5.79)	(4.36)	(0.89)	(0.1)	(2.15)	(3.52)	(-3.91)
	hml	-0.06	0.15	0.17	0.17	0.04	0.10	-0.14	0.12	0.06	0.06	-0.11	0.03
		(-0.45)	(2.05)	(3.51)	(3.91)	(0.84)	(1)	(-0.96)	(1.33)	(1.34)	(1.72)	(-4.08)	(0.23)
	rmw	-1.04	-0.43	0.02	0.20	0.30	1.33	-0.98	-0.63	-0.04	0.06	0.35	1.33
		(-8.45)	(-4.06)	(0.24)	(3.37)	(4.27)	(14.43)	(-4.75)	(-4.23)	(-0.46)	(1.42)	(8.66)	(7.17)
	cma	0.00	0.01	0.10	-0.04	-0.09	-0.09	0.07	-0.01	0.22	-0.03	-0.03	-0.10
		(-0.01)	(0.05)	(1.41)	(-0.86)	(-1.7)	(-0.61)	(0.36)	(-0.09)	(3.46)	(-0.43)	(-0.79)	(-0.51)
	adj-R^2	0.81	0.90	0.95	0.97	0.96	0.52	0.85	0.89	0.91	0.94	0.97	0.62
HXZQ	α	0.77	0.60	0.25	0.21	0.20	-0.57	0.29	0.29	0.08	0.13	0.06	-0.23
		(3.09)	(4.22)	(3.13)	(3.18)	(2.58)	(-2.54)	(1.52)	(2.84)	(1)	(2.05)	(1.1)	(-1.15)
	mkt	101.04	95.51	92.31	95.06	102.27	1.23	124.59	111.88	98.13	100.39	101.84	-22.75
		(21.49)	(31.8)	(40.27)	(42.23)	(39.58)	(0.27)	(25.8)	(48.32)	(51.48)	(51.33)	(66.24)	(-4.22)
	me	109.45	74.19	62.59	64.41	66.76	-42.69	35.80	1.71	-3.17	2.05	-0.43	-36.23
		(12.81)	(11.33)	(9.56)	(9.32)	(8.26)	(-3.87)	(3.82)	(0.28)	(-0.68)	(0.69)	(-0.13)	(-4)
	ia	-21.12	7.71	21.10	8.29	-8.37	12.75	-18.27	11.40	28.51	4.07	-13.81	4.47
		(-1.08)	(0.68)	(2.71)	(1.12)	(-1.04)	(0.68)	(-1.17)	(1.32)	(4.69)	(0.83)	(-2.82)	(0.26)
	roe	-112.99	-63.13	-24.11	-5.26	3.84	116.82	-89.81	-57.67	-12.03	-4.26	20.74	110.55
		(-7.22)	(-7.05)	(-3.87)	(-0.84)	(0.56)	(7.63)	(-6.67)	(-8.14)	(-2.45)	(-0.81)	(4.36)	(7.74)
	eg	-3.38	-10.85	-4.55	-7.58	-12.74	-9.36	-3.18	-17.28	-11.15	-8.40	-6.01	-2.83
		(-0.29)	(-1.85)	(-1.03)	(-1.61)	(-2.28)	(-0.75)	(-0.33)	(-1.78)	(-2.43)	(-2.15)	(-1.48)	(-0.25)
	adj-R^2	0.82	0.92	0.94	0.95	0.94	0.48	0.84	0.90	0.91	0.94	0.95	0.55

Note: This table reports the summary information of the portfolios sorted by tax burden (panel A), interest burden (Panel B), margin (Panel C), turnover (panel D), leverage (Panel E) and ROE (Panel F). Column (1) - (6) represent equal-weighted portfolios and column (7) - (12) are calculated value-weighted. The first part shows the monthly average excess returns on the specific portfolios from first quintile, P1 to the fifth quintile, P5. *P5 - P1* represents the excess returns between the largest quintile portfolio and the smallest quintile portfolio. The rest row panels sequentially displays the factor loadings of regressions of sorted portfolios' returns on Fama-French three-factor model (FF3) with α (intercept), mkt (market factor), smb (small-minus-big), hml (high-minus-low), Carhart four-factor model (C4) with α (intercept), mkt (market factor), smb (small-minus-big), hml (high-minus-low), mom (high-minus-low), Fama-French five-factor model (FF5) with α (intercept), mkt (market factor), smb (small-minus-big), hml (high-minus-low), rmw (robust-minus-weak), cma (conservative-minus-aggressive) and Hou et al. (2015) q-theory five-factor model (HXZQ) with α (intercept), mkt (market factor), me (size factor, high-minus-low), ia (investment factor, high-minus-low), roe (profitability factor, high-minus-low), eg (expected growth factor, high-minus-low). adj-R^2 is the adjusted R-square. Financial firms (SIC: 6000-6999) are excluded from the regressions. Explanatory variables are winsorized at 1 % and 99% for every cross-sectional period. The sample covers from July 1964 to June 2019. *t-stats* is calculated with the Standard errors adjusted by Newey and West (1987) adjustment.

nificant explanatory power on both EW and VW portfolios sorted by interest ratel or margin, except margin-based VW portfolios. Like tax burden, FF5 and HXZQ leave a significant intercept in the EW case of margin, but the intercept is not significant in the VW case. On the contrary, FF5 and HXZQ can only capture the interest burden without an unexplained intercept under EW case.

In Table 1.4 Panel D, turnover shows a clear incremental trend in the excess return from a small quintile to a large one under both EW and VW sorting, which echo the finding of Patin et al. (2020) that the asset turnover is positively correlated with stock returns. Unlike previous ratios, the intercepts of FF3 and FF4 are significant under EW and VW cases, but asset turnover anomaly is digested by the profitability factor in both FF5 and HXZQ.

Table 1.4 Panel E shows a negative relationship between financial leverage and stock returns in EW portfolios but not in VW portfolios. The excess return sorted by financial leverage, which has the largest unexplained intercept in five DuPont ratios, is inexplicable by any multi-factor asset pricing model. Combining the above findings, leverage is the most suitable variable to construct the factor since the popular asset pricing factor models cannot fully absorb the leverage anomaly.

Besides the DuPont ratios, Table 1.4 Panel F exhibits the result of the ROE portfolio on asset pricing models. Five of the eight scenarios can have a good explanation. Only EW FF5 and VW FF3 cannot explain the high-minus-low excess returns of ROE portfolios. Furthermore, HXZQ with profitability cannot, surprisingly, measure the ROE sorted portfolios without the unexplained part.

From Table 1.4, only asset turnover has a different pattern, which can be captured by pricing model with profitability and investment factors. This finding is consistent with previous studies related to profitability (Novy-Marx, 2013; Ball et al., 2015, 2016) which show that asset turnover has a similar structure

Table 1.5: Correlation between the return of long-short portfolios sorted by DuPont ratios and factor returns

Panel A: Fama and French (2015) five-factor model (FF5)

	EW					VW				
	mkt	smb	hml	rmw	cma	mkt	me	ia	roe	eg
tax burden	0.19	0.27	-0.27	-0.64	-0.14	0.27	0.14	-0.20	-0.61	-0.17
interest burden	0.06	0.01	-0.38	-0.42	-0.19	-0.02	-0.14	-0.45	-0.29	-0.31
margin	-0.23	-0.53	0.16	0.66	0.07	-0.42	-0.56	0.22	0.73	0.16
turnover	-0.04	0.15	0.18	0.46	0.05	0.11	0.30	-0.05	0.36	-0.07
leverage	0.05	-0.04	0.55	0.37	0.30	-0.15	-0.12	0.61	0.41	0.44

Panel B: Hou et al. (2015) q-theory five-factor model (HXZQ)

	EW					VW				
	mkt	smb	hml	rmw	cma	mkt	me	ia	roe	eg
tax burden	0.19	0.23	-0.21	-0.49	-0.30	0.27	0.13	-0.21	-0.43	-0.33
interest burden	0.06	0.01	-0.23	-0.20	-0.11	-0.02	-0.15	-0.34	-0.06	-0.05
margin	-0.23	-0.50	0.16	0.63	0.39	-0.42	-0.52	0.23	0.65	0.51
turnover	-0.04	0.14	0.10	0.32	0.16	0.11	0.30	-0.07	0.26	0.02
leverage	0.05	-0.05	0.36	0.08	0.06	-0.15	-0.10	0.47	0.15	0.16

Note: This table shows the Pearson correlation between the return of long-short portfolios sorted by DuPont ratios and factor returns. Panel A is the correlation coefficient between the return of long-short portfolios sorted by DuPont ratios and factor returns in the FF5 model. The mkt, smb, hml, rmw, and cma are the market, size, book-to-market, profitability, and investment factors, respectively. Panel B is the correlation coefficient between five ratios and factor returns in the HXZQ model. The me, ia, roe, and eg represents the size, investment, profitability, and growth factors. In each panel, the left-hand side (EW) is the equal-weighted underlying portfolio and the right-hand side (VW) is the value-weighted underlying portfolio. The sample covers from July 1964 to June 2019.

to other proxies of profitability, earnings over total assets. For rest ratios, the disparate outcomes may be a result of the correlation between sorted portfolio returns and factor returns. I show the correlation between long-short portfolio returns and all factor returns in FF5 and HXZQ in Table 1.5. When the profitability and investment factors are included in the multi-factor model, they produce varying degrees of incremental explanation on long-short portfolios, resulting in a significant intercept. The first four ratios (tax burden, interest burden, margin, and turnover) only have a strong relationship with profitability and a weak relationship with investment. The portfolio sorted by financial leverage, on the other hand, has strong positive correlations with both factors, resulting in a large and significant intercept in FF5. Nonetheless, According to Table 1.5 Panel B, the correlation between financial leverage and q-theory factors is low in HXZQ compared to FF5, implying that financial leverage may be considered a genuine anomaly that HXZQ cannot explain.

1.4.2.2 Fama-MacBeth Analysis

Here, I run the FM regressions to examine the independent performance of DuPont decomposed ratios with four control variables following Novy-Marx (2013), which are the natural logarithm of book-to-market ratio ($log(BE/ME)$), the natural logarithm of the market value of equity ($log(ME)$), the prior one month return ($r_{1,1}$) as the short-term measurement and the prior annual return skip last month ($r_{12,2}$), momentum, as the long-term measurement.

Table 1.6 Panel A shows the result including all available firms. Five ratios fail to capture the expected stock returns since their t-stats are all under the significant threshold except turnover. This result confirms the significant relation between *size* and DuPont decomposed ratios. Due to the low connection between turnover and *size*, the performance of the turnover is not absorbed by the *size* factor, and it can have a significant t-stats (2.39). Surprisingly, ROE entirely loses the explanatory power in predicting the expected stock returns ($t = 0.11$).

This chapter follows Ball et al. (2015) to separate the sample size with the 20 percentile market capitalisation of NYSE into the All-but-Microcap group and Microcap group. Table 1.6 Panel B gives details about the All-but-Microcap scenario. Comparing the result column (1) with Ball et al. (2015), the coefficients and t-statistics of four control variables are lower than Ball et al. (2015), except $r_{1,1}$. In my results, *size* is insignificant (-0.78) in the regression and also

Table 1.6: FM regression with one DuPont decomposition ratio for three scenarios

Panel A: All data

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Independent Variables								
Tax	Mean		-0.021					
	t-stats		(-0.14)					
Interest	Mean			0.028				
	t-stats			(1.12)				
Margin	Mean				0.227			
	t-stats				(1.00)			
Turnover	Mean					0.119		
	t-stats					(2.39)		
Leverage	Mean						-0.016	
	t-stats						(-1.03)	
ROE	Mean							0.021
	t-stats							(0.11)
Control Variables								
size	Mean	-0.12	-0.12	-0.12	-0.12	-0.12	-0.12	-0.13
	t-stats	(-2.59)	(-2.72)	(-2.56)	(-2.85)	(-2.48)	(-2.61)	(-3.01)
btm	Mean	0.29	0.29	0.29	0.29	0.29	0.28	0.28
	t-stats	(3.87)	(3.99)	(3.92)	(3.94)	(3.82)	(3.45)	(3.79)
$r_{1,1}$	Mean	-5.81	-5.87	-5.86	-5.9	-5.91	-5.86	-5.87
	t-stats	(-11.86)	(-11.97)	(-11.86)	(-11.94)	(-12.1)	(-12.01)	(-12.15)
$r_{12,2}$	Mean	0.36	0.35	0.37	0.37	0.33	0.35	0.34
	t-stats	(1.73)	(1.69)	(1.77)	(1.81)	(1.61)	(1.71)	(1.63)

Panel B: All-but-Micro

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Independent Variables								
Tax	Mean		-0.451					
	t-stats		(-2.89)					
Interest	Mean			0.097				
	t-stats			(1.71)				
Margin	Mean				0.376			
	t-stats				(1.28)			
Turnover	Mean					0.133		
	t-stats					(2.56)		
Leverage	Mean						-0.033	
	t-stats						(-1.68)	
ROE	Mean							1.137
	t-stats							(2.94)
Control Variables								
size	Mean	-0.03	-0.03	-0.03	-0.05	-0.02	-0.03	-0.04
	t-stats	(-0.78)	(-0.89)	(-0.86)	(-1.19)	(-0.62)	(-0.85)	(-1.2)
btm	Mean	0.17	0.17	0.18	0.17	0.18	0.16	0.22
	t-stats	(2.3)	(2.34)	(2.43)	(2.3)	(2.39)	(2.01)	(2.67)
$r_{1,1}$	Mean	-3.43	-3.51	-3.49	-3.57	-3.6	-3.52	-3.56
	t-stats	(-7.29)	(-7.42)	(-7.25)	(-7.41)	(-7.68)	(-7.57)	(-7.64)
$r_{12,2}$	Mean	0.57	0.57	0.58	0.58	0.55	0.57	0.57
	t-stats	(2.3)	(2.28)	(2.35)	(2.4)	(2.24)	(2.32)	(2.31)

Panel C: Micro sample

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Independent Variables								
Tax	Mean		0.018					
	t-stats		(0.13)					
Interest	Mean			0.012				
	t-stats			(0.52)				
Margin	Mean				0.108			
	t-stats				(0.47)			
Turnover	Mean					0.097		
	t-stats					(1.77)		
Leverage	Mean						0.004	
	t-stats						(0.2)	
ROE	Mean							-0.1
	t-stats							(-0.58)
Control Variables								
size	Mean	-0.37	-0.37	-0.36	-0.36	-0.37	-0.37	-0.37
	t-stats	(-5.38)	(-5.6)	(-5.34)	(-5.52)	(-5.29)	(-5.38)	(-5.83)
btm	Mean	0.34	0.34	0.34	0.34	0.33	0.34	0.33
	t-stats	(4.28)	(4.53)	(4.29)	(4.46)	(4.22)	(3.86)	(4.22)
$r_{1,1}$	Mean	-6.77	-6.82	-6.8	-6.82	-6.86	-6.82	-6.82
	t-stats	(-11.55)	(-11.64)	(-11.52)	(-11.63)	(-11.73)	(-11.69)	(-11.76)
$r_{12,2}$	Mean	0.3	0.29	0.3	0.31	0.26	0.29	0.26
	t-stats	(1.53)	(1.48)	(1.54)	(1.62)	(1.36)	(1.49)	(1.36)

Note: This table displays the summary information of Fama-MacBeth (FM) regression of expected stock returns on single DuPont decomposed ratio with four control variables under three scenarios. I includes five DuPont ratios: tax burden (Net Income/Pretax Income), interest burden (Pretax Incomde/EBIT), margin (EBIT/Sales), turnover (Sales/Assets), leverage (Assets/Equities). FM regressions includes four control variables: natural logarithm of book-to-market ratio (*btm*, *log(BE/ME)*), natural logarithm of market equity (*size*, *log(ME)*), prior one year stock returns skip last month ($r_{12,2}$) and prior one month stock return ($r_{1,1}$). The sample is separated by using 20th percentile of ME of the New York Stock Exchange (NYSE) into All-but-Microcaps (Medium and large firms above 20th percentile) and Microcaps (tiny and small firms under 20th percentile). Panel A includes all available firms without filter. Panel B includes all medium and large size firms (All-but-Microcaps) and Panel C repeats the same Fama-MacBeth regressions with Microcaps. Financial firms (SIC: 6000-6999) are excluded from the regressions. Explanatory variables are winsorized at 1 % and 99% for every cross-sectional period. The sample covers from July 1964 to June 2019. *t-stats* is calculated with the Standard errors adjusted by Newey and West (1987) adjustment.

has no contribution to the component of predicting stock returns in regression (2) through (7). However, *size* is highly significant in Table 1.6 Panel C in the Microcap group. From regression (2) to (7) in Table 1.6 Panel B, only tax burden, turnover and ROE are significant at the 5% level. ROE has the highest correlation with expected stock returns. 1% increase in ROE will raise the expected stock returns by 1.137%. Interest burden and leverage have marginal explanatory power to capture future stock returns at a 10% significance level. In comparison, the profit margin cannot help to measure monthly stock returns regardless of the degree since it has a high correlation with *size*. The negative coefficients of tax burden and leverage are consistent with the findings of Dhaliwal et al. (2007) that the future stock return will decrease with the increase of the tax burden and leverage for the firm with medium or large market capitalisation.

Table 1.6 Panel C presents the results of the Fama-MacBeth regression under the Microcap data sample. In this subsample, turnover has marginal explanatory power in predicting future stock returns at a 10% significance level ($t = 1.77$). In contrast, the other four ratios in five explanatory variables and ROE are insignificant in regression (2) through (7). Control variables absorb the performance of five explanatory ratios if comparing the regression results of (1) with other regressions in Table 1.6 Panel C. It also gives evidence that small firms are not sensitive to react to the change in ROE or DuPont components.

Overall, the univariate portfolio sorting and FM regressions with single variables can give different findings. From FM regressions, tax burden and turnover are the better choices compared with other three ratios while the portfolio sorting argues the leverage is better than others.

1.4.3 Multi-variable analysis

The single-variable analysis provides evidence that DuPont ratios cannot capture the expected stock returns for Microcaps. The multivariate regressions will only focus on All-but-Micro caps to check whether DuPont ratios can have significant explanatory power in predicting the expected stock returns under multi-variable cases.

Table 1.7 Panel A displays the results of regression in which I combine two different ratios with control variables to examine the power of DuPont decomposition. Regressions with two ratios have similar results with one ratio in Table 1.6 Panel B. However, the significance of decomposed ratios decreases when adding more explanatory variables into the Fama-MacBeth regression. The significance of the tax burden in regression (2) declines remarkably from significant effect (-2.89) to marginally significant (-1.67), which may be caused by the high correlation between tax burden and margin. In regression (4), the interest burden and leverage are correlated with each other, so the power of the interest burden is taken over by leverage. Margin is still not significant in all regressions since a high correlation exists between margin and the natural logarithm of market equity.

Then I take the Fama-MacBeth regression one more step further by including an additional explanatory variable. Hence, regressions (1) to (10) in Table 1.7 Panel B have three explanatory variables and four control variables. The significance of DuPont ratios decreases significantly. Only in regression (6) two variables, tax burden and turnover, capture expected stock returns at the 5% significant level. Interest burden, margin and leverage completely fail to provide incremental explanatory power in predicting future stock returns in all other regressions. Finally, I test the performance with four and five ratios in Table

Table 1.7: FM regression of two and more decomposed ratios

Panel A: Two variables

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Independent Variables											
Tax	Mean	-0.332	-0.275	-0.309	-0.397						
	t-stats	(-1.93)	(-1.67)	(-2.21)	(-2.54)						
Interest	Mean	0.059				0.06	0.064	0.053			
	t-stats	(1.16)				(1.16)	(1.25)	(1.07)			
Margin	Mean		0.355			0.403			0.506	0.383	
	t-stats		(1.24)			(1.39)			(1.61)	(1.31)	
Turnover	Mean			0.125			0.101		0.1		0.131
	t-stats			(2.53)			(1.91)		(1.9)		(2.48)
Leverage	Mean				-0.032			-0.008		-0.009	-0.028
	t-stats				(-1.65)			(-0.43)		(-0.49)	(-1.52)
Control Variables											
size	Mean	-0.04	-0.05	-0.03	-0.04	-0.05	-0.03	-0.04	-0.04	-0.05	-0.03
	t-stats	(-1)	(-1.26)	(-0.74)	(-1.02)	(-1.23)	(-0.76)	(-0.91)	(-1.08)	(-1.24)	(-0.75)
btm	Mean	0.18	0.18	0.19	0.16	0.18	0.19	0.17	0.18	0.16	0.17
	t-stats	(2.44)	(2.38)	(2.47)	(2.08)	(2.4)	(2.46)	(2.16)	(2.32)	(2.08)	(2.13)
$r_{1,1}$	Mean	-3.69	-3.75	-3.77	-3.71	-3.7	-3.77	-3.68	-3.84	-3.77	-3.8
	t-stats	(-7.55)	(-7.65)	(-7.95)	(-7.86)	(-7.6)	(-7.82)	(-7.67)	(-7.94)	(-7.83)	(-8.17)
$r_{12,2}$	Mean	0.6	0.6	0.58	0.59	0.6	0.58	0.6	0.58	0.6	0.57
	t-stats	(2.44)	(2.47)	(2.34)	(2.4)	(2.47)	(2.37)	(2.48)	(2.38)	(2.49)	(2.35)

Panel B: Three variables

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Independent Variables											
Tax	Mean	-0.277	-0.269	-0.332	-0.18	-0.27	-0.287				
	t-stats	(-1.69)	(-1.75)	(-1.93)	(-1.19)	(-1.64)	(-2.04)				
Interest	Mean	0.058	0.06	0.054				0.061	0.056	0.053	
	t-stats	(1.13)	(1.18)	(1.1)				(1.17)	(1.12)	(1.06)	
Margin	Mean	0.368			0.456	0.35		0.526	0.396		0.491
	t-stats	(1.28)			(1.44)	(1.22)		(1.65)	(1.35)		(1.57)
Turnover	Mean		0.087		0.089		0.116	0.1		0.097	0.094
	t-stats		(1.74)		(1.76)		(2.31)	(1.88)		(1.8)	(1.76)
Leverage	Mean			-0.004		-0.006	-0.024		-0.007	-0.007	-0.007
	t-stats			(-0.2)		(-0.29)	(-1.32)		(-0.34)	(-0.38)	(-0.37)
Control Variables											
size	Mean	-0.05	-0.03	-0.04	-0.04	-0.05	-0.03	-0.04	-0.05	-0.03	-0.04
	t-stats	(-1.29)	(-0.87)	(-1.02)	(-1.14)	(-1.3)	(-0.85)	(-1.12)	(-1.26)	(-0.8)	(-1.13)
btm	Mean	0.18	0.19	0.17	0.18	0.16	0.17	0.18	0.17	0.17	0.17
	t-stats	(2.42)	(2.49)	(2.17)	(2.32)	(2.09)	(2.17)	(2.35)	(2.12)	(2.18)	(2.03)
$r_{1,1}$	Mean	-3.76	-3.82	-3.76	-3.89	-3.83	-3.85	-3.86	-3.77	-3.84	-3.92
	t-stats	(-7.66)	(-7.87)	(-7.76)	(-7.97)	(-7.88)	(-8.23)	(-7.95)	(-7.82)	(-8.05)	(-8.2)
$r_{12,2}$	Mean	0.6	0.58	0.6	0.58	0.6	0.57	0.57	0.6	0.58	0.58
	t-stats	(2.46)	(2.38)	(2.47)	(2.39)	(2.48)	(2.35)	(2.37)	(2.49)	(2.39)	(2.4)

Panel C: Four and five variables

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)
Independent Variables							
Tax	Mean	-0.182	-0.277	-0.274	-0.181		-0.186
	t-stats	(-1.22)	(-1.69)	(-1.77)	(-1.19)		(-1.24)
Interest	Mean	0.057	0.056	0.052		0.052	0.049
	t-stats	(1.09)	(1.14)	(1.04)		(0.99)	(0.96)
Margin	Mean	0.475	0.365		0.444	0.514	0.466
	t-stats	(1.48)	(1.26)		(1.41)	(1.61)	(1.46)
Turnover	Mean	0.089		0.083	0.084	0.095	0.084
	t-stats	(1.74)		(1.63)	(1.62)	(1.77)	(1.63)
Leverage	Mean		-0.003	-0.004	-0.005	-0.005	-0.003
	t-stats		(-0.14)	(-0.21)	(-0.24)	(-0.24)	(-0.13)
Control Variables							
size	Mean	-0.05	-0.05	-0.03	-0.05	-0.04	-0.05
	t-stats	(-1.18)	(-1.32)	(-0.9)	(-1.2)	(-1.16)	(-1.22)
btm	Mean	0.18	0.17	0.17	0.17	0.17	0.17
	t-stats	(2.36)	(2.13)	(2.2)	(2.04)	(2.07)	(2.08)
$r_{1,1}$	Mean	-3.9	-3.83	-3.89	-3.97	-3.93	-3.97
	t-stats	(-7.97)	(-7.87)	(-8.1)	(-8.22)	(-8.19)	(-8.21)
$r_{12,2}$	Mean	0.58	0.6	0.58	0.58	0.57	0.58
	t-stats	(2.38)	(2.49)	(2.4)	(2.41)	(2.4)	(2.41)

Note: This table displays the summary information of Fama-MacBeth (FM) regression of expected stock returns on two DuPont decomposed ratios with four control variables under three scenarios. I include five DuPont ratios: tax burden (Net Income/Pretax Income), interest burden (Pretax Incomde/EBIT), margin (EBIT/Sales), turnover (Sales/Assets), leverage (Assets/Equities). FM regressions includes four control variables: natural logarithm of book-to-market ratio (*btm*, *log(BE/ME)*), natural logarithm of market equity (*size*, *log(ME)*), prior one year stock returns skip last month ($r_{12,2}$) and prior one month stock return ($r_{1,1}$). The sample is separated by using 20th percentile of ME of the New York Stock Exchange (NYSE) into All-but-Microcaps (Medium and large firms above 20th percentile) and Microcaps (tiny and small firms under 20th percentile). Panel A includes all available firms without filter. Panel B includes all medium and large size firms (All-but-Microcaps) and Panel C repeats the same Fama-MacBeth regressions with Microcaps. Financial firms (SIC: 6000-6999) are excluded from the regressions. Explanatory variables are winsorized at 1 % and 99% for every cross-sectional period. The sample covers from July 1964 to June 2019. *t-stats* is calculated with the Standard errors adjusted by Newey and West (1987) adjustment.

1.7 Panel C. Turnover is the only ratio which can have marginal explanatory power in predicting the future stock returns at the 10% significant level. If the margin is excluded, the tax burden can have marginal significant performance on prediction ($|t| > 1.69$). But it will be insignificant after adding the margin ($t = -1.22$). The correlation between DuPont ratios will considerably interact and weakens the statistical significance of each other.

1.4.4 Comparison with proxies of profitability

In this subsection, I add three proxies (GP, OP and also including ROE) of profitability into FM regression.[10]

Table 1.1 gives details of three proxies of profitability, GP, OP and CBOP. GP and OP hold the same sample range with DuPont ratios from July 1964 to June 2019, while CBOP only has available data from July 1988 to December 2019. These three variables are more stable and normally distributed than the turnover. The mean values of GP (0.475) and OP (0.163) are significantly smaller than turnover (1.292), which means that the revenue, numerator of turnover, contains redundant information related to earnings compared with pure variable gross profit (Novy-Marx (2013)).

In the beginning, I combine the DuPont ratios with ROE together to check the robustness of the results since ROE is the measurement of profitability in Hou et al. (2015).

[10]Since operating profit should be calculated from gross profit, I follow Koh and Reeb (2015) to set unavailable XSGA and XRD equal to zero. All denominators of three ratios are the one-year lagged total asset, which is the deflating variable for calculating the measurement of profit in Novy-Marx (2013) and Ball et al. (2015).

Table 1.8: FM regression with ROE

Panel A: One variable

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)
Independent Variables							
Tax	Mean		-0.319				
	t-stats		(-2.25)				
Interest	Mean			0.048			
	t-stats			(0.96)			
Margin	Mean				0.128		
	t-stats				(0.42)		
Turnover	Mean					0.106	
	t-stats					(2.2)	
Leverage	Mean						-0.033
	t-stats						(-1.6)
ROE	Mean		1.107	0.953	0.943	1.007	1.33
	t-stats		(2.98)	(2.72)	(2.42)	(2.83)	(3.41)
Control Variables							
size	Mean	-0.03	-0.05	-0.05	-0.05	-0.04	-0.05
	t-stats	(-0.83)	(-1.26)	(-1.31)	(-1.42)	(-1.06)	(-1.35)
btm	Mean	0.17	0.23	0.22	0.21	0.23	0.22
	t-stats	(2.29)	(2.76)	(2.66)	(2.65)	(2.71)	(2.58)
$r_{1,1}$	Mean	-3.54	-3.74	-3.73	-3.77	-3.83	-3.78
	t-stats	(-7.5)	(-7.92)	(-7.79)	(-7.81)	(-8.16)	(-8.2)
$r_{12,2}$	Mean	0.6	0.59	0.6	0.59	0.57	0.58
	t-stats	(2.4)	(2.41)	(2.47)	(2.45)	(2.34)	(2.4)

Panel B: Two variables

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Independent Variables											
Tax	Mean	-0.204	-0.18	-0.221	-0.261						
	t-stats	(-1.32)	(-1.2)	(-1.71)	(-1.89)						
Interest	Mean	0.05				0.042	0.053	0.04			
	t-stats	(1)				(0.82)	(1.05)	(0.84)			
Margin	Mean		0.124			0.139			0.256	0.118	
	t-stats		(0.41)			(0.45)			(0.79)	(0.38)	
Turnover	Mean			0.096			0.067		0.069		0.093
	t-stats			(2.06)			(1.38)		(1.42)		(1.91)
Leverage	Mean				-0.029			-0.004		-0.005	-0.026
	t-stats				(-1.42)			(-0.2)		(-0.26)	(-1.31)
ROE	Mean	0.884	0.887	0.966	1.246	0.932	0.858	1.048	0.766	1.022	1.173
	t-stats	(2.62)	(2.37)	(2.73)	(3.25)	(2.39)	(2.63)	(2.8)	(2.1)	(2.48)	(3.14)
Control Variables											
size	Mean	-0.05	-0.05	-0.04	-0.05	-0.05	-0.04	-0.05	-0.05	-0.05	-0.04
	t-stats	(-1.35)	(-1.44)	(-1.11)	(-1.37)	(-1.44)	(-1.19)	(-1.36)	(-1.27)	(-1.47)	(-1.21)
btm	Mean	0.21	0.21	0.23	0.22	0.21	0.22	0.21	0.21	0.2	0.22
	t-stats	(2.64)	(2.64)	(2.74)	(2.57)	(2.69)	(2.64)	(2.44)	(2.53)	(2.39)	(2.54)
$r_{1,1}$	Mean	-3.79	-3.83	-3.87	-3.83	-3.79	-3.86	-3.82	-3.9	-3.87	-3.92
	t-stats	(-7.83)	(-7.85)	(-8.2)	(-8.24)	(-7.83)	(-8.05)	(-8.06)	(-8.09)	(-8.1)	(-8.48)
$r_{12,2}$	Mean	0.6	0.6	0.58	0.58	0.59	0.57	0.59	0.57	0.59	0.56
	t-stats	(2.48)	(2.46)	(2.35)	(2.4)	(2.44)	(2.38)	(2.46)	(2.38)	(2.45)	(2.33)

Panel C: Three variables

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Independent Variables											
Tax	Mean	-0.188	-0.177	-0.197	-0.147	-0.167	-0.193				
	t-stats	(-1.25)	(-1.23)	(-1.29)	(-1.01)	(-1.13)	(-1.51)				
Interest	Mean	0.043	0.052	0.043				0.043	0.037	0.04	
	t-stats	(0.83)	(1.03)	(0.92)				(0.81)	(0.76)	(0.83)	
Margin	Mean	0.137			0.226	0.114		0.28	0.129		0.213
	t-stats	(0.45)			(0.69)	(0.38)		(0.85)	(0.41)		(0.67)
Turnover	Mean		0.059		0.062		0.085	0.07		0.06	0.06
	t-stats		(1.27)		(1.29)		(1.78)	(1.41)		(1.24)	(1.23)
Leverage	Mean			-0.002		-0.003	-0.024		-0.004	-0.006	-0.007
	t-stats			(-0.08)		(-0.17)	(-1.19)		(-0.19)	(-0.27)	(-0.33)
ROE	Mean	0.874	0.818	0.971	0.755	0.96	1.119	0.746	1.014	0.97	0.861
	t-stats	(2.32)	(2.55)	(2.71)	(2.09)	(2.43)	(3.02)	(2.03)	(2.45)	(2.77)	(2.26)
Control Variables											
size	Mean	-0.05	-0.05	-0.05	-0.05	-0.06	-0.04	-0.05	-0.05	-0.05	-0.05
	t-stats	(-1.47)	(-1.24)	(-1.4)	(-1.31)	(-1.51)	(-1.26)	(-1.3)	(-1.48)	(-1.27)	(-1.36)
btm	Mean	0.21	0.21	0.2	0.21	0.2	0.22	0.21	0.2	0.21	0.2
	t-stats	(2.67)	(2.65)	(2.41)	(2.55)	(2.36)	(2.54)	(2.56)	(2.43)	(2.43)	(2.28)
$r_{1,1}$	Mean	-3.84	-3.9	-3.87	-3.94	-3.92	-3.95	-3.91	-3.87	-3.94	-3.98
	t-stats	(-7.87)	(-8.07)	(-8.08)	(-8.11)	(-8.13)	(-8.51)	(-8.11)	(-8.09)	(-8.32)	(-8.38)
$r_{12,2}$	Mean	0.59	0.58	0.6	0.58	0.59	0.57	0.57	0.58	0.57	0.57
	t-stats	(2.45)	(2.4)	(2.48)	(2.4)	(2.46)	(2.34)	(2.36)	(2.44)	(2.38)	(2.39)

Panel D: Four and five variables

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)
Independent Variables							
Tax	Mean	-0.154	-0.179	-0.177	-0.145		-0.156
	t-stats	(-1.07)	(-1.23)	(-1.24)	(-1)		(-1.1)
Interest	Mean	0.039	0.039	0.041		0.03	0.027
	t-stats	(0.73)	(0.81)	(0.85)		(0.56)	(0.53)
Margin	Mean	0.247	0.128		0.187	0.239	0.21
	t-stats	(0.74)	(0.42)		(0.58)	(0.72)	(0.63)
Turnover	Mean	0.062		0.054	0.053	0.062	0.055
	t-stats	(1.27)		(1.13)	(1.1)	(1.25)	(1.12)
Leverage	Mean		-0.002	-0.003	-0.005	-0.006	-0.004
	t-stats		(-0.08)	(-0.17)	(-0.24)	(-0.3)	(-0.22)
ROE	Mean	0.739	0.947	0.923	0.844	0.846	0.831
	t-stats	(2.02)	(2.4)	(2.7)	(2.24)	(2.19)	(2.19)
Control Variables							
size	Mean	-0.05	-0.06	-0.05	-0.05	-0.05	-0.05
	t-stats	(-1.33)	(-1.51)	(-1.32)	(-1.4)	(-1.36)	(-1.4)
btm	Mean	0.21	0.2	0.21	0.2	0.2	0.2
	t-stats	(2.58)	(2.4)	(2.42)	(2.28)	(2.32)	(2.33)
$r_{1,1}$	Mean	-3.95	-3.92	-3.98	-4.02	-3.99	-4.02
	t-stats	(-8.13)	(-8.12)	(-8.34)	(-8.39)	(-8.38)	(-8.39)
$r_{12,2}$	Mean	0.57	0.59	0.57	0.57	0.57	0.57
	t-stats	(2.39)	(2.46)	(2.4)	(2.41)	(2.38)	(2.4)

Note: This table displays the summary information of Fama-MacBeth (FM) regression of expected stock returns on two DuPont decomposed ratios and ROE with four control variables under three scenarios. I includes five DuPont ratios: tax burden (Net Income/Pretax Income), interest burden (Pretax Incomde/EBIT), margin (EBIT/Sales), turnover (Sales/Assets), leverage (Assets/Equities). ROE is the return on equity (Net Income/ Equities). FM regressions includes four control variables: natural logarithm of book-to-market ratio (*btm*, *log(BE/ME)*), natural logarithm of market equity (*size*, *log(ME)*), prior one year stock returns skip last month ($r_{12,2}$) and prior one month stock return ($r_{1,1}$). The sample is separated by using 20th percentile of ME of the New York Stock Exchange (NYSE) into All-but-Microcaps (Medium and large firms above 20th percentile) and Microcaps (tiny and small firms under 20th percentile). Panel A includes all available firms without filter. Panel B includes all medium and large size firms (All-but-Microcaps) and Panel C repeats the same Fama-MacBeth regressions with Microcaps. Financial firms (SIC: 6000-6999) are excluded from the regressions. Explanatory variables are winsorized at 1 % and 99% for every cross-sectional period. The sample covers from July 1964 to June 2019. *t-stats* is calculated with the Standard errors adjusted by Newey and West (1987) adjustment.

Table 1.8 shows the results after adding ROE to FM regressions.

$$r_{i,t+1} = \alpha_{i,t} + \beta_{1,t}d_{1,t} + \beta_{2,t}roe_{i,t} + \beta_{3,t}size_{i,t} + \beta_{4,t}r_{1,1} + \beta_{5,t}r_{12,2} + \epsilon_{i,t} \tag{1.6}$$

where $d_{1,t}$ is the any one of five DuPont components.

The result in Table 1.8 Panel A shows that the significance of the five ratios drops after adding ROE. However, the tax burden and turnover are still significant in predicting expected stock returns. Interest burden, margin and leverage do not have predictive power. However, the absolute t-value of leverage decreases to 1.6 because of the positive comovement with ROE.

Then, I add one more ratio in the FM regressions to establish dual-ratio regression with ROE shown in Table 1.8 Panel B.

$$r_{i,t+1} = \alpha_{i,t} + \beta_{1,t}d_{1,t} + \beta_{2,t}d_{2,t} + \beta_{3,t}roe_{i,t} + \beta_{4,t}size_{i,t} + \beta_{5,t}r_{1,1} + \beta_{6,t}r_{12,2} + \epsilon_{i,t} \tag{1.7}$$

where $d_{1,t}$ and $d_{2,t}$ are any two components of DuPont decomposition.

After including ROE, the results are worse than non-ROE FM regression Table 1.7 Panel A. Only regression (3) (tax burden and turnover) can have a significant turnover ($t = 2.06$) and marginal significant tax burden ($t = -1.89$) to predict the future stock returns. The other regressions cannot provide efficient results since ROE absorbs the predictive power of most ratios. Table 1.8 Panel C and Table 1.8 Panel D includes at least three decomposed ratios with ROE in regressions. Like before, all variables are insignificant due to the interrelationship between five DuPont decomposed ratios in Table 1.3. The tax burden will reduce the margin and turnover performance, and the interest burden will decrease the effects on the leverage.

Then, I extend the comparison to another two proxies of profitability, gross profit margin and operating profit margin in Table 1.9.

$$r_{i,t+1} = \alpha_{i,t} + \sum_{i=1}^{k} \beta_{i,t} d_{i,t} + \beta_{k+1,t} p_{i,t} + \beta_{k+2,t} size_{i,t} + \beta_{k+3,t} r_{1,1} + \beta_{k+4,t} r_{12,2} + \epsilon_{i,t} \tag{1.8}$$

where k is the maximum number of DuPont composition in the regression, $d_{i,t}$ is the selected variable from DuPont decomposition and $p_{i,t}$ is the underlying proxy of profitability, GP or OP.

From Table 1.8, the multi-ratio with roe regressions cannot provide further information than single-ratio regressions. So I only run the regression of the expected stock returns on individual DuPont ratios with the proxy of profitability and four control variables.

Table 1.9 Panel A gives details on whether the DuPont ratios can survive the regressions with GP. Unfortunately, all five ratios fail to provide the incremental explanatory power to capture the future stock returns. It is worth noting that the interest burden and profit margin will diminish the significance level of GP. The decrease is caused by the similar structure between margin and turnover, including the information related to the profit before subtracting operating expenses.

Table 1.9 Panel B replaces the GP to OP to run the FM regressions with DuPont ratios. The result shows that the OP can have a robust explanatory power in predicting the future stock returns (around $t = 4$) better than GP (around $t = 2.9$) in general. Still, four ratios cannot predict the future stock returns at a 5% significance level. However, turnover survives from the regression with OP and it can be used for pricing the firm-level stock returns.

Besides the above two profitability variables, I also add the cash-based operating profitability derived by Ball et al. (2016) as the fifth control variable. Cash-Based operating profitability only has data available from July 1988 to

Table 1.9: FM regression with significant profitability factors

Panel A: GP

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)
Independent Variables							
Tax	Mean		-0.219				
	t-stats		(-1.52)				
Interest	Mean			0.009			
	t-stats			(0.18)			
Margin	Mean				0.241		
	t-stats				(0.87)		
Turnover	Mean					0.072	
	t-stats					(1.48)	
Leverage	Mean						0.008
	t-stats						(0.42)
GP	Mean		0.387	0.255	0.229	0.329	0.431
	t-stats		(2.91)	(1.99)	(1.73)	(2.61)	(2.96)
Control Variables							
size	Mean	-0.04	-0.04	-0.04	-0.05	-0.03	-0.04
	t-stats	(-1.09)	(-1.01)	(-1.05)	(-1.19)	(-0.86)	(-0.97)
btm	Mean	0.21	0.26	0.24	0.23	0.26	0.27
	t-stats	(2.92)	(3.82)	(3.33)	(3.28)	(3.8)	(3.55)
$r_{1,1}$	Mean	-3.46	-3.86	-3.91	-3.96	-3.93	-3.92
	t-stats	(-7.11)	(-7.84)	(-7.79)	(-7.85)	(-7.95)	(-8.03)
$r_{12,2}$	Mean	0.62	0.59	0.62	0.61	0.59	0.58
	t-stats	(2.48)	(2.36)	(2.49)	(2.48)	(2.37)	(2.36)

Panel B: OP

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)
Independent Variables							
Tax	Mean		-0.183				
	t-stats		(-1.18)				
Interest	Mean			0.007			
	t-stats			(0.14)			
Margin	Mean				0.13		
	t-stats				(0.4)		
Turnover	Mean					0.107	
	t-stats					(2.11)	
Leverage	Mean						0.014
	t-stats						(0.82)
OP	Mean		1.133	0.764	0.699	0.981	1.255
	t-stats		(4.09)	(3.17)	(2.48)	(4.12)	(4.38)
Control Variables							
size	Mean	-0.04	-0.05	-0.05	-0.05	-0.04	-0.05
	t-stats	(-1.09)	(-1.24)	(-1.25)	(-1.27)	(-1.06)	(-1.23)
btm	Mean	0.21	0.29	0.26	0.25	0.29	0.31
	t-stats	(2.92)	(4.13)	(3.55)	(3.69)	(4.03)	(3.98)
$r_{1,1}$	Mean	-3.46	-3.77	-3.81	-3.88	-3.89	-3.8
	t-stats	(-7.11)	(-7.65)	(-7.6)	(-7.7)	(-7.88)	(-7.8)
$r_{12,2}$	Mean	0.62	0.61	0.64	0.63	0.6	0.61
	t-stats	(2.48)	(2.4)	(2.55)	(2.55)	(2.39)	(2.42)

Panel C: CBOP

Varaibles		(1)	(2)	(3)	(4)	(5)	(6)
Independent Variables							
Tax	Mean		-0.053				
	t-stats		(-0.33)				
Interest	Mean			-0.016			
	t-stats			(-0.32)			
Margin	Mean				-0.046		
	t-stats				(-0.29)		
Turnover	Mean					0.051	
	t-stats					(0.71)	
Leverage	Mean						0.005
	t-stats						(0.21)
CBOP	Mean		1.238	1.237	1.231	1.213	1.318
	t-stats		(4.45)	(4.45)	(4.17)	(4.4)	(4.97)
Control Variables							
size	Mean	-0.01	-0.03	-0.02	-0.02	-0.02	-0.03
	t-stats	(-0.14)	(-0.59)	(-0.43)	(-0.44)	(-0.47)	(-0.61)
btm	Mean	0.1	0.17	0.17	0.16	0.17	0.19
	t-stats	(1.2)	(2.11)	(2.03)	(2.15)	(2.05)	(1.98)
$r_{1,1}$	Mean	-1.56	-1.71	-1.72	-1.71	-1.82	-1.79
	t-stats	(-2.79)	(-3.08)	(-3.08)	(-3.07)	(-3.33)	(-3.28)
$r_{12,2}$	Mean	0.08	0.03	0.03	0.05	0.03	0.04
	t-stats	(0.2)	(0.07)	(0.08)	(0.14)	(0.1)	(0.1)

Note: This table displays the summary information of Fama-MacBeth (FM) regression of expected stock returns on single DuPont decomposed ratios and gross profit margin (GP), operating profit margin (OP) and cash-based operating profit margin (CBOP) with four control variables under three scenarios. I includes five DuPont ratios: tax burden (Net Income/Pretax Income), interest burden (Pretax Incomde/EBIT), margin (EBIT/Sales), turnover (Sales/Assets), leverage (Assets/Equities). ROE is the return on equity (Net Income/ Equities). FM regressions includes four control variables: natural logarithm of book-to-market ratio (*btm*, *log(BE/ME)*), natural logarithm of market equity (*size*, *log(ME)*), prior one year stock returns skip last month ($r_{12,2}$) and prior one month stock return ($r_{1,1}$). The sample is separated by using 20th percentile of ME of the New York Stock Exchange (NYSE) into All-but-Microcaps (Medium and large firms above 20th percentile) and Microcaps (tiny and small firms under 20th percentile). Panel A includes all available firms without filter. Panel B includes all medium and large size firms (All-but-Microcaps) and Panel C repeats the same Fama-MacBeth regressions with Microcaps. Financial firms (SIC: 6000-6999) are excluded from the regressions. Explanatory variables are winsorized at 1 % and 99% for every cross-sectional period. The sample covers from July 1964 to June 2019. *t-stats* is calculated with the Standard errors adjusted by Newey and West (1987) adjustment.

December 2017. Thus I run the Fama-MacBeth regression for this subsample. Table 1.9 Panel C shows the Fama-MacBeth average coefficients and t-values for different regressions. Regressions (2) to (6) represent the regression of stock returns on five control variables and different explanatory variables. All five ratios do not have the explanatory power to predict the expected stock returns. Furthermore, cash-based operating profitability is always a significant stock returns predictor.

Overall, the DuPont ratios do not have a good performance in predicting future stock returns after adding significant proxies of profitability. Only turnover can acceptably price the expected stock returns mixed with profitability and four control variables.

1.5 Conclusion

Hou et al. (2015) apply return on equity (ROE) as the proxy of profitability to construct the q-factor asset pricing model. This chapter follows them to decompose ROE into five ratios using DuPont analysis entirely. These five parts are tax burden, interest burden, operating margin, asset turnover and leverage.

I examine the predictive power of five ratios on future stock returns and compare the performance with ROE. Also, I follow Novy-Marx (2013), and Ball et al. (2015) to test whether these ratios can outperform the previous cash-based factor or earning-based factor. In the single-variable analysis, the result of the univariate-sorted portfolio are interesting. FF3 and FFC4 have a considerable explanation on DuPont ratios, except asset turnover. If the portfolio is sorted by tax burden and profit margin, FF5 and HXZQ in the equal-weighted portfolio case cannot explain the high-minus-low excess returns, while FF3 and FFC4 can achieve this without an unexplained part. In contrast, FF5 and HXZQ can produce a significant intercept in value-weighted portfolio case. The only leverage-sorted high-minus-low portfolio is an outlier in all asset pricing models, particularly in the HXZQ, where profitability and investment factors are insignificant in explaining the high-minus-low excess returns.

The results are susceptible to the data sample selection in single-ratio FM regressions. If including all available samples or only selecting Microcaps with market capitalisation under the 20 percentile of NYSE, five ratios cannot predict future stock returns. Under the All-but-Microcaps sample, only tax burden and turnover can accurately measure the stock returns in the FM regressions. At the same time, the other three ratios cannot provide incremental predictive power. The only ratio with marginal significance in predicting the expected returns is turnover when the FM regression has all five DuPont ratios. Besides, the leverage can only marginally predict stock returns, which contradicts the results from the univariate portfolio analysis. When extending the analysis from one-ratio to multi-ratio regression, ratios will interact with each other and reduce the predictive power of capturing expected stock returns. The result is that most of the regressions fail to verify the prediction, and only turnover can have significant performance at the 10% significant level.

In addition, I present DuPont ratios in conjunction with profitability proxies. I have some findings from comparing the profitability factors:

First, the turnover is the only ratio that can have predictive power in explaining expected stock returns under the specific situation with OP. This can be explained by the similar turnover and profitability factors structure. Asset turnover is measured as sales scaled by total assets, similar to gross and operating profitability, which are constructed by the specific profit over lagged total assets.

Second, other ratios cannot survive in the competition with profitability.

The reason is that DuPont ratios only contain some independent parts of information related to operating activities. In contrast, the earnings-related information cannot be wholly reflected in the DuPont ratios. So the decomposed ratios give the evidence that the power of the individual ratios cannot overtake the joint variable, GP, OP and CBOP.

Third, although the DuPont ratios do not have significant performance, CBOP can predict the future stock returns far better than GP and OP since the CBOP includes the information related to cash flow, which is a "pure" proxy of profitability than gross profit and operating profit.

Chapter 2

Earnings-to-market ratios, long-term change in earnings and expected stock returns

2.1 Introduction

In asset pricing, scholars are interested in the value strategies since Graham et al. (1934), where value strategies can deliver higher expected stock returns relative to superior fundamentals (Lakonishok et al. (1994)). In previous decades, research have established the total book equity over total market equity (BE/ME, BTM) as a traditional proxy for the value component, demostrating the relationship between BTM and expected stock returns through two directions.

The first channel is based on the theory of risk premium. Researchers believe that the high BTM (value) firms are riskier than low BTM (growth) firms becasue high BTM firms have lower future earnings or growth rates, and the excess return compensates for holding the risk. Fama and French (1992) demonstrate that BTM has a robust positive performance in explaining average returns by using Fama and MacBeth (1973) regression (FM regression), which may also absorb the explanatory power of leverage and earnings per share (E/P). Later, Fama and French (1993) construct the multi-factor model to price the asset with three systematic risk factors, and they suggest that BTM can capture the stock risk exposure, like *size* (market equity, log(ME)) and market factor. They also contend that value investors have higher expectations for stock returns and utilise BTM as a value gauge.

Another group views the excess return as mispricing, in which the fundamentals are either over- or under-reacted to by investors, leading to an over- or underpriced asset. Lakonishok et al. (1994) argues the positive correlation between BTM and expected stock returns is due to the behavioural nature of value investors who overestimate the performance of past firms' information. They do not discover any proof, however, that value stocks have a higher beta or standard deviation than growth stocks.

Nevertheless, Daniel and Titman (2006) (DT) argue that expected stock returns can only be partially explained by BTM via previous studies. DT define the BTM as the proxy of all historical information based on the correlation between current returns and past performance. Then, they creatively decompose long-term stock returns into *tangible returns*, the proportion can be measured by past accounting-based variables, and *intangible returns*, which are orthogonal to tangible returns and cannot be explained by accounting measurements.[1]

Their empirical studies show that past growth based on accounting has only a limited ability to predict future stock returns. The BTM is insufficient as a proxy for prior performance. The intangible return, which can be approximated

[1]In this chapter, All variables are related to the values, prices and returns. *Tangible* and *intangible* do not refer to the tangible or intangible assets listed on balance sheets. They are essentially a description of expected stock returns explained by the past proxies (*tangible returns*)or unexplained (*intangible returns*).

by past return reversals or issuance activity, effectively captures the negative correlation between financial distress and expected stock returns.[2]

In other words, the postive correlation between BTM and future stock returns is on account of the intangible information missed by the BTM and growth in BE. Gerakos and Linnainmaa (2014) further test DT decomposition model. They argue that the significant performance of tangible returns is caused by price-related variables rather than book returns, which indicates that the BTM is ineffective for measuring past tangible information. Asness et al. (2015) find that value factors should be quantified by a number of value-related variables rather of using only one accounting variable.

This chapter is motivated by the idea of "pure". I would like to decompose a variablethat has a strong correlation with BTM and is more pure in its information than book returns. The substituted accounting-based variable should have a similar structure as BTM. Ball (1978) classify the BTM as a price-scaled variable, and price-scaled variables can be defined as exploring the relationship between stock prices and stock returns using underlying financial information. According to Gerakos and Linnainmaa (2016), decomposed BTM in DT are polluted by corporate events. They discover a novel way to calculate book return using return on equity (ROE) and intangible stock return with lagged stock returns without dividends.

Ball et al. (2020) (BGLN) decompose the total value of book equity (BE) to introduce retained earnings (RE) and contributed capital (CC). They also propose an indicator to distinguish between earning variables' negative and positive values. Their results show that the explanatory power of BTM is attributed to RE, which can reflect the past earnings condition of a company. RE is a good

[2]The *book return* is defined as the log change in book equity between period t to $t-\tau$, $log(\frac{BE_t}{BE_{t-\tau}})$.

substitute for earnings yield and a predictor of earnings growth, and the indicator performs well in FM regressions. According to Ball (1978), retained earnings-to-market ratio (RTM) and contributed capital-to-market ratio (CTM) in BGLN can both be regarded as the price-scaled variables, which means that both of them meet the requirement of decomposition in DT and can be a proxy of past information.

According to the BGLN theory, RE and CC are two totally different variables that represent two different perspectives on accounting data. CC records the issuance operations and the change in net capitals. RE, on the other hand, includes all prior earnings going back to the beginning of the record.

Furthremore, BGLN propose that the information in CC cannot accurately reflect the firm's sensitivity to risk, which implies that CC may not have much of an effect on future stock returns. But CC may be a decent proxy for net issuances since net issuances and expeceted stock returns are adversely correlated. DeAngelo et al. (2006) measure the relationship between CC and the dividends. They find that higher proportions of CC in total equity will reduce the willingness to pay dividends, and firms with lower CC will be less profitable and have a smaller book-to-market ratio. The conclusion implies that companies with larger CC percent may experience lower expected stock returns. Based on the foregoing, this chapter presumes that the predictive power of CC can be replaced by other proxies of expected stock returns, but the change in CC can strongly correlate to the issuance activities. This correlation can be explained by the risk-premium channel: with a decrease in contributed capital growth, the firm will be more risky and have a higher expected return.

RE, in contrast, stands for the distinction between dividends and cumulative earnings. BGLN demonstrates two benefits of RTM as a stand-in for the value factor. Frist, the accumulative information can decrease the sensitivity of the short-term value on expected stock returns. Second, they discover that RTM can serve as a reliable substitute for the earnings yield, which exhibits a strong positive correlation with future stock returns (Ball, 1978; Reinganum, 1981; Basu, 1983; Fama and French, 1992; Lamont, 1998).[3]

Based on the empirical results of BGLN, I assume that the past decomposition components of RTM can be a proxy for tangible returns and will adequately explain expected stock returns. Conversely, variables from CTM should be subsumed by BTM or RTM components.

This chapter's central topic is whether the breakdown components of RTM can serve as a better proxy for expected stock returns than BTM. Additionally, I focus on some important questions: (1) whether the various data filtering criteria can have an impact on the performance of the dcomposed components;[4] (2) whether variations in estimates can result from how frequently data is updated.[5]

In order to address the above questions, I start this chapter by replicating the decomposition from $log(BE/ME)$ (btm) to $log(RE/ME)$ (rtm) and $log(CC/ME)$ (ctm) with the full data sample. I then expand the estimation to the data samples containing only shares with a price greater than 5 dollars or only ME greater than the 20th percentile of NYSE market capitalization.

Some conclusions can be taken from the first section of empirical analysis. First, relative to the data sample till 2003 in DT, this chapter finds that extending the data range to 2020 boosts the explanatory power of price-related

[3]Higher earnings' yield (earnings-to-market, E/P) can generate higher risk-adjusted stock returns. Banz (1981) show earning yield is proxy of size effect. With the control of firm size, the average firm size of low earnings' yield is larger than it in a high earnings' yield portfolio, which lead to the difference in expected stock returns.

[4]Data filtering criterion involves choosing a cutoff point to limit the data sample, such as only considering companies' share price larger than 5 dollars in DT.

[5]Data updating freqency is about using different data frequency to construct variables, where the price-based variables will be collected from annually to monthly. The new ratio will be updated monthly.

variables(*btm* and past stock returns) while decreasing the explanatory power of book returns. Second, the results of BGLN cannot be explained by the five-year lagged price-scaled value ratios. On the other hand, the current period estimate is consistent with BGLN's results that price-scaled value ratios can accurately forecast predicted stock returns. Third, the decomposed variables may be affected in a number of different ways by the data constraint criteria. Most lag variables may have less predictive ability when data filtering constraints are imposed to the data sample. The last and most significant finding in relation to BGLN is that contributed capital returns perform better than the other variables used to explain predicted stock returns. Then, I raise the data updating frequency of price ratios from yearly to monthly. Cutler et al. (1988) compare the correlation between the impact of macro news and stock returns on a monthly and annual basis. The finding indicates that the monthly updated data will reduce the corresponding macroeconomic fundamentals information. It interprets that the past information should perform better in cross-section regressions if the data are rolled over monthly. Huang and Jo (1995) investigate whether the risk premium has consistent performance across data frequencies. They discover that the relevance of the risk premium grows as the time-series interval shrinks. Ghysels et al. (2004) propose a method for matching data with mixed frequencies. They show that a greater time-frequency with a shorter time gap can increase the accuracy of the estimation regressions. Narayan and Sharma (2015) finds that data frequency substantially affects the forward prediction. The exchange rate with a higher frequency will result in a larger forward premium.

I attempt to construct the past variables at the end of each month. Whenever price-related variables are reported on a monthly basis, they will be rebalanced at the end of each month and made available at the starting of the next month. The annual data, however, will not be modified. The outcome indicates that regression using monthly updated data can improve the explanatory power of price-related variables, lagged price-scaled ratios, and prior stock returns on monthly stock returns. The finding aligns with Blackburn and Cakici (2019), who discover that time lags will considerably affect the significance of past ratios.

The third part of the empirical analysis follows DT to partition past five-year variables into prior one-year (medium-term) and prior five-year skip the last year (long-term). Then, following Asness (1995), I split the prior one-year stock return into a prior one-month return (short-term) and a prior one-year stock momentum (medium-term). These results suggest that data frequency considerably affects short-term stock returns, which will have a robust and consistent capacity to predict expected stock returns with the monthly updated data build. Long-term and prior one-year returns, however, are not usually statistically significant. The CC-based regression has the best estimation compared with BE- and RE-based regressions, leaving the most insignificant unexplained portion of the future stock returns in residual. In explaining future stock returns, past CC returns can occasionally outweigh other variables, perhaps even issuance effects.

In the final part, I include current period *log(ME)* ($size_{i,t}$), in order to construct the regressions with four control variables like Novy-Marx (2013) and BGLN. Consistent with previous studies, the conclusion indicates that the size is marginally relevant in describing future stock returns. The performance of $size_{i,t}$ is only significant in CC-based regressions with prior medium-term stock returns.

In addition to the conclusions discussed previously, this chapter addresses issuance activities and negative value indicators. I analyse the issuance effects with decomposed past variables. They usually have strong explanatory power in explaining the expected monthly stock returns and dominate other variables. There are two situations where CC returns can marginally outperform issuance activities: data samples with only positive RE and CC and sample with only share prices above 5 dollars.

The RE indicator, on the other hand, distinguishes between positive and negative ratios. Also, the past information included in stock returns cannot be explained by lagged *rtm* due to the low significance of the past variables and the robust intangible return in the RE-based regressions. The finding corresponds to Gerakos and Linnainmaa (2014) that value returns can provide significant explanatory power in predicting future stock returns.

In addition to the research already described, other literature can be divided into three groups.

The first section is on price-scaled ratios, such as BTM, RTM, and CTM. According to Stattman (1980) and Rosenberg et al. (1985), the book value per share (BEPS) or BTM strategies can provide substantial evidence of market inefficiency, indicating that book-to-market can cause anomalies in asset pricing. Ball (1978) reveals that the price-to-earnings ratio can explain the performance of undiscovered components in expected stock returns. Rosenberg et al. (1985) first discover that market pricing efficiency is related to BTM. They develop a book/price strategy to examine the abnormal performance of the stock market, and the results imply that book-to-market can aid in identifying stock mispricing and generating large profits. Gerakos and Linnainmaa (2014) and Daniel and Titman (2016) find that the total change in book equity can perform better in predicting expected stock return. Daniel and Titman (2016) point out that an operating performance measure of essential information will be preferable to BTM in the absence of external financing options. Additionally, some researchers investigate the connection between RE and stock performance. Graham et al. (1962) mention that RE of value firms might affect stock prices less than growth firms. Harkavy (1953) explores how the earnings distribution can affect stock prices in a given year and discovers that a rise in RE can have a major beneficial effect on stock prices.

The second aspect is the relationship between medium-term and short-term returns and predicted stock returns. De BONDT and Thaler (1985) find a negative correlation between past long-term and future stock returns. Lower past returns predict higher future stock returns. Asness (1995) summarises the previous studies of the strategies of past stock return. He decomposes the long-horizon past stock returns into three components: short-term, medium-term and long-term. He includes three variables in the monthly cross-sectional regressions to compare the explanatory power. He shows that all variables have a substantial tendency to predict the future stock returns and that the one-year momentum component, along with the value factor and size factor, is necessary to explain future stock returns. Jegadeesh and Titman (1993) also discover that the prior one-year return will continue to influence future stock returns. The past short-term return has a positive correlation with the expected stock return. Fama and French (1996) utilise FF3 to evaluate the anomalies and conclude that the past twelve months' stock return cannot be explained by the three-factor model, although sorting the portfolio by value, size, and market factor can absorb other anomalies. Carhart (1997) add the momentum to build a four-factor model (FFC4) based on FF3 as the fourth standard factor to FF3, which is a very good way to measure average returns. In predicting future stock returns, Novy-Marx (2013), Ball et al. (2015) and Bali et al. (2016) all show strong significant prior one-year stock return. Short-term and medium-term returns are represented by prior one-month stock return and one-year stock momentum, prior one-year return skip last month. In addition, long-term return in further decomposition accumulates the past five-year stock return skip the last year.

The last section covers studies on the *size* factor. Banz (1981) and Fama and French (1992) examine a high negative correlation between stock returns. The average returns on tiny stocks are excessively high, while those on large stocks are poor. Berk (1995) contends that *size* can considerably capture the unexplained portion of future stock returns and that size-related information can assist in eliminating anomalies in cross-sectional regressions. Further, Fama and French (1996) shows that the portfolio sorted by *size* and BTM in the three-factor model can explain the stock returns well. In addition, the size factor performs better than the BTM value factor, which is redundant after adding profitability and investment factors in Fama and French (2015) five-factor model (FF5). Fama and French (2008b) explore the return anomalies through cross-section regressions. The size effect can explain the exceptionally high micro-cap stock return, but it has only minimal significance for All-but-microcaps stocks. Hence, Novy-Marx (2013)introduce *size*, *btm* with prior stock returns as the control variables to predict the explanatory power of profitability variables on expected stock returns. Based on Novy-Marx (2013), Ball et al. (2015) and Bali et al. (2016) derive new profitability factor related to earnings, and discover that the significance of *size* is lower than in Novy-Marx (2013).

The remainder of this chapter is organised as follows. Section 2.2 reviews the decomposition of current period *btm*, and describes how to extend the decomposition with *rtm* and *ctm* to run the FM regressions. Then, Section 2.3 shows how to build variables in FM regressions using different data samples, as well as the summary statistics and average correlation coefficients between variables. Following that, Section 2.4 discusses the main results of FM regression of monthly stock returns on the decomposed five-year lagged variables based on several data filtering criteria and two data frequencies. Section 2.5 extends tests on a further decomposition of past variables into long-term, medium-term and short-term components. Finally, Section 2.6 summarises the key findings.

2.2 Model Construction

This section presents the procedure of decomposing logarithm of book-to-market (*btm*) from DT to construct variables as the proxy of past long-term information. Then, I describe how to use the decomposed variables to predict future stock returns. After that, the tangible and intangible returns are determined by extending the decomposition to logarithm of retain earnings-to-market(*rtm*) and logarithm of contributed capital-to-market (*ctm*) from BGLN.

2.2.1 Notation description

This part provides a quick introduction of the chapter's notation in order to set the stage before illustrating the model's construction. There are two classes into which variables can be placed:

1. Stock return related variables:

- $r_{i,t+1}$, monthly stock returns of firm i at year t. Since the variables in the right hand side of FM regression cover a year, I use this notation whenever the data being presented is annual.
- $r^i_{t-\tau,t}$, $r^i_{t-\tau,t-1}$ and $r^i_{t-1,t}$ are yearly stock returns covering the underlying period.
- $r^i_{t-1,t-1_{/12}}$ and $r^i_{t-1_{/12},t}$ are monthly updated one-year momentum and reversal returns.

2. Value ratio related variables:

- $bm_{i,t-\tau}$, $rm_{i,t-\tau}$ and $cm_{i,t-\tau}$ are *btm*, *rtm* and *ctm* at the end June of year $t-\tau$.

- $r^{i,*}_{t-\tau,t}$ and $r^{i,*}_{t-\tau,t-1}$ are the growth rate of book-to-market (BTM), retained earning-to-market (RTM) and contributed capital-to-market (CTM) ratios for underlying time range. $*$ can be BE, RE and CC.

- $r^{i,T*}_{t-\tau,t}$ and $r^{i,I*}_{t-\tau,t}$ are the tangible return and intangible returns computed from the regression of past long-term returns. $*$ can be B (BE), R (RE) and C (CC).

In addition to the above-mentioned factors, this chapter may also present new ones. To the extent that the notation is consistent with the aforementioned variables, the same structure shall be attributed to each.

2.2.2 Links between book equity, retained earnings and contributed capital

This chapter has two main objectives. The first one is combining the ideas of DT and BGLN to assess the predictive power of RTM and CTM compared with BTM under annual regressions. I follow BGLN to construct RE and CC. They describe the correlation between book value of common equity with retained earnings (RE) and contributed capitals (CC) as follows:

CEQ (Common Equity) =
RE (Retained Earnings) + CC (Contributed Capital)+ Other

RE[6] (Retained Earnings)=

[6]BGLN mention that the Compustat includes the ACOMINC into retained earnings variables, and they believe the former variable does not provide information about expected stock returns. So I follow them to remove the accumulative other comprehensive income from reported retained earnings.

Reported-RE (Reported Retained Earnings) - ACOMINC (Accumulated Other Comprehensive Income)

CC (Contributed Capital) =
CSTK (Common/Ordinary Stock capital) + CAPS (Capital Surplus/Share Premium Reserve) - TSTK (Total capital of Treasury Stock)

Other = ACOMINC (Accumulated Other Comprehensive Income)

BGLN mention that RE represents the net earnings on the income statement, which is the difference between past accumulated earnings and the current fiscal year's distributed dividends. RE has a positive relation with firms' operation and book earnings. However, larger dividends will diminish RE because cash and stock dividends reduce a firm's total net earnings. Some RE can be reinvested or used to increase operational activity. Besides, they find that RE can dilute the low informativeness caused by accounting timing issues. RE can be useful for measuring past stock information to predict future expected returns. Paulo (2018) illustrates that RE strongly correlates with a firm's growth potential. As a result, after the 2008 financial crisis, RE of the majority of firms declines as they lose growth possibilities. Negative RE will arise if firms do not perform efficiently or if the economy is in a recession, both resulting in considerable income losses. Additionally, stock repurchases will result in negative retained earnings (BGLN).

Moreover, BGLN define CC, which is the net increment capital received from short-term issuance and repurchase financing activities. CC is a summary of the total additive capital acquired from the common stock and capital surplus over par value and the subtractive capital lost through the repurchase of the common stock from shareholders. Most firms will retain the repurchased shares as treasury capital for future reissue activities. The objective of repurchasing actions is to adjust the undervaluation of the company's prospects, reduce the possibility of a takeover, and dispose of excess money. The stock repurchasing significantly affect stock returns (Ikenberry and Vermaelen, 1996; Ikenberry et al., 2000; Dittmar, 2000; Grullon and Michaely, 2004; Yook, 2010). Alternatively, some studies find the abnormal returns associated with composite issuance effects (Hovakimian et al., 2001; Daniel and Titman, 2006; Lyandres et al., 2008).

Because BE incorporates RE, according to BGLN, BTM is significant for predicting monthly stock returns. BE is highly correlated with RE, which will significantly lower the performance of BE. They also find that all three accounting factors can be used to measure earnings. Therefore, RTM and CTM may be able to substitute BTM as to the values proxy.

2.2.3 The decomposition of book-to-market ratio

DT separate the τ-year realised return into three parts:

$$\tilde{r}_{t-\tau,t} = E_{t-\tau}[\tilde{r}_{t-\tau,t}] + \tilde{r}^{T}_{t-\tau,t} + \tilde{r}^{I}_{t-\tau,t}, \tag{2.1}$$

where $E_{t-\tau}[\tilde{r}_{t-\tau,t}]$ is the expected return at the beginning of the period, $\tilde{r}^{T}_{t-\tau,t}$ and $\tilde{r}^{I}_{t-\tau,t}$ are the tangible and intangible returns, reflect the past information or not, over the period from $t-\tau$ to t. They assume that investors cannot rationally anticipate the future information, thus the decomposition can be used to determine if the components of past returns have predictive power for future stock returns.

DT apply the decomposition to *btm*, which can be decomposed into τ-year lagged btm, the change of BE and the price changes from period $t-\tau$ to t.

$$\begin{aligned} bm_t &= \log\left(\frac{\mathrm{BE}_t}{\mathrm{ME}_t}\right) = \log\left(\frac{B_t}{P_t}\right) \\ &= \log\left(\frac{B_{t-\tau}}{P_{t-\tau}}\right) + \log\left(\frac{B_t}{B_{t-\tau}}\right) - \log\left(\frac{P_t}{P_{t-\tau}}\right), \end{aligned} \tag{2.2}$$

where BE_t and ME_t are the total value of BE and the total value of ME in period t. B_t and P_t are the book value per share and the stock price, respectively, for the same period.

Log return on the book value of equity is derived from the relation between stock returns and prices:

$$r_{t-\tau,t} = \sum_{s=t-\tau+1}^{t} \log\left(\frac{P_s f_s + D_s}{P_{s-1}}\right) = \log\left(\frac{P_t}{P_{t-\tau}}\right) + n_{t-\tau,t}, \tag{2.3}$$

where f_s is the factor to adjust price in period from $s-1$ to s. D_s is total cash distribution per share between $s-1$ and s. Cumulative log share adjustment factor, $n_{t-\tau,t}$, equals to the sum of log price adjustment factors n_s from $t-\tau+1$ to t.[7]

Equation 2.3 presents the relationship between realised returns and the change of prices via the factor to adjust price, and it can also be extended to measure the return on book equity with the changes of BE and the price adjustment factor covering periods $t-\tau$ to t.

$$r^{BE}_{t-\tau,t} = \log\left(\frac{B_t}{B_{t-\tau}}\right) + n_{t-\tau,t}. \tag{2.4}$$

[7]Where single period adjustment factor, n_s equals to the log of factor to adjust price in period plus the log of dividend adjustment. $n_s = \log(f_s) + \log\left(1 + \frac{D_s}{P_s f_s}\right)$.

Therefore, Equation 2.2 can be rewritten as:

$$\begin{aligned} bm_t &= \log\left(\frac{B_{t-\tau}}{P_{t-\tau}}\right) + \log\left(\frac{B_t}{B_{t-\tau}}\right) - \log\left(\frac{P_t}{P_{t-\tau}}\right) \\ &= bm_{t-\tau} + \underbrace{\log\left(\frac{B_t}{B_{t-\tau}}\right) + n_{t-\tau,t}}_{r^{BE}_{t-\tau,t}} - \underbrace{\left(\log\left(\frac{P_t}{P_{t-\tau}}\right) + n_{t-\tau,t}\right)}_{r_{t-\tau,t}}, \end{aligned} \tag{2.5}$$

where book return $r^{BE}_{t-\tau,t}$ is the log changes of book value of equity between $t-\tau$ to t plus the adjustment factor $n_{t-\tau,t}$ for the same period. According to the assumption, the log changes of book value and share price are composed of tangible and intangible information. Both will be utilised for forecasting future monthly stock returns.

2.2.4 Decomposition of retained earnings and contributed capital

The first goal is to determine whether *rtm* and *ctm* can provide incremental information relative to *btm*. Thus, I reproduce the DT decomposition for *rtm* and *ctm* from BGLN because they have the same structure (earnings-to-price) as *btm*:

$$\begin{aligned} rm_t &= \log\left(\frac{\mathrm{RE}_t}{\mathrm{ME}_t}\right) = \log\left(\frac{R_t}{P_t}\right) \\ &= \log\left(\frac{R_{t-\tau}}{P_{t-\tau}}\right) + \log\left(\frac{R_t}{R_{t-\tau}}\right) - \log\left(\frac{P_t}{P_{t-\tau}}\right), \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} cm_t &= \log\left(\frac{\mathrm{CC}_t}{\mathrm{ME}_t}\right) = \log\left(\frac{C_t}{P_t}\right) \\ &= \log\left(\frac{C_{t-\tau}}{P_{t-\tau}}\right) + \log\left(\frac{C_t}{C_{t-\tau}}\right) - \log\left(\frac{P_t}{P_{t-\tau}}\right). \end{aligned} \tag{2.7}$$

Then, using Equation 2.3, *rtm* and *ctm* can be expressed as follows:

$$rm_t = rm_{t-\tau} + r^{RE}_{t-\tau,t} - r_{t-\tau,t}, \tag{2.8}$$

and

$$cm_t = cm_{t-\tau} + r^{CC}_{t-\tau,t} - r_{t-\tau,t}, \tag{2.9}$$

where $r^{RE}_{t-\tau,t}$ and $r^{CC}_{t-\tau,t}$ are the long-term value returns of underlying measurement (RE and CC) from period $t-\tau$ to t.

2.2.5 Tangible and intangible return

In this chapter, I compute the tangible and intangible returns using *rtm* and *ctm* and compare the results to *btm*. Before comparing the results, I will describe the procedure for calculating the tangible and intangible returns from the long-term stock returns.

Following DT's model, I run the cross-sectional regression of stock returns on prior earning ratios and the BE return:

$$r^{i}_{t-\tau,t} = \gamma_0 + \gamma_1 \cdot bm_{i,t-\tau} + \gamma_2 \cdot r^{i,BE}_{t-\tau,t} + u_{i,t}, \tag{2.10}$$

where the tangible return is

$$r^{i,TB}_{t-\tau,t} = \gamma_0 + \gamma_1 bm_{i,t-\tau} + \gamma_2 r^{i,BE}_{t-\tau,t},$$

and intangible return is

$$r^{i,IB}_{t-\tau,t} = u_{i,t}.$$

Computing the tangible and intangible return on *rtm* and *ctm* is similar to

btm:

$$r^{i}_{t-\tau,t} = \underbrace{\gamma_0 + \gamma_1 \cdot rm_{i,t-\tau} + \gamma_2 \cdot r^{i,RE}_{t-\tau,t}}_{r^{i,TR}_{t-\tau,t}} + \underbrace{u_{i,t}}_{r^{i,IR}_{t-\tau,t}} \tag{2.11}$$

and

$$r^{i}_{t-\tau,t} = \underbrace{\gamma_0 + \gamma_1 \cdot cm_{i,t-\tau} + \gamma_2 \cdot r^{i,CC}_{t-\tau,t}}_{r^{i,TC}_{t-\tau,t}} + \underbrace{u_{i,t}}_{r^{i,IC}_{t-\tau,t}} . \tag{2.12}$$

2.2.6 Fama-MacBeth two-step regressions

The core methodology in this chapter is to run FM regressions of future monthly returns on the past information. Twelve cross-sectional regressions will share the same information from July to June of next year, which will be collected at the end of June. In the regressions, I analyse the past information derived from price-scaled earning ratios, and the regression is as follows:

$$r_{i,t+1} = \gamma_0 + \gamma_1 bm_{i,t-5} + \gamma_2 r^{i,BE}_{t-5,t} + \gamma_3 r^{i}_{t-5,t} + e_{i,t+1}, \tag{2.13}$$

where $e_{i,t+1}$ is the estimation residual in the regression.

Alternatively, we can write Equation 2.13 using:

$$r_{i,t+1} = \gamma_0 + \gamma_1 bm_{i,t-5} + \gamma_2 r^{i,BE}_{t-5,t} + \gamma_3 r^{i,IB}_{t-5,t} + e_{i,t+1}, \tag{2.14}$$

which gives equivalent results.

DT introduce the five-year issuance effects, $iss^{i}_{t-5,t}$ as a measure of intangible information. They find that the issuance effect can predict the unexplained portion of predicted stock returns. I apply the same method to insert the issuance effect into FM regression with other past information:

$$r_{i,t+1} = \gamma_0 + \gamma_1 bm_{i,t-5} + \gamma_2 r^{i,BE}_{t-5,t} + \gamma_3 r^{i}_{t-5,t} + \gamma_4 iss^{i}_{t-5,t} + e_{i,t+1}. \tag{2.15}$$

In short, for each earnings-to-market ratio, I will examine six explanatory variables in total, which are earning-to-market in period t, prior ratio in period $t-\tau$, long-term return on value variable, prior long-term stock returns, long-term issuance effects and intangible returns. The regressions for BTM are displayed in Equation 2.13, Equation 2.14 and Equation 2.15.

The complete FM regressions group for RE-related variables will be as follows:

$$\begin{aligned} r_{i,t+1} &= \gamma_0 + \gamma_1 rm_{i,t-5} + \gamma_2 r^{i,RE}_{t-5,t} + \gamma_3 r^{i}_{t-5,t} + e_{i,t+1} \\ r_{i,t+1} &= \gamma_0 + \gamma_1 rm_{i,t-5} + \gamma_2 r^{i,RE}_{t-5,t} + \gamma_3 r^{i}_{t-5,t} + \gamma_4 iss^{i}_{t-5,t} + e_{i,t+1} \\ r_{i,t+1} &= \gamma_0 + \gamma_1 rm_{i,t-5} + \gamma_2 r^{i,RE}_{t-5,t} + \gamma_3 r^{i,IR}_{t-5,t} + e_{i,t+1}. \end{aligned} \tag{2.16}$$

In addition, CC-based variables will run the regression as:

$$\begin{aligned} r_{i,t+1} &= \gamma_0 + \gamma_1 cm_{i,t-5} + \gamma_2 r^{i,CC}_{t-5,t} + \gamma_3 r^{i}_{t-5,t} + e_{i,t+1} \\ r_{i,t+1} &= \gamma_0 + \gamma_1 cm_{i,t-5} + \gamma_2 r^{i,CC}_{t-5,t} + \gamma_3 r^{i}_{t-5,t} + \gamma_4 iss^{i}_{t-5,t} + e_{i,t+1} \\ r_{i,t+1} &= \gamma_0 + \gamma_1 cm_{i,t-5} + \gamma_2 r^{i,CC}_{t-5,t} + \gamma_3 r^{i,IC}_{t-5,t} + e_{i,t+1}. \end{aligned} \tag{2.17}$$

2.2.7 Monthly-updated ratios

The above models are established by the annual variables at the end of June each year. There will be at least a six-month lag between expected stock returns and right-hand side independent variables. Pontiff and Schall (1998) and Lewellen (1999) argue that the monthly updated BTM can provide more market information with no delay. Thus, I extend the computing process from an annual to a monthly basis. In the annually updated regressions, all earning ratios are calculated at the end of June for the next 12 months of FM regressions. Instead, the monthly basis regressions increase the denominator collection frequency from once per year to once per month, meaning the market equity will be collected at the end of each month and made available for the following period. While the numerator (BE, RE, and CC) will be balanced annually at the end of June.

The second target is to divide the return-related variables into medium-term and long-term measures by following Asness (1995). I refer to the information as short-term if the variable spanned no more than one year. Instead, long-term variables will encompass at least one year. Consequently, the five-year stock return, returns on BE, RE, and CC will be divided into long-term and short-term components:

$$\begin{aligned} r^i_{t-5,t} &= r^i_{t-5,t-1} + r^i_{t-1,t}, \\ r^{i,BE}_{t-5,t} &= r^{i,BE}_{t-5,t-1} + r^{i,BE}_{t-1,t}, \\ r^{i,RE}_{t-5,t} &= r^{i,RE}_{t-5,t-1} + r^{i,RE}_{t-1,t}, \\ r^{i,CC}_{t-5,t} &= r^{i,CC}_{t-5,t-1} + r^{i,CC}_{t-1,t}. \end{aligned} \tag{2.18}$$

Previous research prefers to divide the preceding one-year return into two opposing components: the short-term prior one-month return and medium-term prior one-year return skip the last one month. Accordingly, I decompose the "medium-term" one-year stock return as follows:

$$r^i_{t-5,t} = r^i_{t-5,t-1} + r^i_{t-1,t-1_{/12}} + r^i_{t-1_{/12},t} \tag{2.19}$$

where $r^i_{t-1_{/12},t}$ is prior one-month stock return (short-term) and $r^i_{t-1,t-1_{/12}}$ is prior one-year momentum (medium-term), prior one-year stock return skip the last month.

After the decomposition of prior long-run returns, I estimate the long-run tangible and intangible returns on BE, RE and CC by the new long-term past returns $r^i_{t-5,t-1}$:

$$r^i_{t-5,t-1} = \underbrace{\gamma_0 + \gamma_1 bm_{i,t-5} + \gamma_2 r^{i,BE}_{t-5,t-1}}_{r^{i,TB}_{t-5,t-1}} + \underbrace{u_{i,t}}_{r^{i,IB}_{t-5,t-1}}, \tag{2.20}$$

$$r^i_{t-5,t-1} = \underbrace{\gamma_0 + \gamma_1 rm_{i,t-5} + \gamma_2 r^{i,RE}_{t-5,t-1}}_{r^{i,TR}_{t-5,t-1}} + \underbrace{u_{i,t}}_{r^{i,IR}_{t-5,t-1}}, \tag{2.21}$$

$$r^i_{t-5,t-1} = \underbrace{\gamma_0 + \gamma_1 cm_{i,t-5} + \gamma_2 r^{i,CC}_{t-5,t-1}}_{r^{i,TC}_{t-5,t-1}} + \underbrace{u_{i,t}}_{r^{i,IC}_{t-5,t-1}}, \tag{2.22}$$

where $r^{i,T*}_{t-5,t-1}$ is the long-run tangible return of underlying measurement, and $r^{i,I*}_{t-5,t-1}$ represents the intangible returns, which measures the proportion of monthly stock returns unexplained by the specific price-scaled ratio.

In all, the extended monthly cross-sectional FM regression will include the decomposed past information for different time periods, such as prior one-month stock returns, one-year momentum, prior five-year stock returns skip last year, short-term and long-term return on BE

$$\begin{aligned} r_{i,t+1} = \gamma_0 + \gamma_1 bm_{i,t-5} + \gamma_2 r^{i,BE}_{t-5,t-1} + \gamma_3 r^{i,BE}_{t-1,t} \\ + \gamma_4 r^i_{t-5,t-1} + \gamma_5 r^i_{t-1,t-1/12} + \gamma_6 r^i_{t-1/12,t} + e_{i,t+1}, \end{aligned} \tag{2.23}$$

where the long-term stock return $r^i_{t-5,t-1}$ can be replaced by the intangible return of BE $r^{i,IB}_{t-5,t-1}$ or the issuance effect $iss^i_{t-5,t-1}$.

The following regressions will estimate the variables based on RE and CC:

$$
\begin{aligned}
r_{i,t+1} = {} & \gamma_0 + \gamma_1 rm_{i,t-5} + \gamma_2 r^{i,RE}_{t-5,t-1} + \gamma_3 r^{i,RE}_{t-1,t} \\
& + \gamma_4 r^{i}_{t-5,t-1} + \gamma_5 r^{i}_{t-1,t-1/12} + \gamma_6 r^{i}_{t-1/12,t} + e_{i,t+1},
\end{aligned} \tag{2.24}
$$

and

$$
\begin{aligned}
r_{i,t+1} = {} & \gamma_0 + \gamma_1 cm_{i,t-5} + \gamma_2 r^{i,CC}_{t-5,t-1} + \gamma_3 r^{i,CC}_{t-1,t} \\
& + \gamma_4 r^{i}_{t-5,t-1} + \gamma_5 r^{i}_{t-1,t-1/12} + \gamma_6 r^{i}_{t-1/12,t} + e_{i,t+1}.
\end{aligned} \tag{2.25}
$$

2.2.8 Indicator for negative retained earnings

BGLN, DeAngelo et al. (2006), and researchers find that around a quarter of RE are negative, which will be unavailable if applying the logarithm. Therefore, BGLN suggest introducing the indicator for negative RE and CC. If RE or CC are negative, set *rtm* or *ctm* to zero and add the dummy indicator, which equals one in the Fama-MacBeth regression. I set the condition for RE and CC as follows:

- If the $RE_t \leq 0$, $RE_{t-1} \leq 0$ or $RE_{t-5} \leq 0$, any condition is satisfied

$$rm_{i,t-5} = rm_{i,t-1} = rm_{i,t} = r^{i,RE}_{t-5,t} = r^{i,RE}_{t-5,t-1} = r^{i,RE}_{t-1,t} = 0.$$

- If the $CC_t \leq 0$, $CC_{t-1} \leq 0$ or $CC_{t-5} \leq 0$,

$$cm_{i,t-5} = cm_{i,t-1} = cm_{i,t} = r^{i,CC}_{t-5,t} = r^{i,CC}_{t-5,t-1} = r^{i,CC}_{t-1,t} = 0.$$

I add the indicator $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$. If one of the conditions mentioned is met, the value of the indicators will be one; otherwise, it will be zero. The FM regression for RE with indicator will be written as:

$$\begin{aligned} r_{i,t+1} = \gamma_0 + \gamma_1 rm_{i,t-5} + \gamma_2 r^{i,RE}_{t-5,t-1} + \gamma_3 r^{i,RE}_{t-1,t} + \gamma_4 r^{i}_{t-5,t-1} \\ + \gamma_5 r^{i}_{t-1,t-1/12} + \gamma_6 r^{i}_{t-1/12,t} + \gamma_7 ind^{RE}_{i,t} + u_{i,t}, \end{aligned} \tag{2.26}$$

and CC-based regression:

$$\begin{aligned} r_{i,t+1} = \gamma_0 + \gamma_1 cm_{i,t-5} + \gamma_2 r^{i,CC}_{t-5,t-1} + \gamma_3 r^{i,CC}_{t-1,t} + \gamma_4 r^{i}_{t-5,t-1} \\ + \gamma_5 r^{i}_{t-1,t-1/12} + \gamma_6 r^{i}_{t-1/12,t} + \gamma_7 ind^{CC}_{i,t} + u_{i,t}. \end{aligned} \tag{2.27}$$

2.3 Data

This section describes the data samples, data generation procedure, and data filtering criteria. Before beginning the empirical analysis, we must define the time horizon covered in the past information. DT and Asness (1995) examine the returns and accounting information over the past five years from period t through period $t-5$. They claim that five-year ($\tau = 5$) is a reasonable time frame that can provide evidence for return reversals.

2.3.1 Basic Data

The first part of data aligns with DT and BGLN. The whole period will cover the time range from January 1964 to December 2020. In this chapter, the empirical study spans the periods from July 1969 to December 2020 since five-year lagged variables are computed from the full period.

Data samples are split into two different groups. One category includes information from the Center for Research in Security Prices (CRSP) database regarding stock prices and trading activities. From CRSP, I collect the monthly stock returns (CRSP item, RET), unadjusted and adjusted price (PRC, ADJPRC), market capitalisation (TCAP), the factor to adjust price (FACPRC) and common share outstanding (SHR). Following Shumway (1997), the stock returns and prices should be adjusted to eliminate the delisting bias. Bali et al. (2016) provide detailed instructions for adjusting stock returns according to the delisting file in the CRSP database. If the delisting return is reported in CRSP before the last available trading date, it will be replaced by the available delisting return. If the delisting activity is announced after the final month of trading, the delisting return will be counted as the next monthly return after the last month of trading. The data is missing in the CRSP database if firms fail to report the delisting information. If the delisting code is 500, 520, 551 to 573, 574, 580 or 584, -30% can be substituted for the missing return. If the above conditions cannot be met, the return in the delisting month will be marked as -100%.

In addition, some firms may experience abnormal returns in a given period for various reasons.[8] If the stock return code matches the missing stock code, they will be adjusted to missing returns and excluded from the data samples.

Once handling the missing returns, all prior returns in regressions can be calculated via the equation by Bali et al. (2016):

$$r_{t-\tau,t} = log(\prod_{m=t-\tau}^{t} (r_{i,m} + 1)) \tag{2.28}$$

where $t - \tau$ is the start time of the calculation period, t is the end year of the stock returns and $r_{i,m}$ is the monthly stock return in period m for stock i.

They can be allocated into three time periods:

- Short-term: prior one-month return $r^i_{t-1_{/12},t}$;

[8]CRSP guidebook gives the missing return code. -66.0 means that the data only have an available current period price but no previous price. -77.0 indicates that the firm is not at time t. -88.0 happens in the first or the last period, which is out of the time range of stock. -99.0 is due to the missing price.

- Medium-term: prior one-year return $r^i_{t-1,t}$ and one year momentum, prior one-year return skip the last month $r^i_{t-1,t-1_{/12}}$;
- Long-term: prior five-year return $r^i_{t-5,t}$ and prior five-year skip the last year return $r^i_{t-5,t-1}$.

Another group consists of accounting-based variables from COMPUSTAT. I collect all components of the total values of book equity (BE), the retained earnings (RE) and contributed capitals (CC) that I mentioned in section 2.2. Following BGLN, I use the total value of common equity (COMPUSTAT item, CEQ) as the proxy of book equity (BE). The adjusted RE equals retained earnings (RE), subtracting accumulated other comprehensive income (ACOMINC). The contributed capital (CC) is the sum of the capital of common stock (CSTK) and capital surplus (CAPS) minus the total capital of treasury stock (TSTK).[9] All accounting-based variables are on a fiscal-year basis from 1968 to 2019.

To maintain firm list's uniformity, this chapter combines the price-related information from CRSP with accounting variables from COMPUSTAT via KYPERMNO code. In contrast to exclude the financial firms with SIC one-digit six (Ball et al., 2015, 2020), I follow DT and BGLN to cover all stock firms listed on New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and NASDAQ.

Section 2.2 provides two distinct data frequencies to construct the right-hand side explanatory variables. Novy-Marx (2013) and BGLN find that number of firms will report the fiscal year t's statement by the end of December of the calendar year $t-1$. Therefore, there must be a six-month lag between the publication date and construction date, and annual basis value ratios are calculated at the end of June of the calendar year t.

[9] Koh and Reeb (2015) suggest setting missing research and development expense (XRD) to zero, as this value will not bias in the regression. Similarly, I employ the approach to set ACOMINC and TSTK's missing values to zero.

I collect the BE, RE and CC of fiscal year $t-1$ end at anywhere of calendar year $t-1$ and match with the ME at the last trading day of calendar year $t-1$ to build the value ratios. At the same time as ME, the factor to adjust the price in period $t-1$ and long-term past stock returns are also collected. From July of calendar year t to June of calendar year $t+1$, the independent variables are, intuitively, unchanged.

The earnings ratios are created differently every month. BE, RE and CC are still obtained at the end of the fiscal year anywhere calendar year $t-1$, which will be rebalanced in June of the calendar year t. The values are available between July in year t to June in year $t+1$. Instead, the monthly items reported in CRSP, such as TCAP, will be collected at the end of each month and used as the deflator for the next month. Thus, monthly updated accounting variables divided over TCAP can provide more timely information without latency.

2.3.2 Data filtering criterion

In addition to the above calculation requirements, I filter the data with some criteria to keep the sample available.

First, in line with DT, all firms should have available returns, BE, RE and CC, in year t and prior five years ago $t-5$, and positive BE from beginning to end. Second, I employ two techniques to lessen the effect of small firms. One is to remove firms with share prices less than 5 dollars. Another approach is to select firms with ME above the 20th percentile of NYSE. Two methods target two separate directions. DT filter the sample by the price that low-priced stocks lack liquidity and may have a severe bid-ask bounce issue. On the other hand, Fama and French (2008a) find that small firms only account for a negligible proportion of total market values but have a massive effect on the estimation. Thus, they separate the firms by 20th percentile of ME for NYSE stocks, which can produce more steady results than including all stocks. Notably, I utilise ME and stock price at the end of June in year t to break the data sample rather than the information at the end of $t-1$. Finally, the cross-sectional data samples are winsorized at 1% and 99% period-by-period to reduce the impact of outliers on all explanatory variables.

2.3.3 Summary statistics

This subsection presents the summary statistics as well as the correlation between all deconstructed components of the price-scaled variables. I compare the results for different scenarios with data filtering conditions.

2.3.3.1 The average

I begin by comparing whether the data sample contains negative RE and CC. The negative indicator will be applied if negative values are included.

Table 2.1 reports the average summary statistics for the full sample from July 1969 to December 2020 without data restriction. Under the full sample condition, the average available number of firms can reach about 2750. The mean share price is about 32.5 dollars, and around 22% of share prices are less than 5 dollars. While around 47% of ME is under the average 20th percentile of NYSE market cap of 277 million dollars. If the sample is filtered using the ME criterion, the data sample size will be reduced significantly.

Intuitively, the mean values of negative indicators, $ind_{i,t}^{RE}$ and $ind_{i,t}^{CC}$, indicate that about 25% of RE are negative while no more than 7% of CC are less than zero. The RE indicator has no explanatory power to distinguish positive from negative RE during half of the periods. Remarkably, there will be 540 out of 618 periods with all zero indicators in the FM regressions. The low percentage of negative values in CC can cause bias when running FM regression with

Table 2.1: Summary statistics of the annual basis returns and accounting information

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Prices	32.46	449.89	33.01	1717.77	0.36	0.97	2.17	8.03	17.73	32.74	68.19	119.81	25750.11	2749.51	0.00
Shares	68941.50	226714.08	10.43	172.13	152.38	997.59	2174.62	7756.54	19039.58	50147.74	257144.70	855741.20	5249873.20	2749.51	0.00
ME	2740.39	11182.21	12.55	230.57	0.59	3.79	10.88	71.58	330.71	1422.38	11146.30	45430.99	262200.10	2749.51	0.00
BE	1488.44	5536.37	11.71	224.89	0.18	2.68	8.00	49.22	194.07	756.50	6328.05	24622.85	110615.74	2749.51	0.00
RE	963.14	4882.68	11.39	215.84	-41551.67	-1053.88	-232.01	-6.36	58.64	393.03	4497.20	18497.34	126399.85	2749.51	608.58
CC	562.67	3159.30	14.61	519.90	-52696.53	-1892.33	-100.00	21.06	87.70	339.89	2618.25	9879.37	84712.25	2749.51	170.60
$bm_{i,t}$	-0.33	0.93	0.87	10.87	-6.24	-2.55	-1.72	-0.82	-0.33	0.10	0.98	2.74	6.34	2749.51	1926.41
$rm_{i,t}$	-0.59	0.95	-0.03	9.74	-7.22	-3.19	-2.14	-1.12	-0.43	-0.02	0.38	2.12	5.79	2749.51	1659.70
$cm_{i,t}$	-0.94	1.38	-0.34	5.90	-9.36	-4.75	-3.21	-1.70	-0.84	-0.14	1.13	2.47	5.79	2749.51	1994.04
$iss^i_{t-5,t}$	0.06	0.47	1.74	29.17	-3.68	-0.84	-0.40	-0.17	-0.02	0.19	0.82	1.75	4.93	2749.51	1400.48
$bm_{i,t-5}$	-0.43	0.96	0.81	11.51	-6.51	-2.76	-1.87	-0.92	-0.40	0.03	0.86	2.68	6.59	2749.51	1964.01
$rm_{i,t-5}$	-0.74	1.06	-0.29	9.16	-7.72	-3.79	-2.53	-1.30	-0.55	-0.09	0.27	2.04	6.01	2749.51	1696.72
$cm_{i,t-5}$	-1.00	1.26	-0.19	6.70	-8.44	-4.45	-3.08	-1.70	-0.91	-0.23	0.81	2.28	5.78	2749.51	2077.21
$r_{i,t}$	0.01	0.13	2.82	55.44	-0.61	-0.28	-0.16	-0.05	0.00	0.06	0.20	0.42	1.93	2749.51	1262.86
$r^i_{t-5,t}$	0.28	0.91	-0.68	6.59	-5.09	-2.36	-1.31	-0.18	0.37	0.83	1.58	2.26	3.68	2749.51	885.19
$r^{i,BE}_{t-5,t}$	0.73	1.02	-0.12	10.94	-6.09	-2.26	-0.79	0.24	0.67	1.24	2.33	3.42	7.11	2749.51	484.21
$r^{i,RE}_{t-5,t}$	0.73	0.94	0.99	10.96	-4.99	-1.37	-0.25	0.16	0.52	1.19	2.41	3.69	8.10	2749.51	215.88
$r^{i,CC}_{t-5,t}$	0.65	1.02	0.39	12.30	-6.80	-2.01	-0.62	0.09	0.49	1.11	2.38	3.69	8.22	2749.51	387.38
ind^i_{RE}	0.25	0.40	1.57	5.15	0.00	0.00	0.00	0.00	0.00	0.60	0.96	1.00	1.00	2749.51	0.00
ind^i_{CC}	0.07	0.21	4.15	72.86	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.73	0.75	2749.51	0.00
$r^{i,TB}_{t-5,t}$	0.28	0.59	-0.34	9.25	-3.56	-1.50	-0.65	0.00	0.28	0.58	1.18	1.78	3.74	2749.51	849.98
$r^{i,IB}_{t-5,t}$	0.00	0.68	-0.46	6.38	-4.07	-1.89	-1.18	-0.37	0.04	0.41	1.03	1.59	2.97	2749.51	1286.71
$r^{i,TR}_{t-5,t}$	0.28	0.41	0.55	11.57	-2.52	-0.79	-0.22	0.07	0.19	0.48	0.99	1.50	3.23	2749.51	812.39
$r^{i,IR}_{t-5,t}$	0.00	0.80	-0.72	8.20	-5.53	-2.40	-1.37	-0.38	0.07	0.45	1.15	1.83	3.40	2749.51	1253.57
$r^{i,TC}_{t-5,t}$	0.28	0.32	-0.08	10.74	-2.07	-0.62	-0.19	0.13	0.25	0.44	0.80	1.17	2.35	2749.51	785.38
$r^{i,IC}_{t-5,t}$	0.00	0.84	-0.62	4.95	-4.68	-2.45	-1.52	-0.44	0.09	0.52	1.21	1.81	3.10	2749.51	1230.56
$r^{i,TT}_{t-5,t}$	0.41	0.58	-0.25	9.57	-3.46	-1.31	-0.51	0.14	0.39	0.70	1.31	1.90	3.87	2749.51	522.33
$r^{i,IT}_{t-5,t}$	0.00	0.66	-0.34	6.21	-3.92	-1.80	-1.12	-0.37	0.03	0.40	1.03	1.59	3.10	2749.51	1308.86

Note: This table displays the summary statistics for the full sample with retained earnings (RE) and contributed capital (CC) indicator (value < 0), from July 1969 to December 2020. The table contains all stock information for the first three rows, Stock price, number of shares in thousands, market values of the equity in a million. Then BE, RE and CC represents the book equity, retained earnings and contributed capital, which are the fundamental measurement collected from the end of each year for earnings. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ are the book-to-market, retained earnings-to-market and contributed capital-to-market ratio collected at the end of year t. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the prior five-year lagged ratios collected at the end of year$t-5$. All returns with $r^{i,*}_{t-5,t}$ are the five years specific returns from the last trading day of year $t-6$ to the last trading day of year t. $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$ are the indicators to distinguish the negative and positive value of retained earning and contributed capital. T and I denote the tangible and intangible, respectively.

the negative value indicator.[10]

In addition, the timely variables have higher mean values than data collected five years ago. The average of *btm*, *rtm* and *ctm* in period t is about -0.33, -0.59 and -0.94, respectively. But the same variables in period $t-5$ are -0.43, -0.74, -1.00. The mean values of past five-year value returns deviate greatly from the average of five-year stock return (0.28). BE and RE returns ($r^{i,BE}_{t-5,t}$ and $r^{i,RE}_{t-5,t}$) 0.73, while the CC return ($r^{i,CC}_{t-5,t}$) is 0.65.

Table 2.2 presents the results of the data sample without the indicators of RE and CC. Clearly, the average sample size falls to 1765 for each period. The average share price increases to 41.5 dollars from 32.5 dollars. The elimination of negative RE and CC considerably impacts the average values of BE, RE, and CC, which increase from 1488, 963, and 562 to 1952, 1240, and 751 million, respectively. In addition, all past long-run returns grow to a varying degree, with $r^{i,RE}_{t-5,t}$ exhibiting the largest increase to 0.97.

Table 2.3 provides the summary statistics of all explanatory variables with monthly frequency constructed at the end of each month. The reduction of time lag improves the mean of ME, which increases from 2635 to 2824 million, which results in an reduction by 0.05 to 0.06 of the short-term and long-term price-scaled variables. While five-year value returns are unaffected, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$ remain at 0.73, 0.73 and 0.65, respectively. Nevertheless, the mean value of five-year stock return and all four tangible returns decrease from 0.28 to 0.25.

In a nutshell, including the negative RE and CC can have a significant influence on all variables, such that the mean values of price-scaled ratios and long-run past variables will increase more or less. For the above reason, the negative RE and CC with indicator variables should be included in the data sample.

[10] All zero indicators will be included in the FM regression, and the estimation coefficient for the indicator, in periods with no negative RE or CC, will be zero.

Table 2.2: Summary statistics of the annual basis of returns and accounting information for only positive RE and CC data sample

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Prices	41.46	597.97	26.39	1057.09	0.64	1.76	3.84	11.64	21.52	35.38	70.22	125.41	25667.86	1764.78	0.00
Shares	72854.52	236645.40	9.48	141.90	161.26	966.28	2019.96	6946.97	17904.05	51645.07	281936.86	922837.98	4746540.34	1764.78	0.00
ME	3039.83	11640.79	11.62	202.74	1.42	5.45	16.45	103.60	435.13	1702.09	12426.41	50312.31	233714.94	1764.78	0.00
BE	1952.60	6398.50	10.64	190.11	2.25	7.06	17.18	83.11	281.47	1085.15	8829.72	29173.05	109520.01	1764.78	0.00
RE	1240.75	4405.90	10.05	145.73	0.12	1.79	6.86	40.00	152.11	617.07	5618.55	19834.53	75476.75	1764.78	0.00
CC	751.11	2764.36	14.42	388.05	0.05	0.74	3.75	25.23	92.17	385.12	3371.04	11440.75	55659.57	1764.78	0.00
$bm_{i,t}$	-0.21	0.91	1.66	12.15	-3.59	-2.02	-1.44	-0.69	-0.27	0.13	1.21	3.11	6.21	1764.78	1163.61
$rm_{i,t}$	-0.85	1.06	0.64	9.86	-6.57	-3.47	-2.38	-1.39	-0.85	-0.36	0.71	2.59	5.79	1764.78	1489.68
$cm_{i,t}$	-1.40	1.36	-0.28	6.75	-9.35	-5.43	-3.60	-2.11	-1.30	-0.64	0.47	2.28	5.28	1764.78	1599.72
$iss^{i}_{t-5,t}$	-0.02	0.36	0.52	33.58	-3.41	-0.86	-0.39	-0.18	-0.05	0.09	0.54	1.11	3.46	1764.78	1046.64
$bm_{i,t-5}$	-0.31	0.93	1.65	12.95	-3.82	-2.21	-1.60	-0.79	-0.34	0.06	1.07	3.02	6.52	1764.78	1198.25
$rm_{i,t-5}$	-1.03	1.16	0.35	9.21	-7.45	-4.13	-2.83	-1.61	-0.98	-0.46	0.55	2.50	6.00	1764.78	1519.74
$cm_{i,t-5}$	-1.38	1.27	-0.03	7.65	-8.44	-5.04	-3.43	-2.04	-1.30	-0.69	0.35	2.20	5.59	1764.78	1609.69
$r_{i,t}$	0.01	0.11	1.63	25.18	-0.50	-0.24	-0.14	-0.05	0.01	0.06	0.18	0.34	1.11	1764.78	794.77
$r^{i}_{t-5,t}$	0.41	0.73	-0.20	4.56	-2.93	-1.54	-0.84	-0.01	0.44	0.85	1.56	2.19	3.39	1764.78	454.85
$r^{i,BE}_{t-5,t}$	0.88	0.79	1.04	11.40	-2.78	-0.68	-0.13	0.38	0.75	1.28	2.29	3.25	6.34	1764.78	149.28
$r^{i,RE}_{t-5,t}$	0.97	1.03	0.54	10.25	-4.94	-1.70	-0.48	0.42	0.87	1.46	2.69	4.06	8.10	1764.78	198.03
$r^{i,CC}_{t-5,t}$	0.76	1.05	0.52	11.88	-6.14	-1.87	-0.48	0.18	0.61	1.25	2.53	3.93	7.99	1764.78	242.78
$r^{i,TB}_{t-5,t}$	0.41	0.43	0.83	8.96	-1.68	-0.49	-0.18	0.14	0.36	0.63	1.16	1.70	3.29	1764.78	371.47
$r^{i,IB}_{t-5,t}$	0.00	0.59	-0.49	6.37	-3.27	-1.62	-1.00	-0.33	0.03	0.36	0.90	1.35	2.21	1764.78	837.90
$r^{i,TR}_{t-5,t}$	0.41	0.43	-0.02	8.93	-2.23	-0.79	-0.24	0.19	0.39	0.63	1.10	1.56	2.96	1764.78	350.93
$r^{i,IR}_{t-5,t}$	0.00	0.59	-0.27	5.90	-3.05	-1.58	-0.99	-0.34	0.03	0.36	0.92	1.42	2.59	1764.78	850.74
$r^{i,TC}_{t-5,t}$	0.41	0.26	-0.01	9.04	-1.13	-0.28	0.03	0.26	0.39	0.55	0.85	1.14	1.92	1764.78	303.84
$r^{i,IC}_{t-5,t}$	0.00	0.68	-0.32	4.44	-3.24	-1.84	-1.17	-0.38	0.04	0.42	1.04	1.57	2.52	1764.78	838.71
$r^{i,TT}_{t-5,t}$	0.50	0.45	0.73	9.74	-1.61	-0.55	-0.13	0.23	0.44	0.73	1.29	1.81	3.36	1764.78	170.12
$r^{i,IT}_{t-5,t}$	0.00	0.57	-0.44	6.52	-3.16	-1.56	-0.97	-0.32	0.03	0.35	0.88	1.33	2.24	1764.78	844.09

Note: This table reports the summary statistics the data sample only including the positive retained earnings and contributed capital, from July 1969 to December 2020. The table contains all stock information for the first three rows, Stock price, number of shares in thousands, market values of the equity in a million. Then BE, RE and CC represents the book equity, retained earnings and contributed capital, which are the fundamental measurement collected from the end of each year for earnings. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ are the book-to-market, retained earnings-to-market and contributed capital-to-market ratio collected at the end of year t. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the prior five-year lagged ratios collected at the end of year$t-5$. All returns with $r^{i,*}_{t-5,t}$ are the five years specific returns from the last trading day of year $t-6$ to the last trading day of year t. $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$ are the indicators to distinguish the negative and positive value of retained earning and contributed capital. T and I denote the tangible and intangible, respectively.

Table 2.3: Summary statistics of the monthly basis returns and accounting information

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Prices	33.09	475.53	32.98	1717.87	0.32	0.85	2.00	7.82	17.61	32.93	69.56	123.88	27214.75	2749.51	0.00
Shares	70480.81	233376.92	10.53	178.85	154.82	999.48	2188.20	7851.88	19419.70	51385.02	262784.54	871548.23	5557319.46	2749.51	0.00
ME	2824.03	11660.29	12.63	233.98	0.58	3.48	10.29	70.61	331.49	1450.24	11466.18	46681.99	277540.45	2749.51	0.00
$bm_{i,t}$	-0.39	1.00	0.74	9.37	-6.30	-2.76	-1.89	-0.91	-0.39	0.08	1.09	2.82	6.41	2749.51	1964.99
$rm_{i,t}$	-0.64	0.99	-0.06	8.70	-7.25	-3.33	-2.27	-1.20	-0.46	-0.02	0.42	2.15	5.74	2749.51	1673.61
$cm_{i,t}$	-0.99	1.42	-0.28	5.60	-9.37	-4.86	-3.32	-1.80	-0.90	-0.17	1.18	2.52	6.11	2749.51	2008.17
$iss^i_{t-5,t}$	0.05	0.48	1.89	33.47	-3.61	-0.87	-0.41	-0.17	-0.02	0.18	0.81	1.78	5.22	2749.51	1416.94
$bm_{i,t-5}$	-0.52	1.01	0.58	9.83	-6.63	-3.05	-2.10	-1.04	-0.48	-0.02	0.88	2.63	6.34	2749.51	2028.67
$rm_{i,t-5}$	-0.81	1.10	-0.40	8.32	-7.82	-4.05	-2.70	-1.40	-0.62	-0.09	0.25	1.95	5.77	2749.51	1726.90
$cm_{i,t-5}$	-1.09	1.30	-0.22	6.22	-8.56	-4.58	-3.23	-1.82	-1.00	-0.28	0.78	2.20	5.57	2749.51	2104.22
$r_{i,t}$	0.01	0.13	2.82	55.44	-0.61	-0.28	-0.16	-0.05	0.00	0.06	0.20	0.42	1.93	2749.51	1262.86
$r^i_{t-5,t}$	0.25	0.90	-0.85	7.65	-5.60	-2.48	-1.35	-0.19	0.35	0.79	1.51	2.18	3.58	2749.51	896.64
$r^{i,BE}_{t-5,t}$	0.73	1.02	-0.12	10.94	-6.09	-2.26	-0.79	0.24	0.67	1.24	2.33	3.42	7.11	2749.51	484.21
$r^{i,RE}_{t-5,t}$	0.73	0.94	0.99	10.96	-4.99	-1.37	-0.25	0.16	0.52	1.19	2.41	3.69	8.10	2749.51	215.88
$r^{i,CC}_{t-5,t}$	0.65	1.02	0.39	12.30	-6.80	-2.01	-0.62	0.09	0.49	1.11	2.38	3.69	8.22	2749.51	387.38
ind^i_{RE}	0.25	0.40	1.57	5.15	0.00	0.00	0.00	0.00	0.00	0.60	0.96	1.00	1.00	2749.51	0.00
ind^i_{CC}	0.07	0.21	4.15	72.86	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.73	0.75	2749.51	0.00
$r^{i,TB}_{t-5,t}$	0.25	0.48	-0.31	10.25	-2.98	-1.23	-0.52	0.03	0.26	0.49	0.96	1.48	3.21	2749.51	836.67
$r^{i,IB}_{t-5,t}$	0.00	0.76	-0.59	6.43	-4.77	-2.20	-1.31	-0.40	0.06	0.45	1.12	1.73	3.07	2749.51	1266.15
$r^{i,TR}_{t-5,t}$	0.25	0.33	0.52	12.73	-2.06	-0.63	-0.16	0.09	0.18	0.41	0.81	1.21	2.68	2749.51	806.31
$r^{i,IR}_{t-5,t}$	0.00	0.83	-0.87	8.51	-5.97	-2.55	-1.45	-0.40	0.08	0.48	1.17	1.84	3.31	2749.51	1233.22
$r^{i,TC}_{t-5,t}$	0.25	0.22	-0.16	10.73	-1.41	-0.39	-0.09	0.15	0.24	0.36	0.59	0.84	1.69	2749.51	775.43
$r^{i,IC}_{t-5,t}$	0.00	0.87	-0.77	6.15	-5.29	-2.63	-1.57	-0.43	0.10	0.53	1.23	1.84	3.20	2749.51	1217.47
$r^{i,TT}_{t-5,t}$	0.32	0.49	-0.36	9.54	-3.09	-1.17	-0.46	0.11	0.32	0.57	1.07	1.58	3.26	2749.51	545.67
$r^{i,IT}_{t-5,t}$	0.00	0.74	-0.48	6.28	-4.60	-2.09	-1.25	-0.39	0.04	0.43	1.12	1.75	3.14	2749.51	1293.08

Note: This table displays the summary statistics for the full sample with retained earnings (RE) and contributed capital (CC) indicator (value < 0) based on monthly construction, from July 1969 to December 2020. The table contains all stock information for the first three rows, Stock price, number of shares in thousands, market values of the equity in a million. Then BE, RE and CC represents the book equity, retained earnings and contributed capital, which are the fundamental measurement collected from the end of each year for earnings. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ are the book-to-market, retained earnings-to-market and contributed capital-to-market ratio constructed by the value information collected at the prior fiscal year divided over the ME at the end of each month. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the prior five-year lagged ratios collected at the end of year$t-5$. All returns with $r^{i,*}_{t-5,t}$ are the five years specific returns from the last trading day of year $t-6$ to the last trading day of year t. $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$ are the indicators to distinguish the negative and positive value of retained earning and contributed capital. T and I denote the tangible and intangible, respectively.

2.3.3.2 Comparison of data with and without data filtering

In this part, I add the data filtering criteria. Table 2.4 reports the summary statistics by following DT to remove firms with share prices below 5 dollars.

This scenario reduces the average number of firms to 2252, which indicates that about 500 firms are excluded from the data sample each period. ME, BE, RE, and CC significantly increase by about 21%, 19.5%, 23% and 13.3%, respectively. The low stock price affects the value of RE remarkably that only 18% of RE is negative compared to 25% if stock prices under 5 dollars are included. However, the percentage of CC has increased by 1%. This result interprets that CC is typically positive when RE is negative. Erasing the negative RE cannot remove the negative CC simultaneously but the positive CC. It results in a decrease in the mean of $ind_{i,t}^{RE}$ and an increase in $ind_{i,t}^{CC}$. Besides, all six timely and past value-to-price ratios are decreased, except an increased $bm_{i,t-5}$. Apart from the above findings, the average long-term stock returns grow substantially due to the removal of extreme low stock prices.

From the analysis of the ME distribution in Table 2.1, adjusting the data sample by the 20th percentile of the NYSE market cap can exclude nearly half of the available data obtained by 5 dollars share price threshold.

Table 2.5 summarise the average information of firms of the All-but-micro firms. The average number of All-but-micro firms is 1421. The average share price increases to around 53, and the number of shares, ME and three value variables (BE, RE and CC) are almost doubled compared to the data sample without a data filtering criterion. The performance of other variables is broadly similar and greater than those in the data sample containing just stock prices exceeding 5 dollars. Notably, $bm_{i,t-5}$ decreases to -0.54 as well as the other ratios and does not increase as in Table 2.4. Additionally, the percentage of negative RE and CC further changes in line with the trend in the case of share prices no

Table 2.4: Summary statistics of the annual basis of returns and accounting information with share prices greater than 5 dollars

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Prices	38.81	501.55	29.96	1383.88	5.00	5.26	6.44	12.57	22.04	36.64	72.84	129.79	25750.11	2252.28	0.00
Shares	78181.18	242269.26	9.49	137.39	162.18	997.39	2287.70	8737.18	21615.44	58416.73	299380.52	949357.23	4894468.14	2252.28	0.00
ME	3325.90	12278.43	11.55	196.78	2.77	11.62	29.25	147.36	530.92	1936.52	13399.18	53334.54	262200.10	2252.28	0.00
BE	1777.08	6026.95	10.81	192.41	0.55	7.00	19.31	88.63	288.90	1013.25	7575.63	27349.66	110615.74	2252.28	0.00
RE	1184.99	5340.26	10.55	181.85	-40666.04	-1053.05	-214.78	14.88	111.77	559.91	5520.95	21852.38	126399.85	2252.28	310.63
CC	637.51	3418.90	13.65	458.70	-52696.52	-2286.62	-165.33	22.73	105.98	423.06	3021.41	10925.72	84712.25	2252.28	165.47
$bm_{i,t}$	-0.39	0.90	1.04	11.77	-6.14	-2.47	-1.71	-0.85	-0.38	0.01	0.80	2.79	6.08	2252.28	1668.43
$rm_{i,t}$	-0.68	0.97	0.15	9.25	-7.03	-3.23	-2.20	-1.22	-0.62	-0.05	0.36	2.28	5.67	2252.28	1549.91
$cm_{i,t}$	-1.19	1.30	-0.52	6.56	-9.36	-4.99	-3.37	-1.89	-1.06	-0.33	0.47	2.13	5.14	2252.28	1807.85
$iss^i_{t-5,t}$	0.01	0.41	1.16	28.13	-3.49	-0.85	-0.41	-0.19	-0.04	0.14	0.67	1.44	4.02	2252.28	1263.21
$bm_{i,t-5}$	-0.41	0.94	0.95	11.71	-6.27	-2.58	-1.78	-0.89	-0.40	0.01	0.84	2.78	6.49	2252.28	1613.62
$rm_{i,t-5}$	-0.81	1.07	-0.14	8.98	-7.64	-3.82	-2.56	-1.37	-0.71	-0.11	0.31	2.21	5.95	2252.28	1556.77
$cm_{i,t-5}$	-1.10	1.26	-0.28	6.49	-8.41	-4.57	-3.18	-1.79	-1.01	-0.31	0.59	2.20	5.51	2252.28	1755.51
$r_{i,t}$	0.01	0.10	1.09	15.32	-0.52	-0.23	-0.14	-0.04	0.01	0.06	0.18	0.32	0.98	2252.28	1018.08
$r^i_{t-5,t}$	0.48	0.73	-0.17	6.33	-3.40	-1.37	-0.69	0.06	0.49	0.90	1.65	2.32	3.66	2252.28	522.13
$r^{i,BE}_{t-5,t}$	0.88	0.90	0.48	12.74	-4.70	-1.41	-0.29	0.37	0.76	1.32	2.41	3.48	7.00	2252.28	251.61
$r^{i,RE}_{t-5,t}$	0.85	0.94	1.14	11.12	-4.20	-1.08	-0.18	0.23	0.69	1.31	2.51	3.82	8.10	2252.28	156.22
$r^{i,CC}_{t-5,t}$	0.72	1.02	0.50	12.31	-6.47	-1.82	-0.45	0.11	0.55	1.19	2.45	3.79	8.08	2252.28	280.30
ind^i_{RE}	0.17	0.35	2.39	9.66	0.00	0.00	0.00	0.00	0.00	0.38	0.81	1.00	1.00	2252.28	0.00
ind^i_{CC}	0.08	0.22	3.75	62.96	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.73	0.75	2252.28	0.00
$r^{i,TB}_{t-5,t}$	0.48	0.42	0.21	10.47	-2.10	-0.61	-0.12	0.25	0.45	0.69	1.18	1.67	3.15	2252.28	351.65
$r^{i,IB}_{t-5,t}$	0.00	0.58	-0.19	6.02	-3.22	-1.50	-0.96	-0.34	0.01	0.35	0.94	1.46	2.56	2252.28	1105.83
$r^{i,TR}_{t-5,t}$	0.48	0.33	0.54	10.94	-1.58	-0.35	0.05	0.31	0.43	0.65	1.05	1.46	2.72	2252.28	299.45
$r^{i,IR}_{t-5,t}$	0.00	0.64	-0.21	8.14	-4.07	-1.64	-1.01	-0.35	0.00	0.35	1.02	1.67	3.06	2252.28	1114.40
$r^{i,TC}_{t-5,t}$	0.48	0.26	-0.20	9.78	-1.22	-0.22	0.09	0.35	0.46	0.62	0.91	1.20	1.98	2252.28	255.94
$r^{i,IC}_{t-5,t}$	0.00	0.67	-0.24	5.13	-3.46	-1.76	-1.12	-0.39	0.02	0.40	1.06	1.63	2.77	2252.28	1092.16
$r^{i,TT}_{t-5,t}$	0.44	0.43	0.41	11.20	-2.13	-0.65	-0.15	0.20	0.39	0.65	1.16	1.65	3.19	2252.28	263.57
$r^{i,IT}_{t-5,t}$	0.00	0.57	-0.20	6.29	-3.23	-1.47	-0.94	-0.33	0.01	0.34	0.92	1.43	2.55	2252.28	1102.98

Note: This table displays the summary statistics for the data samples without the share prices lower than five dollars, from July 1969 to December 2020. The table contains all stock information for the first three rows, Stock price, number of shares in thousands, market values of the equity in millions. Then BE, RE and CC represent the book equity, retained earnings and contributed capital, which are the fundamental measurement collected from the end of each year for earnings. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ are the book-to-market, retained earnings-to-market and contributed capital-to-market ratio collected at the end of year t. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the prior five-year lagged ratios collected at the end of year$t-5$. All returns with $r^{i,*}_{t-5,t}$ are the five years specific returns from the last trading day of year $t-6$ to the last trading day of year t. $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$ are the indicators to distinguish the negative and positive value of retained earning and contributed capital. T and I denote the tangible and intangible, respectively.

Table 2.5: Summary statistics of the annual basis returns and accounting information for All-but-Micro data samples

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
Prices	52.85	647.18	24.16	862.60	1.91	5.19	9.63	19.51	29.83	44.51	82.94	156.15	25750.11	1421.11	0.00
Shares	122519.52	307630.31	7.98	99.13	1104.58	5816.27	10274.49	22420.88	44289.85	104563.11	441636.19	1368375.91	5249873.20	1421.11	0.00
ME	5132.56	15025.16	9.50	132.79	277.72	285.69	321.41	587.61	1293.70	3682.25	20026.27	70379.37	262200.10	1421.11	0.00
BE	2627.42	7277.49	8.95	130.41	5.12	41.91	103.60	287.40	639.45	1843.70	10695.08	36482.50	110060.87	1421.11	0.00
RE	1744.95	6528.51	8.63	123.18	-41544.16	-1815.80	-341.03	76.65	315.20	1098.71	8176.88	28913.95	126399.85	1421.11	162.99
CC	954.02	4284.06	10.90	294.78	-52696.23	-3371.24	-398.36	63.76	246.15	770.43	4334.00	14876.94	84712.25	1421.11	129.63
$bm_{i,t}$	-0.55	0.80	0.62	9.42	-4.81	-2.51	-1.79	-0.98	-0.52	-0.15	0.51	2.04	4.39	1421.11	1149.77
$rm_{i,t}$	-0.82	0.90	-0.14	7.78	-6.21	-3.21	-2.25	-1.34	-0.79	-0.17	0.16	1.63	4.16	1421.11	1068.43
$cm_{i,t}$	-1.37	1.29	-0.85	5.84	-8.89	-5.34	-3.58	-2.09	-1.22	-0.43	0.19	1.32	3.15	1421.11	1162.31
$iss^i_{t-5,t}$	0.02	0.39	2.03	23.76	-2.32	-0.72	-0.38	-0.18	-0.05	0.15	0.68	1.42	3.75	1421.11	809.78
$bm_{i,t-5}$	-0.54	0.86	0.75	10.73	-4.98	-2.60	-1.85	-0.99	-0.51	-0.13	0.61	2.21	5.19	1421.11	1114.06
$rm_{i,t-5}$	-0.93	1.02	-0.28	8.29	-7.09	-3.80	-2.61	-1.47	-0.86	-0.19	0.15	1.71	4.79	1421.11	1067.57
$cm_{i,t-5}$	-1.26	1.23	-0.58	6.09	-8.01	-4.84	-3.37	-1.96	-1.14	-0.38	0.26	1.55	4.24	1421.11	1137.27
$r_{i,t}$	0.01	0.09	0.75	12.93	-0.44	-0.22	-0.13	-0.04	0.01	0.06	0.16	0.28	0.72	1421.11	636.63
$r^i_{t-5,t}$	0.56	0.69	0.00	5.33	-2.63	-1.22	-0.54	0.17	0.55	0.94	1.68	2.37	3.61	1421.11	270.74
$r^{i,BE}_{t-5,t}$	1.02	0.89	1.00	12.71	-3.36	-0.91	-0.09	0.47	0.89	1.46	2.59	3.74	6.94	1421.11	116.04
$r^{i,RE}_{t-5,t}$	0.99	0.96	1.27	12.30	-3.77	-0.85	-0.05	0.30	0.83	1.45	2.69	4.04	7.94	1421.11	76.50
$r^{i,CC}_{t-5,t}$	0.84	1.05	0.75	12.20	-5.41	-1.57	-0.30	0.18	0.68	1.33	2.65	4.07	8.05	1421.11	137.97
ind^i_{RE}	0.14	0.31	3.40	20.59	0.00	0.00	0.00	0.00	0.00	0.28	0.65	0.97	1.00	1421.11	0.00
ind^i_{CC}	0.10	0.24	2.33	7.28	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.73	0.73	1421.11	0.00
$r^{i,TB}_{t-5,t}$	0.56	0.39	0.66	9.67	-1.33	-0.32	0.01	0.32	0.52	0.76	1.24	1.71	2.92	1421.11	154.44
$r^{i,IB}_{t-5,t}$	0.00	0.56	-0.28	6.02	-2.88	-1.51	-0.92	-0.31	0.01	0.32	0.88	1.37	2.29	1421.11	693.69
$r^{i,TR}_{t-5,t}$	0.56	0.33	0.63	11.91	-1.34	-0.22	0.12	0.38	0.50	0.72	1.13	1.54	2.58	1421.11	124.94
$r^{i,IR}_{t-5,t}$	0.00	0.60	-0.07	7.14	-3.29	-1.59	-0.94	-0.32	0.00	0.32	0.95	1.60	2.86	1421.11	706.84
$r^{i,TC}_{t-5,t}$	0.56	0.24	-0.11	9.27	-0.71	-0.06	0.20	0.43	0.54	0.68	0.96	1.23	1.85	1421.11	85.55
$r^{i,IC}_{t-5,t}$	0.00	0.64	-0.22	5.03	-3.09	-1.73	-1.05	-0.35	0.01	0.37	1.00	1.57	2.58	1421.11	695.76
$r^{i,TT}_{t-5,t}$	0.49	0.40	0.82	10.86	-1.47	-0.41	-0.04	0.24	0.43	0.69	1.19	1.68	2.94	1421.11	121.95
$r^{i,IT}_{t-5,t}$	0.00	0.54	-0.25	6.14	-2.82	-1.46	-0.90	-0.30	0.01	0.31	0.87	1.36	2.26	1421.11	692.61

Note: This table displays the summary statistics for the data samples only including the firms' ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. The table contains all stock information for the first three rows, Stock price, number of shares in thousands, market values of the equity in millions. Then BE, RE and CC represent the book equity, retained earnings and contributed capital, which are the fundamental measurement collected from the end of each year for earnings. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ are the book-to-market, retained earnings-to-market and contributed capital-to-market ratio collected at the end of year t. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the prior five-year lagged ratios collected at the end of year$t-5$. All returns with $r^{i,*}_{t-5,t}$ are the five years specific returns from the last trading day of year $t-6$ to the last trading day of year t. $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$ are the indicators to distinguish the negative and positive value of retained earning and contributed capital. T and I denote the tangible and intangible, respectively.

less than 5 dollars where $ind_{i,t}^{RE}$ decrease and $ind_{i,t}^{CC}$ increase.

2.3.3.3 Comparing specific years with DT

Following DT, Table 2.6 shows the cross-sectional summary statistics of some specific years. Besides the first five years (1969, 1977, 1978, 1990 and 2000), I add 2010 and 2020. The data collection method is consistent with previous tables.

Table 2.6 Panel A displays the cross-sectional summary statistics on annual data. This chapter differs from DT in that the first three periods (1969, 1977 and 1978) contain few firms than DT, particularly 1969 and 1978. I have 496 and 1517 firms, while DT collect 1030 and 2463 firms, respectively. The average stock price in 2000 is 43.66 dollars compared with 29.73 in DT. From 1978 to 2020, the issuance effect has an upward and then downward trend from -0.12 to 0.15 and then 0.04. The interesting finding is that negative indicators, $ind_{i,t}^{RE}$ and $ind_{i,t}^{CC}$, have a general increasing trend. This result shows that the proportion of firms with negative RE and CC increases over time. In 2020, the number will reach a maximum of 32% and 15%. Three value ratios lack a clear direction and fluctuate between -1.71 and 0.06, with RTM and CTM always less than one. There is no statistically significant difference between other variables with DT data samples.

Then, I perform the same process with monthly variables in Table 2.6 Panel B. The difference is that monthly updated data collects all information at the end of June timely without lag. The result shows a slight disparity between the monthly and annual data. Most notable point is the large gap in $r_{t-5,t}^{i}$ between annual basis (0.10) and monthly basis (0.58) in 1978. There is also some variable variation for other periods, but the difference is not as large as in 1978.

Table 2.6: Summary statistics for specific years on different basis

Panel A: Annual basis

	Prices	Shares	ME	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$iss^i_{t-5,t}$	$r^i_{t-5,t}$	ind^i_{RE}	ind^i_{CC}	$r^{i,TB}_{t-5,t}$	$r^{i,IB}_{t-5,t}$	$r^{i,TR}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	$r^{i,TC}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	$r^{i,TT}_{t-5,t}$	$r^{i,IT}_{t-5,t}$
For 1969(496)																		
Mean	34.14	13082.98	594.82	-0.71	-1.25	-1.71	-0.04	0.81	0.04	0.00	0.81	0.00	0.81	0.00	0.81	0.00	0.48	0.00
Std	23.12	19999.31	2103.21	0.66	0.77	0.89	0.27	0.61	0.19	0.00	0.40	0.46	0.37	0.48	0.34	0.50	0.35	0.45
Min	5.00	109.00	2.73	-3.68	-4.99	-4.78	-0.66	-0.96	0.00	0.00	-0.78	-1.90	-0.53	-1.97	-0.28	-1.02	-1.03	-1.95
Max	337.75	285794.00	38154.94	5.48	5.29	3.75	2.27	2.98	1.00	0.00	2.55	1.50	2.62	1.75	2.45	1.72	2.18	1.46
For 1977(1427)																		
Mean	23.06	15575.85	523.79	0.00	-0.45	-1.27	-0.10	0.14	0.03	0.00	0.14	0.00	0.14	0.00	0.14	0.00	0.56	0.00
Std	16.85	31751.64	1918.99	0.57	0.65	0.99	0.23	0.59	0.18	0.00	0.39	0.44	0.37	0.45	0.24	0.54	0.35	0.43
Min	5.00	313.00	1.57	-1.77	-4.55	-5.52	-1.78	-1.76	0.00	0.00	-1.28	-2.47	-1.51	-2.39	-1.10	-1.85	-1.16	-2.45
Max	264.00	615528.00	39213.24	3.96	3.90	2.59	1.84	2.44	1.00	0.00	2.19	1.72	1.88	2.43	1.23	2.07	2.40	1.76
For 1978(1517)																		
Mean	22.62	15819.60	492.48	0.06	-0.39	-1.22	-0.12	0.10	0.04	0.00	0.10	0.00	0.10	0.00	0.10	0.00	0.58	0.00
Std	15.71	32349.45	1761.07	0.56	0.67	1.00	0.22	0.61	0.21	0.00	0.43	0.43	0.40	0.46	0.28	0.54	0.38	0.42
Min	5.00	517.00	4.99	-3.08	-5.39	-6.81	-1.77	-2.29	0.00	0.00	-1.90	-2.94	-2.51	-2.70	-1.45	-2.22	-1.67	-2.85
Max	257.25	652884.00	39091.43	4.25	4.05	2.56	0.85	3.17	1.00	0.00	2.50	1.89	2.47	2.68	1.46	2.03	3.67	1.98
For 1990(2067)																		
Mean	28.52	36871.56	1381.24	-0.47	-0.82	-1.31	-0.03	0.68	0.12	0.12	0.68	0.00	0.68	0.00	0.68	0.00	0.44	0.00
Std	159.68	81187.19	4213.30	0.86	0.99	1.28	0.48	0.68	0.33	0.32	0.41	0.53	0.35	0.58	0.26	0.62	0.42	0.53
Min	5.00	79.00	2.24	-6.48	-7.30	-7.40	-7.25	-3.42	0.00	0.00	-1.68	-2.00	-1.48	-3.91	-0.51	-3.29	-1.58	-2.19
Max	7200.00	1250000.00	67527.25	6.30	6.18	5.29	2.64	3.57	1.00	1.00	2.92	3.93	2.50	3.08	1.90	2.45	2.90	4.00
For 2000(2956)																		
Mean	43.66	101486.75	4591.13	-0.58	-0.72	-1.17	0.15	0.67	0.28	0.09	0.67	0.00	0.67	0.00	0.67	0.00	0.53	0.00
Std	989.44	392808.25	24298.06	1.16	1.15	1.46	0.53	0.85	0.45	0.29	0.51	0.68	0.31	0.79	0.36	0.77	0.53	0.68
Min	5.00	10.00	0.07	-7.70	-7.70	-9.36	-2.25	-2.96	0.00	0.00	-2.43	-2.33	-0.73	-3.44	-1.78	-2.76	-2.52	-2.51
Max	53800.00	9893426.00	524351.58	7.42	7.17	7.32	5.73	5.81	1.00	1.00	3.83	4.61	2.94	5.34	2.81	4.81	3.95	4.53
For 2010(2549)																		
Mean	74.43	163412.50	4868.55	-0.39	-0.58	-0.86	0.05	0.04	0.27	0.14	0.04	0.00	0.04	0.00	0.04	0.00	0.31	0.00
Std	2376.66	540158.03	16170.56	0.99	1.01	1.42	0.41	0.69	0.45	0.35	0.39	0.57	0.30	0.62	0.16	0.67	0.42	0.56
Min	5.00	238.00	4.23	-8.01	-8.01	-13.38	-2.70	-3.67	0.00	0.00	-2.59	-2.17	-2.11	-3.63	-1.73	-3.26	-2.70	-2.16
Max	120000.00	10676518.00	290959.69	5.62	4.59	4.42	3.93	2.68	1.00	1.00	1.55	3.03	1.68	2.66	0.68	2.75	2.44	3.02
For 2020(2423)																		
Mean	60.19	184988.10	12533.56	-0.75	-0.69	-1.07	0.04	0.32	0.32	0.15	0.32	0.00	0.32	0.00	0.32	0.00	0.19	0.00
Std	140.24	518234.14	61823.21	1.14	1.13	1.63	0.46	0.77	0.47	0.35	0.43	0.64	0.19	0.74	0.20	0.74	0.43	0.63
Min	5.00	136.00	7.25	-7.37	-8.60	-13.75	-2.94	-5.52	0.00	0.00	-4.37	-3.54	-1.84	-5.75	-1.36	-5.20	-4.64	-3.52
Max	3258.75	8747092.00	1562780.89	6.29	6.32	4.81	5.12	3.32	1.00	1.00	2.18	3.05	1.75	3.10	1.44	3.05	2.41	3.17

Panel B: Monthly basis

	Prices	Shares	ME	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$iss^i_{t-5,t}$	$r^i_{t-5,t}$	ind^i_{RE}	ind^i_{CC}	$r^{i,TB}_{t-5,t}$	$r^{i,IB}_{t-5,t}$	$r^{i,TR}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	$r^{i,TC}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	$r^{i,TT}_{t-5,t}$	$r^{i,IT}_{t-5,t}$
For 1969(494)																		
Mean	31.62	13134.37	597.12	-0.54	-1.08	-1.54	-0.03	0.56	0.03	0.00	0.56	0.00	0.56	0.00	0.56	0.00	0.42	0.00
Std	22.23	20036.60	2107.16	0.68	0.78	0.91	0.28	0.55	0.18	0.00	0.32	0.45	0.31	0.46	0.26	0.48	0.30	0.44
Min	5.13	109.00	2.73	-3.12	-4.55	-4.47	-0.66	-0.54	0.00	0.00	-0.69	-1.40	-0.48	-1.56	-0.31	-1.15	-0.76	-1.48
Max	324.50	285794.00	38154.94	5.75	5.55	4.02	2.27	2.91	1.00	0.00	1.95	1.62	1.98	1.79	1.90	1.96	1.91	1.59
For 1977(1427)																		
Mean	22.50	15737.89	523.78	-0.06	-0.50	-1.33	-0.10	0.14	0.03	0.00	0.14	0.00	0.14	0.00	0.14	0.00	0.49	0.00
Std	16.53	32711.05	1918.99	0.56	0.66	0.95	0.24	0.62	0.18	0.00	0.41	0.46	0.38	0.49	0.27	0.55	0.31	0.45
Min	5.00	400.00	3.84	-2.06	-4.50	-5.33	-1.80	-1.92	0.00	0.00	-1.29	-2.34	-1.70	-2.33	-1.26	-1.88	-1.03	-2.32
Max	268.13	628415.00	39213.24	3.42	3.35	2.37	1.83	2.42	1.00	0.00	2.16	1.76	1.98	2.27	1.11	2.01	2.08	1.78
For 1978(1524)																		
Mean	23.65	15876.02	490.30	-0.06	-0.50	-1.34	-0.13	0.58	0.05	0.00	0.58	0.00	0.58	0.00	0.58	0.00	0.49	0.00
Std	16.54	32354.16	1757.31	0.57	0.68	0.99	0.23	0.59	0.21	0.00	0.38	0.46	0.35	0.48	0.24	0.54	0.33	0.43
Min	5.00	517.00	3.68	-3.41	-5.47	-7.06	-1.77	-1.50	0.00	0.00	-1.07	-2.53	-1.82	-2.42	-0.39	-2.23	-1.43	-2.37
Max	281.00	652884.00	39091.43	3.82	3.62	2.51	0.85	2.89	1.00	0.00	2.48	1.73	2.37	2.50	1.32	1.99	2.97	1.70
For 1990(2045)																		
Mean	27.94	37605.62	1395.56	-0.48	-0.81	-1.33	-0.04	0.53	0.12	0.12	0.53	0.00	0.53	0.00	0.53	0.00	0.36	0.00
Std	155.99	84434.80	4233.64	0.90	1.00	1.30	0.48	0.67	0.32	0.32	0.33	0.58	0.28	0.61	0.20	0.64	0.35	0.58
Min	5.00	79.00	2.24	-6.54	-7.27	-7.62	-7.24	-4.73	0.00	0.00	-1.50	-3.22	-1.32	-5.16	-0.45	-4.48	-1.35	-3.46
Max	7000.00	1250000.00	67527.25	6.17	6.06	5.24	2.69	4.08	1.00	1.00	2.67	3.34	2.31	3.65	1.60	2.56	2.44	3.48
For 2000(2937)																		
Mean	43.93	104249.18	4618.23	-0.61	-0.71	-1.21	0.14	0.55	0.27	0.10	0.55	0.00	0.55	0.00	0.55	0.00	0.43	0.00
Std	1016.63	412108.54	24374.17	1.26	1.20	1.48	0.55	0.83	0.45	0.29	0.40	0.73	0.23	0.79	0.30	0.77	0.43	0.72
Min	5.00	10.00	0.07	-7.78	-7.78	-9.43	-3.58	-2.75	0.00	0.00	-1.94	-2.47	-0.54	-3.18	-1.56	-2.66	-1.91	-2.38
Max	55100.00	9893426.00	524351.58	8.92	6.26	10.17	5.93	6.59	1.00	1.00	2.98	5.58	2.25	6.16	2.32	5.67	3.18	5.43
For 2010(2576)																		
Mean	74.19	162501.65	4819.78	-0.38	-0.57	-0.84	0.04	0.02	0.28	0.14	0.02	0.00	0.02	0.00	0.02	0.00	0.23	0.00
Std	2305.07	537666.18	16092.53	1.01	1.03	1.42	0.39	0.65	0.45	0.35	0.30	0.58	0.24	0.60	0.12	0.64	0.33	0.56
Min	5.00	238.00	4.23	-7.86	-7.86	-13.11	-2.65	-3.42	0.00	0.00	-2.02	-2.83	-1.71	-3.38	-1.19	-3.44	-2.17	-2.77
Max	117000.00	10676518.00	290959.69	5.75	5.03	5.16	2.90	2.77	1.00	1.00	1.23	2.27	1.33	2.81	0.49	2.73	1.97	2.29
For 2020(2440)																		
Mean	62.96	185347.58	12449.28	-0.55	-0.52	-0.89	0.03	0.10	0.33	0.14	0.10	0.00	0.10	0.00	0.10	0.00	0.17	0.00
Std	154.09	518060.31	61615.62	1.31	1.16	1.66	0.49	0.83	0.47	0.35	0.37	0.75	0.17	0.82	0.17	0.82	0.37	0.75
Min	5.01	136.00	3.05	-7.13	-8.09	-13.22	-3.14	-6.47	0.00	0.00	-3.89	-4.39	-1.79	-6.48	-1.34	-5.65	-3.85	-4.57
Max	3930.15	8753289.00	1562780.89	7.25	6.63	7.21	5.43	3.41	1.00	1.00	1.89	3.65	1.25	3.40	1.10	3.68	2.00	3.50

Note: This table displays the summary statistics for the observation in 1969, 1977, 1978, 1990, 2000, 2010 and 2020. Panel A shows the information with annual basis. Panel B gives information for the sample constructed by monthly basis. The table contains stock information for the first three rows, Stock price, number of shares in thousands, market values of the equity in millions. Then $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ are the book-to-market, retained earnings-to-market and contributed capital-to-market ratio collected at the end of year t. All returns with $(t-5,t)$ are the five years returns from the last trading day of year $t-6$ to the last trading day of year t. $ind^{RE}_{i,t}$ and $ind^{CC}_{i,t}$ are the indicators to distinguish the negative and positive value of retained earning and contributed capital.

2.3.3.4 Correlation information

Table 2.7 shows the average correlation coefficients between all available decomposed past information on annual basis (Table 2.7 Panel A) and monthly basis (Table 2.7 Panel B).

For both panels, the upper triangle above the diagonal shows the Pearson's correlation and the lower triangle displays the Spearman's ranking correlation. From the table, the higher frequency of data collection can provide more contemporaneous information, which can reinforce the relationship between past price-scaled ratios and the long-term value returns. For example, the absolute correlation coefficient of $bm_{i,t-5}$ and $r^{i,BE}_{t-5,t}$ in monthly updated data increases from 22.9% to 36.42%. However, three value returns lose the connection with the past five-year stock returns, where the correlation coefficient between the return on book equity and stock returns falls the most, to 16%. Like DT, the past five-year return ($r^{i}_{t-5,t}$) has a negative correlation with btm, rtm and ctm in period t while it has the same movements with the past information, such as $bm_{i,t-5}$ or $r^{i,BE}_{t-5,t}$. In all past variables, only $rm_{i,t-5}$ is negatively correlated with $r^{i}_{t-5,t}$ in both annual and monthly updated data. The coefficient closing to zero indicates that changes in $rm_{i,t-5}$ cannot affect the past long-term stock returns.

Table 2.8 presents the average correlation coefficients after excluding data with a share price below 5 dollars, which results in a higher correlation between variables relative to those in Table 2.7, which is consistent with the summary statistics. 7.9% of $rm_{i,t-5}$ has insignificantly positive correlation with $r^{i}_{t-5,t}$. In addition, if data filtering constraints are applied, the issuance effect may have a more significant positive correlation with long-run stock returns.

Table 2.7: Average correlation coefficients of decomposed components with all firms

Panel A: Panel A: Annual basis

Variables	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$
$bm_{i,t}$	100	61.48	-17.87	55.49	34.32	-18.71	52.78	33.65	-15.59	-38.21	-9.6
$bm_{i,t-5}$	53.77	100	-22.9	36.35	54.1	-12.29	30.85	56.8	-14.31	18.73	-7.21
$r^{i,BE}_{t-5,t}$	-24.29	-25.17	100	-25.87	-23.09	62.68	-34.72	-22.28	70.93	55.36	12.07
$rm_{i,t}$	47.78	27.53	-30.16	100	70.46	-28.99	36.66	29.69	-26.24	-30.62	-1.35
$rm_{i,t-5}$	26.07	47.81	-25.09	72.46	100	-47.16	28.53	36.35	-11.28	-0.55	6.33
$r^{i,RE}_{t-5,t}$	-19	-12.23	64.3	-42.23	-50.24	100	-34.97	-18.9	38.69	42.42	0
$cm_{i,t}$	48.5	23.95	-34.76	38.25	30.3	-40.39	100	71.28	-10.07	-47.47	19.53
$cm_{i,t-5}$	25.3	51.9	-22.69	28.74	36.66	-23.37	69.17	100	-30.06	5.78	15.59
$r^{i,CC}_{t-5,t}$	-16.01	-14.76	67.86	-25.75	-11.31	32	-16.39	-30.74	100	30.51	22.07
$r^{i}_{t-5,t}$	-43.63	18.34	50.96	-34	-1.02	42.76	-46.9	6.77	25.54	100	-14
$iss^{i}_{t-5,t}$	-10.36	-12.05	20.78	0.38	7.99	-6.29	17.01	8.79	33.13	-7.63	100

Panel B: Panel B: Monthly basis

Variables	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$
$bm_{i,t}$	100	55.86	-13.83	56.97	32	-16.47	56.32	31.62	-11.75	-44.37	-12.09
$bm_{i,t-5}$	49.18	100	-36.42	34.64	54.34	-21.21	31.24	59.44	-24.23	17.07	-8.06
$r^{i,BE}_{t-5,t}$	-19.84	-39.12	100	-23.9	-30.3	62.68	-31.72	-32.71	70.93	39.16	7.75
$rm_{i,t}$	49.92	27.2	-27.88	100	68.28	-27.68	39.72	29.63	-23.63	-34.35	-2.13
$rm_{i,t-5}$	25.12	49.07	-32.23	70.85	100	-55.19	29.03	39.08	-17.64	-1.12	5.9
$r^{i,RE}_{t-5,t}$	-16.74	-20.93	64.3	-40.5	-57.35	100	-33.15	-25.18	38.69	32.45	-1.69
$cm_{i,t}$	52.28	25.71	-31.68	40.57	30.74	-38.03	100	69.67	-7.95	-50.45	17.43
$cm_{i,t-5}$	24.22	55.22	-33.07	28.49	38.99	-28.49	68.06	100	-37.73	4.75	13.04
$r^{i,CC}_{t-5,t}$	-12.06	-24.58	67.86	-23.21	-16.74	32	-14.06	-38.3	100	17.58	19.86
$r^{i}_{t-5,t}$	-48.91	15.82	35.21	-37.39	-2.22	33.49	-49	5.59	13.77	100	-17.53
$iss^{i}_{t-5,t}$	-12.45	-12.66	17.57	0.05	7.91	-7.76	14.93	6.36	31.46	-10.36	100

Note: This table presents the Pearson average and Spearman rank correlation coefficients for accounting information and returns on stock or accounting measurement based on full data sample. Panel A shows the information with annual basis. Panel B gives information for the sample constructed by monthly basis. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ represent earnings-to-market ratios in year t and $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ is the lagged earning-to-market ratios in year $t-5$. $r^{i,BE}_{t-5,t}$ is the book equity return from the end of year $t-6$ to the end of year. $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$ have the similar meanings. $r^{i}_{t-5,t}$ is the past five-year cumulative stock returns over the same period. $iss^{i}_{t-5,t}$ represent the issuance effect for the last five year from $t-5$ to t.

Table 2.8: Average correlation coefficients of decomposed components without share prices lower than 5 dollars

Panel A: Annual basis

Variables	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$
$bm_{i,t}$	100	67.19	-20.79	61.87	40.89	-18.1	54.03	36.54	-14.95	-39.5	-10.2
$bm_{i,t-5}$	58.17	100	-26.36	43.86	62.66	-15.56	37.96	58.69	-14.92	15.32	-4.57
$r^{i,BE}_{t-5,t}$	-25.1	-27.28	100	-23.42	-20.11	65.08	-23.36	-18.74	72	48.68	28.73
$rm_{i,t}$	51.71	32.94	-26.65	100	71.75	-28.48	33.13	29.15	-25.82	-26.77	-8.07
$rm_{i,t-5}$	29.65	55.57	-20.9	72.84	100	-46.09	24.63	34.83	-10.01	7.9	0.73
$r^{i,RE}_{t-5,t}$	-18.24	-15.75	64.59	-40.5	-48.58	100	-27.5	-15.69	38.93	38.24	8.35
$cm_{i,t}$	47.28	28.27	-24.51	30.31	22.3	-30.98	100	74.61	-3.39	-33.25	11.55
$cm_{i,t-5}$	25.74	51.96	-19.06	25.73	32.76	-19.2	72.99	100	-27.88	12.35	10.91
$r^{i,CC}_{t-5,t}$	-14.28	-14.01	69.41	-26.22	-10.24	32.63	-12.28	-28.94	100	28.17	32.98
$r^{i}_{t-5,t}$	-44.2	17.15	45.43	-28.08	9.03	35.83	-34.09	13.32	24.02	100	1.17
$iss^{i}_{t-5,t}$	-11.98	-10.29	33.95	-8.7	-0.4	4.74	6.54	2.47	40.13	4.18	100

Panel B: Monthly basis

Variables	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$
$bm_{i,t}$	100	63.02	-15.58	61.45	37.92	-14.37	54.95	34.04	-11.23	-43.34	-15.21
$bm_{i,t-5}$	54.09	100	-39.03	41.47	62.52	-24.51	37.64	60.56	-24.2	15.38	-7.22
$r^{i,BE}_{t-5,t}$	-19.72	-40.6	100	-21.15	-28.26	65.36	-20.82	-28.92	71.9	29.43	25.06
$rm_{i,t}$	51.84	31.44	-24.04	100	70.21	-27.05	35.14	28.95	-23.25	-29.47	-9.69
$rm_{i,t-5}$	27.38	56.24	-29.1	71.65	100	-54.24	24.89	37.57	-16.67	8.79	-0.4
$r^{i,RE}_{t-5,t}$	-14.11	-24.5	65.02	-38.56	-56.08	100	-25.12	-22.07	39.17	24.83	6.9
$cm_{i,t}$	48.92	28.94	-21.78	31.63	22.31	-27.92	100	73.34	-1.91	-34.21	7.74
$cm_{i,t-5}$	23.79	54.5	-29.06	25.16	35.05	-24.39	71.83	100	-35.38	13.3	7.27
$r^{i,CC}_{t-5,t}$	-10.45	-23.28	69.36	-23.56	-16.3	32.97	-10.62	-36.25	100	14.01	31.23
$r^{i}_{t-5,t}$	-47.79	16.44	27.36	-30.23	9.19	23.56	-34.83	13.91	11.21	100	0.17
$iss^{i}_{t-5,t}$	-16.75	-12.54	30.8	-10.03	-1.28	3.61	2.94	-0.71	38.46	2.86	100

Note: This table presents the Pearson average and Spearman rank correlation coefficients for accounting information and returns on stock or accounting measurement only containing share prices greater than five dollars. Panel A shows the information with annual basis. Panel B gives information for the sample constructed by monthly basis. $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$ represent earnings-to-market ratios in year t and $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ is the lagged earning-to-market ratios in year $t-5$. $r^{i,BE}_{t-5,t}$ is the book equity return from the end of year $t-6$ to the end of year. $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$ have the similar meanings. $r^{i}_{t-5,t}$ is the past five-year cumulative stock returns over the same period. $iss^{i}_{t-5,t}$ represent the issuance effect for the last five year from $t-5$ to t.

2.4 Empirical results

In this section, I summarise the major findings of the empirical work based on two different groups of FM regressions. The first subsection runs FM regressions with decomposition until 2003 by following DT with share prices greater than 5 dollars. In the remaining two subsections, we will examine the FM regression covering the full period of time from July 1969 to December 2020. In the second subsection, I cover the complete data samples, and in the third subsection, I add the data filtering conditions. I construct 12 cross-sectional FM regressions on an annual basis for a rolling cycle during the 12 months from July in calendar year t to June in calendar year $t+1$. Every year, all values on the right will be rolled at the end of June. All tables in this part consist of four panels. Three panels consist of variables related to btm, rtm, and ctm, respectively, and Panel D shows the joint performance of three ratios.

2.4.1 Replication of DT until 2003

Table 2.9 presents the FM regression results from July 1969 to December 2003 according to the data sample only including share prices greater than 5 dollars.

Table 2.9 Panel A contains one current ratio and five long-term variables, $bm_{i,t}$, $bm_{i,t-5}$, $r^{i,BE}_{t-5,t}$, $r^{i}_{t-5,t}$, $iss^{i}_{t-5,t}$ and $r^{i,IB}_{t-5,t}$. According to the univariate regressions (1) to (5), all decompositions have significant explanatory power, except for the five-year lagged btm. The result is slightly inconsistent with DT, who find robust $bm_{i,t-5}$ and weak $r^{i,BE}_{t-5,t}$ at the 5% significant level. I also found a correlation between the past information and monthly stock returns. A positive relationship exists between $bm_{i,t}$ and $bm_{i,t-5}$, while the lagged return related variables (five-year book equity returns, stock returns, and issuance activities)

Table 2.9: FM regressions of monthly return on decomposition of price-scaled variables from 1969 to 2003

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.291	0.296					
	(5.11)	(3.64)					
(2)	1.257		0.109				
	(4.93)		(1.62)				
(3)	1.436			-0.194			
	(5.94)			(-3.38)			
(4)	1.326				-0.24		
	(5.47)				(-2.87)		
(5)	1.235					-0.606	
	(4.7)					(-3.42)	
(6)	1.403		0.051	-0.17			
	(5.85)		(0.8)	(-3.21)			
(7)	1.378		0.116	-0.042	-0.229		
	(6.06)		(1.67)	(-0.79)	(-2.25)		
(8)	1.33		0.057	-0.102		-0.493	
	(5.18)		(0.92)	(-2.12)		(-3.06)	
(9)	1.403		0.051	-0.17			-0.229
	(5.85)		(0.8)	(-3.21)			(-2.25)
(10)	1.296		0.139	0.062	-0.278	-0.591	
	(5.37)		(2.01)	(1.21)	(-2.8)	(-3.86)	

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.458	0.282						-0.466
	(5.88)	(4.59)						(-2.76)
(2)	1.323		0.078					-0.332
	(5.2)		(1.59)					(-1.99)
(3)	1.391			-0.106				-0.4
	(5.9)			(-2.02)				(-2.15)
(4)	1.368				-0.248			-0.351
	(5.78)				(-3)			(-2.37)
(5)	1.253					-0.564		-0.14
	(4.84)					(-3.42)		(-1.08)
(6)	1.382		0.047	-0.085				-0.391
	(5.79)		(1)	(-1.71)				(-2.09)
(7)	1.409		0.128	0.066	-0.301			-0.393
	(6.18)		(2.65)	(1.56)	(-3.28)			(-2.2)
(8)	1.306		0.041	-0.037		-0.502		-0.2
	(5.13)		(0.92)	(-0.85)		(-3.52)		(-1.31)
(9)	1.382		0.047	-0.085			-0.301	-0.391
	(5.79)		(1)	(-1.71)			(-3.28)	(-2.09)
(10)	1.322		0.122	0.122	-0.31	-0.551		-0.191
	(5.5)		(2.62)	(3.33)	(-3.48)	(-4.21)		(-1.28)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.271	0.036						0.091
	(4.8)	(0.99)						(1.15)
(2)	1.249		0.005					0.09
	(4.68)		(0.14)					(1.22)
(3)	1.427			-0.22				-0.011
	(5.78)			(-4.46)				(-0.2)
(4)	1.321				-0.241			0.078
	(5.42)				(-2.88)			(1.27)
(5)	1.239					-0.61		-0.039
	(4.7)					(-3.46)		(-0.78)
(6)	1.386		-0.046	-0.231				0.022
	(5.38)		(-1.38)	(-4.78)				(0.32)
(7)	1.39		-0.028	-0.176	-0.158			0.024
	(5.74)		(-0.84)	(-3.94)	(-1.81)			(0.37)
(8)	1.368		-0.012	-0.167		-0.412		-0.059
	(5.14)		(-0.36)	(-4)		(-2.48)		(-0.91)
(9)	1.386		-0.046	-0.231			-0.158	0.022
	(5.38)		(-1.38)	(-4.78)			(-1.81)	(0.32)
(10)	1.379		0.013	-0.095	-0.185	-0.49		-0.053
	(5.59)		(0.39)	(-2.45)	(-2.19)	(-3.16)		(-0.84)

Panel D: Combination information

Reg \Var	Constant	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	ind^{i}_{RE}	ind^{i}_{CC}
(1)	1.414	0.116	0.192					-0.404	
	(5.67)	(1.39)	(4.18)					(-2.76)	
(2)	1.127	0.437		-0.142					0.231
	(4.21)	(4.39)		(-3.37)					(1.8)
(3)	1.47		0.267	0.019				-0.493	-0.003
	(5.86)		(4.77)	(0.61)				(-2.94)	(-0.05)
(4)	1.292	0.233	0.133	-0.068				-0.223	0.142
	(4.76)	(1.98)	(2.31)	(-1.64)				(-2)	(1.14)
(5)	1.299				0.068	0.029		-0.294	
	(5.1)				(0.98)	(0.66)		(-1.88)	
(6)	1.184				0.165		-0.066		0.162
	(4.41)				(1.88)		(-1.48)		(1.43)
(7)	1.345					0.067	0.023	-0.353	0.023
	(5.21)					(1.46)	(0.77)	(-2.09)	(0.41)
(8)	1.285				0.078	0.024	-0.008	-0.288	0.063
	(4.75)				(0.89)	(0.54)	(-0.23)	(-2.23)	(0.69)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past annual variables with earning indicator. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$, past five-year earning returns, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$, five years stock returns $r^{i}_{t-5,t}$ and intangible returns, $r^{i,IB}_{t-5,t}$, $r^{i,IR}_{t-5,t}$ and $r^{i,IC}_{t-5,t}$. Panel A, B and C only include the variables under one measurement, respectively. Panel D combines the current and past earnings ratios together to compare the performance with the indicators. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment. Regressions only covers the firms with share prices greater than five dollars from July 1969 to December 2003.

may decrease future monthly stock returns. However, lagged BE returns can be more accurate in predicting monthly returns than five-year lagged stock returns (t-statistic $|-3.38| > |-2.87|$). In regression (6), $r^{i,BE}_{t-5,t}$ remains significant in capturing expected monthly stock returns and decreases the significance of $btm_{i,t-5}$. DT finds that $bm_{i,t-5}$ dominates the BE return. $bm_{i,t-5}$, $r^{i,BE}_{t-5,t}$ and $r^{i}_{t-5,t}$ are all significant in a single multivariate regression. In contrast, $r^{i,BE}_{t-5,t}$ will lose its the ability to explain the stock returns after adding $r^{i}_{t-5,t}$ in regression (7) (t-value, -0.79) or have marginally significant performance with $iss^{i}_{t-5,t}$ in regression (8) (t-value, -2.12). Because the intangible return is derived from the FM regression of the past return on book equity information, it has the same t-statistic with $r^{i}_{t-5,t}$ (-2.25). The result indicates that the decomposition cannot fully explain the monthly stock returns with a significant unexplained proportion in the FM regression. In regression (10), the decomposed variables all have marginal significant explanatory power in predicting future stock returns at 5% level compared, compared to strongly predictive power in DT. Generally speaking, the FM regression performance in this chapter is slightly inferior to that in DT, which may be caused by the different sample size of the regressions.

The contradiction is also present in the RE-based regressions in Table 2.9 Panel B. $rm_{i,t}$ with a significant negative indicator (t-value, -2.76) performs better than $bm_{i,t}$ for the same period (t-value $4.59 > 3.64$). $r^{i,RE}_{t-5,t}$ can also give marginally incremental explanatory power for predicting future stock returns at 5% significance levels. However, $rm_{i,t-5}$ with t-statistic of 1.59 cannot offer similar level of prediction as $rm_{i,t}$. Similarly, RE return dominates $rm_{i,t-5}$ in the multivariate regression with only the past stock return participating in the game of regressions (6) and (9). In regression (7), unlike BE-based regressions in Table 2.9 Panel A, both $rm_{i,t-5}$ and $r^{i}_{t-5,t}$ are significant on explaining the monthly returns at 5% levels. Nonetheless, past five-year lagged returns play a bigger role in predicting stock returns than BE-based regressions, which indicates that RE leaves more expected stock returns unexplained. According to the result, the decomposition *rtm* is superior to those of *btm* to measure value, which shows an opposite result with BGLN. Furthermore, the RE indicator, ind^i_{RE}, plays an important role on identifying the negatives and positives in the FM regressions, particularly when the issuance activities are not included. When $iss^i_{t-5,t}$ joins the equation, the significant level of $rm_{i,t-5}$ and $r^{i,RE}_{t-5,t}$ will depend on whether the regression contains $r^i_{t-5,t}$ or not. If prior long-term stock returns exist in regression (10), all four past decomposition components can have a significantly explanatory power on measuring monthly stock returns.

Table 2.9 Panel C presents the results of CC-based FM regressions, which differ from the first two Panels. Both $cm_{i,t}$ and $cm_{i,t-5}$ are completely inconsequential whereas $r^{i,CC}_{t-5,t}$ is highly important in predicting the expected stock return in all regressions. Surprisingly, CC return can absorb the performance of past five-year stock returns and issuance activities in regressions (7) to (9). Specifically, intangible return in regression (9) loses its explanatory power on future stock returns at the 5% significance level, indicating that past CC returns can absorb and fully explain the future stock returns without leaving a substantial unexplained portion in the regression. Moreover, the CC indicator is incapable of identifying negative CC due to the low percentage of negative CC and all positive in most periods. Briefly, *ctm* produces the best predictive performance whilst $r^{i,RE}_{t-5,t}$ in RE-based regression indicates the worst prediction of expected stock return after the effectiveness of *btm*.

Lastly, Table 2.9 Panel D displays the t-statistics when merging three price-scaled current and past ratios in a single regression. Regression (1) to (4) can support the finding of BGLN that $rm_{i,t}$ will absorb the significance of $bm_{i,t}$ and $cm_{i,t}$. The t-statistics for $rm_{i,t}$ with $bm_{i,t}$ in regression (1) and $cm_{i,t}$ in regression (3) are 4.18 and 4.77, respectively. Only $rm_{i,t}$ can have t-value larger than 2 when predicting the stock returns with all three current period ratios. The performance of $bm_{i,t}$ and $cm_{i,t}$ will be greatly enhanced in regression (2). Because $bm_{i,t}$ and $cm_{i,t}$ behave differently in the regression, BE related variables have the positive impact on the expected stock returns, while an increase in CC-related variables decreases the expected stock returns. Due to the highly positive correlation between $bm_{i,t}$ and $cm_{i,t}$, the performance of both variables will be extended in their initial directions. In regression (5) to (8), the link between past price-scaled ratios and timely ratios is inconsistent, such that $rm_{i,t-5}$ can still dominate $cm_{i,t-5}$ in regression (7), but cannot subsume $bm_{i,t-5}$ in regression (5). Furthermore, all lagged value-to-market ratios are meaningless in monthly stock returns.

2.4.2 Extended time range with full sample size

This subsection extends the cross-sectional time period to December 2020 and only runs regressions based on the full data sample.

2.4.2.1 All information with indicator

Table 2.10 replicates the FM regressions of monthly returns on the past decomposition components with the negative indicator (if available) after removing all data and period restriction. The FM regressions comprise all available data from July 1969 to December 2020.

Table 2.10 Panel A and Table 2.10 Panel B show the similar results that $bm_{i,t-5}$ and $rm_{i,t-5}$ are always non-significant in the univariate regressions. All other past returns are significant for capturing the future monthly stock returns. Meanwhile, $iss^i_{t-5,t}$ are considerably influenced by ind^i_{RE}, such that its t-statistics in RE-based regression are stronger than in BE-based regression (t-

Table 2.10: **FM regression of monthly return on decomposition of price-scaled variables with all available information**

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.279	0.25					
	(5.43)	(4.06)					
(2)	1.232		0.028				
	(5.32)		(0.53)				
(3)	1.435			-0.235			
	(5.86)			(-4.66)			
(4)	1.257				-0.297		
	(5.46)				(-3.54)		
(5)	1.266					-0.405	
	(5.45)					(-2.79)	
(6)	1.379		-0.035	-0.234			
	(5.76)		(-0.64)	(-4.42)			
(7)	1.293		0.061	-0.059	-0.276		
	(5.67)		(1.21)	(-1.64)	(-3.04)		
(8)	1.367		-0.038	-0.207		-0.343	
	(5.62)		(-0.75)	(-3.84)		(-2.45)	
(9)	1.379		-0.035	-0.234			-0.276
	(5.76)		(-0.64)	(-4.42)			(-3.04)
(10)	1.279		0.077	0.012	-0.34	-0.502	
	(5.53)		(1.54)	(0.29)	(-3.94)	(-4.01)	

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.346	0.204						-0.048
	(6.05)	(4.35)						(-0.31)
(2)	1.221		0.025					0.077
	(5.54)		(0.7)					(0.47)
(3)	1.332			-0.115				-0.034
	(6.09)			(-2.96)				(-0.22)
(4)	1.276				-0.301			-0.103
	(5.92)				(-3.85)			(-0.8)
(5)	1.204					-0.482		0.23
	(5.47)					(-4.04)		(1.76)
(6)	1.292		-0.019	-0.129				0.006
	(5.92)		(-0.52)	(-3.2)				(0.04)
(7)	1.285		0.086	0.058	-0.349			-0.13
	(6.01)		(2.4)	(1.68)	(-4.14)			(-0.89)
(8)	1.244		-0.02	-0.092		-0.449		0.184
	(5.48)		(-0.61)	(-2.36)		(-4.1)		(1.38)
(9)	1.292		-0.019	-0.129			-0.349	0.006
	(5.92)		(-0.52)	(-3.2)			(-4.14)	(0.04)
(10)	1.231		0.087	0.111	-0.377	-0.538		0.068
	(5.61)		(2.51)	(3.57)	(-4.59)	(-5.64)		(0.55)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.34	0.108						-0.031
	(5.15)	(2.38)						(-0.3)
(2)	1.283		0.022					-0.008
	(5.14)		(0.7)					(-0.09)
(3)	1.415			-0.221				-0.082
	(5.93)			(-5.64)				(-1.06)
(4)	1.254				-0.298			0.063
	(5.39)				(-3.58)			(1.04)
(5)	1.276					-0.416		-0.096
	(5.45)					(-2.91)		(-1.93)
(6)	1.393		-0.031	-0.231				-0.071
	(5.61)		(-1.05)	(-6.3)				(-0.79)
(7)	1.326		-0.011	-0.136	-0.251			0.008
	(5.54)		(-0.35)	(-3.78)	(-2.86)			(0.11)
(8)	1.404		-0.001	-0.19		-0.295		-0.155
	(5.71)		(-0.04)	(-5.23)		(-2.1)		(-2.22)
(9)	1.393		-0.031	-0.231			-0.251	-0.071
	(5.61)		(-1.05)	(-6.3)			(-2.86)	(-0.79)
(10)	1.364		0.039	-0.061	-0.3	-0.48		-0.097
	(5.72)		(1.38)	(-1.98)	(-3.57)	(-4.27)		(-1.62)

Panel D: Combination information

Reg \Var	Constant	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	ind^{i}_{RE}	ind^{i}_{CC}
(1)	1.275	0.187	0.069					0.049	
	(5.9)	(2.91)	(2.12)					(0.31)	
(2)	1.261	0.246		0.011					0.023
	(4.84)	(3.16)		(0.19)					(0.17)
(3)	1.426		0.164	0.089				-0.158	-0.066
	(5.86)		(3.95)	(2.48)				(-1.14)	(-0.88)
(4)	1.269	0.202	0.054	0.008				0.055	0.03
	(5.08)	(2.4)	(1.31)	(0.17)				(0.55)	(0.23)
(5)	1.208				0.023	0.007		0.073	
	(5.53)				(0.45)	(0.26)		(0.49)	
(6)	1.241				0.01		0.019		-0.017
	(4.97)				(0.13)		(0.39)		(-0.13)
(7)	1.22					0.018	0.008	0.07	0.041
	(5.37)					(0.56)	(0.32)	(0.43)	(0.78)
(8)	1.204				0.02	0.008	-0.001	0.077	0.046
	(5.18)				(0.29)	(0.24)	(-0.03)	(0.64)	(0.51)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past annual variables with earning indicator. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$, past five-year earning returns, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$, five years stock returns $r^{i}_{t-5,t}$ and intangible returns, $r^{i,IB}_{t-5,t}$, $r^{i,IR}_{t-5,t}$ and $r^{i,IC}_{t-5,t}$. Panel A, B and C only include the variables under one measurement, respectively. Panel D combines the current and past earnings ratios together to compare the performance with the indicators. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment. Regressions cover all available data from July 1969 to December 2020.

value, -4.04 vs -2.79).

Surprisingly, Table 2.10 Panel C is vastly better than the results of FM regression from 1969 to 2003. In the univariate regressions, all variables, except lagged ctm, predict the future stock returns at the 5% significant level, but the CC indicator is redundant in the regression. Notably, the five-year lagged CC return (t-value, -5.64) has the greatest explanatory power among all other components. But it is easily affected by the past stock returns and issuance activities. $r^{i,CC}_{t-5,t}$ t-statistic increases to -1.98 in regression (10), which has the comparable change to $r^{i,BE}_{t-5,t}$ (t-value, 0.29) in regression (10) of Table 2.10 Panel A.

Table 2.10 Panel D shows the results of multivariate regressions on timely or lagged ratios. Surprisingly, both $bm_{i,t}$ and $bm_{i,t-5}$ can dominate in all three price-scaled ratios, whereas $rm_{i,t}$ loses the dominant position in three timely price-scaled ratios in regression (4). BE will benefit greatly from firms with share prices lower than 5 dollars. $rm_{i,t}$ can have better performance than $cm_{i,t}$, but $rm_{i,t}$ cannot affect the significance of $cm_{i,t}$ (t-value, 2.48) in regression (3). The distribution of RE is skewed, which is consistent with the finding of summary statistics. None of the five-year lagged ratios has predictive power on monthly stock returns from regression (5) to (8), which means that the long-term mispricing may be rectified in the full sample estimation. Moreover, both RE and CC negative indicators cannot distinguish between negative and positive average values since the proportion of negative underlying earnings values in a given period is quite low.

2.4.2.2 Monthly updated data with indicator

Table 2.11 gives details of the FM regressions based on monthly updated data, with variables constructed at the end of each month.

Comparing the results to the first three panels on annual data, the construc-

Table 2.11: FM regression of monthly return on decomposition of price-scaled variables based on monthly basis

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i}_{t-5,t}$	$issm^{i}_{t-5,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.32	0.337					
	(5.56)	(4.68)					
(2)	1.289		0.107				
	(5.54)		(2.18)				
(3)	1.435			-0.235			
	(5.86)			(-4.66)			
(4)	1.207				-0.313		
	(5.16)				(-3.06)		
(5)	1.275					-0.475	
	(5.48)					(-3.28)	
(6)	1.403		0.014	-0.221			
	(5.79)		(0.26)	(-4)			
(7)	1.288		0.108	-0.074	-0.305		
	(5.52)		(2.01)	(-1.49)	(-2.75)		
(8)	1.397		0.004	-0.199		-0.429	
	(5.68)		(0.07)	(-3.62)		(-3.08)	
(9)	1.403		0.014	-0.221			-0.305
	(5.79)		(0.26)	(-4)			(-2.75)
(10)	1.279		0.121	-0.009	-0.375	-0.599	
	(5.41)		(2.28)	(-0.18)	(-3.47)	(-4.95)	

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i}_{t-5,t}$	$issm^{i}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.373	0.237						-0.075
	(5.97)	(4.26)						(-0.48)
(2)	1.282		0.071					0.016
	(5.82)		(2.03)					(0.09)
(3)	1.332			-0.115				-0.034
	(6.09)			(-2.96)				(-0.22)
(4)	1.222				-0.328			-0.086
	(5.53)				(-3.34)			(-0.66)
(5)	1.204					-0.565		0.264
	(5.46)					(-4.72)		(2.04)
(6)	1.326		0.026	-0.108				-0.028
	(6.05)		(0.72)	(-2.74)				(-0.17)
(7)	1.289		0.116	0.047	-0.37			-0.178
	(5.85)		(3.01)	(1.09)	(-3.49)			(-1.23)
(8)	1.271		0.019	-0.072		-0.529		0.192
	(5.57)		(0.57)	(-1.88)		(-4.74)		(1.43)
(9)	1.326		0.026	-0.108			-0.37	-0.028
	(6.05)		(0.72)	(-2.74)			(-3.49)	(-0.17)
(10)	1.227		0.116	0.102	-0.406	-0.636		0.059
	(5.42)		(3.08)	(2.49)	(-3.92)	(-6.64)		(0.47)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$issm^{i}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.379	0.155						-0.063
	(5.23)	(3.05)						(-0.61)
(2)	1.341		0.067					-0.038
	(5.35)		(2.12)					(-0.41)
(3)	1.415			-0.221				-0.082
	(5.93)			(-5.64)				(-1.06)
(4)	1.203				-0.314			0.06
	(5.1)				(-3.09)			(0.98)
(5)	1.286					-0.488		-0.11
	(5.48)					(-3.42)		(-2.18)
(6)	1.417		0.004	-0.213				-0.084
	(5.66)		(0.15)	(-6.09)				(-0.92)
(7)	1.325		0.016	-0.154	-0.288			-0.009
	(5.4)		(0.5)	(-4.26)	(-2.75)			(-0.12)
(8)	1.451		0.043	-0.16		-0.412		-0.2
	(5.85)		(1.65)	(-4.69)		(-2.89)		(-2.78)
(9)	1.417		0.004	-0.213			-0.288	-0.084
	(5.66)		(0.15)	(-6.09)			(-2.75)	(-0.92)
(10)	1.379		0.072	-0.071	-0.347	-0.587		-0.141
	(5.64)		(2.49)	(-2.07)	(-3.39)	(-5.03)		(-2.1)

Panel D: Combination information

Reg \Var	Constant	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	ind^{i}_{RE}	ind^{i}_{CC}
(1)	1.264	0.312	0.005					0.133	
	(5.77)	(4.07)	(0.14)					(0.82)	
(2)	1.3	0.335		0.014					0.065
	(4.97)	(4.21)		(0.24)					(0.48)
(3)	1.5		0.168	0.139				-0.236	-0.131
	(5.88)		(3.73)	(3.27)				(-1.68)	(-1.58)
(4)	1.184	0.383	-0.045	-0.023				0.245	0.18
	(4.75)	(4.45)	(-1.14)	(-0.51)				(2.23)	(1.49)
(5)	1.251				0.094	0.003		0.096	
	(5.72)				(1.9)	(0.12)		(0.63)	
(6)	1.305				0.075		0.031		-0.014
	(5.17)				(1.04)		(0.63)		(-0.11)
(7)	1.321					0.048	0.044	-0.01	-0.002
	(5.8)					(1.55)	(1.72)	(-0.06)	(-0.04)
(8)	1.252				0.085	0.004	0.006	0.103	0.058
	(5.4)				(1.37)	(0.14)	(0.19)	(0.83)	(0.68)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past annual variables with earning indicator. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$, past five-year earning returns, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$, five years stock returns $r^{i}_{t-5,t}$ and intangible returns, $r^{i,IB}_{t-5,t}$, $r^{i,IR}_{t-5,t}$ and $r^{i,IC}_{t-5,t}$. Panel A, B and C only include the variables under one measurement, respectively. Panel D combines the current and past earnings ratios together to compare the performance with the indicators. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment. The variables are constructed at the end of each month with monthly updated market equity. Regressions only covers all available data from July 1969 to December 2020.

tion alteration does not affect the lagged value returns because the accounting information can only be rebalanced annually. Yet, five-year lagged price-scaled ratios will be considerably improved so that all three ratios can significantly explain expected stock returns. $cm_{i,t-5}$ is even better than $rm_{i,t-5}$ (t-value, $2.12 > 2.03$). Also, issuance effects are slightly improved in all three scenarios. Conversely, monthly computed $r^i_(t-5, t)$ has less explanatory power in predicting the future stock returns than the annual construction in Table 2.10. Still, this decrease has no consequence on the significance level. Finally, Table 2.11 Panel D presents three primary outcomes. First, monthly updated data has no effect on the dominance of $bm_{i,t}$ in current and lagged periods. Second, five-year lagged ctm outperforms the lagged rtm (t-statistic, $1.55 < 1.72$). Third, ind^i_{RE} and ind^i_{CC} are insignificant in every instance, except when all current ratios are included in regression (4).

I can draw conclusions from the above FM regressions with the full data sample. First, the monthly calculation can enhance the significance of price-related variables. Second, btm is always more significant than other earning-to-market ratios. Third, the lagged earnings deflated by ME do not provide incremental explanatory power in most multivariate cases. Third, the lagged long-term stock return and intangible returns are often robust at the 1% significant level, implying that predicting the monthly stock returns by the past long-term variables will always preserve a sizeable proportion of unexplained returns.

2.4.3 Adding the data filter criterion

According to the preceding subsection, FM regressions with full data samples do not produce results compatible with the dominance of RTM in BGLN. In this part, I add the data filtering constraints on the data samples to test whether the data adjustment can yield more consistent results.

2.4.3.1 Only positive RE and CC without indicator

Table 2.12 displays the results of FM regression for firms with only positive RE and CC. Hence the FM regressions do not include the negative RE and CC indicators.

I have several common findings compared with the annual data without the restriction. First, the significance of three value ratios in the current period t declines, with the t-statistic of $cm_{i,t}$ falling from 2.38 to 1.75. In contrast, negative RE and CC deletion increase the predictive power of all prior five-year lagged value ratios and value returns for BE and CC-based variables. Third, it is impossible to raise the explanatory power of lagged ratios at the 5% significant level. The insignificance of ind^{i}_{RE} in Table 2.12 Panel B corresponds to the fact that no RE-related variables are affected by the data restriction process. Moreover, regression (10) in Table 2.12 Panel C shows that $r^{i,CC}_{t-5,t}$ (-3.73) can outperform the five-year lagged stock returns (-2.56) and even the issuance activities (-2.55). After decreasing the data distribution, Table 2.12 Panel D indicates that $rm_{i,t}$ totally holds the dominant position in three ratios. At the same time, the lagged ratios remain incapable of predicting the expected stock returns.

2.4.3.2 Annual data with only share prices larger than 5 dollars

Table 2.13 is the extension of Table 2.9 from DT's 2003 version. I remove the firms with share prices below 5 dollars and do FM regressions from July 1969 to December 2020 with negative RE and CC indicators.

From Table 2.13 Panel A of the BE-based regressions, the t-statistics of $bm_{i,t}$, $r^{i,BE}_{t-5,t}$ and $r^{i}_{t-5,t}$ decrease, but are still statistically significant. In contrast, $bm_{i,t-5}$ and $iss^{i}_{t-5,t}$ have greater explanatory power with the data sample restriction,

Table 2.12: FM regression of monthly return on decomposition of price-scaled variables with only positive RE and CC

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.219	0.198					
	(5.58)	(3.13)					
(2)	1.209		0.043				
	(5.55)		(0.89)				
(3)	1.444			-0.26			
	(6.51)			(-5.17)			
(4)	1.265				-0.256		
	(5.9)				(-3.19)		
(5)	1.206					-0.409	
	(5.44)					(-3.15)	
(6)	1.412		-0.024	-0.263			
	(6.44)		(-0.51)	(-5.36)			
(7)	1.337		0.035	-0.133	-0.201		
	(6.41)		(0.73)	(-2.98)	(-2.16)		
(8)	1.37		-0.017	-0.227		-0.22	
	(5.91)		(-0.38)	(-4.32)		(-1.76)	
(9)	1.412		-0.024	-0.263			-0.201
	(6.44)		(-0.51)	(-5.36)			(-2.16)
(10)	1.279		0.056	-0.063	-0.233	-0.346	
	(5.86)		(1.16)	(-1.43)	(-2.54)	(-3.17)	

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IR}_{t-5,t}$
(1)	1.341	0.196					
	(5.98)	(4.25)					
(2)	1.217		0.025				
	(5.47)		(0.71)				
(3)	1.328			-0.114			
	(6.03)			(-3)			
(4)	1.265				-0.256		
	(5.9)				(-3.19)		
(5)	1.206					-0.409	
	(5.44)					(-3.15)	
(6)	1.287		-0.018	-0.127			
	(5.85)		(-0.51)	(-3.22)			
(7)	1.281		0.063	0.024	-0.29		
	(5.98)		(1.74)	(0.72)	(-3.17)		
(8)	1.248		-0.021	-0.095		-0.367	
	(5.44)		(-0.62)	(-2.49)		(-3.09)	
(9)	1.287		-0.018	-0.127			-0.29
	(5.85)		(-0.51)	(-3.22)			(-3.17)
(10)	1.234		0.062	0.065	-0.302	-0.439	
	(5.58)		(1.76)	(2.04)	(-3.38)	(-4.12)	

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IC}_{t-5,t}$
(1)	1.275	0.054					
	(5.46)	(1.75)					
(2)	1.239		0.017				
	(5.46)		(0.66)				
(3)	1.38			-0.207			
	(6.39)			(-5.85)			
(4)	1.265				-0.256		
	(5.9)				(-3.19)		
(5)	1.206					-0.409	
	(5.44)					(-3.15)	
(6)	1.349		-0.024	-0.212			
	(6)		(-0.96)	(-6.17)			
(7)	1.34		-0.004	-0.149	-0.195		
	(6.07)		(-0.15)	(-4.77)	(-2.34)		
(8)	1.327		-0.013	-0.182		-0.182	
	(5.76)		(-0.52)	(-5.58)		(-1.43)	
(9)	1.349		-0.024	-0.212			-0.195
	(6)		(-0.96)	(-6.17)			(-2.34)
(10)	1.317		0.012	-0.103	-0.21	-0.284	
	(5.87)		(0.5)	(-3.73)	(-2.56)	(-2.55)	

Panel D: Combination information

Reg \Var	Constant	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$
(1)	1.306	0.062	0.145				
	(6.09)	(0.84)	(3.67)				
(2)	1.134	0.283		-0.074			
	(4.96)	(4.01)		(-2.83)			
(3)	1.354		0.182	0.022			
	(5.76)		(4.42)	(0.83)			
(4)	1.281	0.099	0.125	-0.015			
	(5.53)	(1.05)	(3.02)	(-0.56)			
(5)	1.197				0.048	-0.008	
	(5.5)				(0.89)	(-0.23)	
(6)	1.212				0.036		0.007
	(5.4)				(0.59)		(0.25)
(7)	1.234					0.017	0.018
	(5.4)					(0.52)	(0.79)
(8)	1.199				0.048	-0.008	0.003
	(5.17)				(0.69)	(-0.26)	(0.1)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past annual variables without indicator. Regressions only covers firms with positive retained earnings and contributed capital from July 1969 to December 2020. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$, past five-year earning returns, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$, five years stock returns $r^{i}_{-5,t}$ and intangible returns, $r^{i,IB}_{t-5,t}$, $r^{i,IR}_{t-5,t}$ and $r^{i,IC}_{t-5,t}$. Panel A, B and C only include the variables under one measurement, respectively. Panel D combines the current and past earnings ratios together to compare the performance with the indicators. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

Table 2.13: FM regression of monthly return on decomposition of price-scaled variables without share prices under 5 dollars

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.158	0.172					
	(5.39)	(2.81)					
(2)	1.138		0.048				
	(5.3)		(0.95)				
(3)	1.292			-0.16			
	(6.2)			(-3.86)			
(4)	1.186				-0.163		
	(5.76)				(-2.42)		
(5)	1.143					-0.481	
	(5.3)					(-3.83)	
(6)	1.255		0.002	-0.15			
	(6.02)		(0.05)	(-3.9)			
(7)	1.219		0.044	-0.072	-0.132		
	(6.04)		(0.86)	(-1.89)	(-1.67)		
(8)	1.207		0.011	-0.097		-0.393	
	(5.56)		(0.25)	(-2.68)		(-3.41)	
(9)	1.255		0.002	-0.15			-0.132
	(6.02)		(0.05)	(-3.9)			(-1.67)
(10)	1.166		0.066	0.008	-0.172	-0.457	
	(5.59)		(1.28)	(0.2)	(-2.2)	(-4.14)	

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.279	0.18						-0.308
	(6.04)	(3.87)						(-2.55)
(2)	1.178		0.034					-0.206
	(5.52)		(0.94)					(-1.74)
(3)	1.248			-0.074				-0.277
	(6.21)			(-1.91)				(-2.11)
(4)	1.214				-0.166			-0.238
	(6.01)				(-2.5)			(-2.27)
(5)	1.15					-0.457		-0.076
	(5.42)					(-3.95)		(-0.83)
(6)	1.226		0.008	-0.074				-0.254
	(5.99)		(0.22)	(-2.01)				(-1.92)
(7)	1.228		0.062	0.025	-0.197			-0.255
	(6.15)		(1.7)	(0.81)	(-2.72)			(-2.02)
(8)	1.17		0.007	-0.035		-0.421		-0.1
	(5.51)		(0.21)	(-1.04)		(-4.18)		(-0.92)
(9)	1.226		0.008	-0.074			-0.197	-0.254
	(5.99)		(0.22)	(-2.01)			(-2.72)	(-1.92)
(10)	1.166		0.062	0.072	-0.206	-0.457		-0.093
	(5.66)		(1.76)	(2.54)	(-2.92)	(-4.9)		(-0.87)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.14	0.008						0.097
	(5.07)	(0.29)						(1.61)
(2)	1.133		-0.01					0.089
	(5.07)		(-0.37)					(1.6)
(3)	1.283			-0.179				-0.009
	(6.08)			(-5.11)				(-0.22)
(4)	1.181				-0.163			0.069
	(5.69)				(-2.44)			(1.55)
(5)	1.145					-0.483		-0.028
	(5.29)					(-3.86)		(-0.75)
(6)	1.24		-0.05	-0.193				0.028
	(5.65)		(-1.97)	(-5.62)				(0.53)
(7)	1.225		-0.038	-0.157	-0.097			0.03
	(5.77)		(-1.55)	(-4.99)	(-1.41)			(0.59)
(8)	1.234		-0.021	-0.144		-0.321		-0.039
	(5.54)		(-0.86)	(-4.75)		(-2.73)		(-0.77)
(9)	1.24		-0.05	-0.193			-0.097	0.028
	(5.65)		(-1.97)	(-5.62)			(-1.41)	(0.53)
(10)	1.225		-0.005	-0.095	-0.118	-0.376		-0.035
	(5.73)		(-0.2)	(-3.43)	(-1.75)	(-3.39)		(-0.71)

Panel D: Combination information

Reg \Var	Constant	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	ind^{i}_{RE}	ind^{i}_{CC}
(1)	1.256	0.039	0.148					-0.304	
	(5.96)	(0.63)	(4.46)					(-2.9)	
(2)	1.039	0.274		-0.106					0.174
	(4.6)	(3.7)		(-3.47)					(1.91)
(3)	1.259		0.176	-0.007				-0.305	0.046
	(5.71)		(4.2)	(-0.3)				(-2.49)	(0.88)
(4)	1.167	0.123	0.108	-0.054				-0.174	0.096
	(5.17)	(1.46)	(2.65)	(-1.85)				(-2.14)	(1.08)
(5)	1.165				0.023	0.015		-0.199	
	(5.51)				(0.44)	(0.49)		(-1.81)	
(6)	1.085				0.09		-0.05		0.122
	(4.85)				(1.42)		(-1.55)		(1.51)
(7)	1.172					0.029	0.003	-0.201	0.056
	(5.32)					(0.87)	(0.13)	(-1.65)	(1.2)
(8)	1.142				0.039	0.007	-0.013	-0.178	0.067
	(5.14)				(0.62)	(0.24)	(-0.51)	(-1.93)	(1.03)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past annual variables with earning indicator. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$, past five-year earning returns, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$, five years stock returns $r^{i}_{t-5,t}$ and intangible returns, $r^{i,IB}_{t-5,t}$, $r^{i,IR}_{t-5,t}$ and $r^{i,IC}_{t-5,t}$. Panel A, B and C only include the variables under one measurement, respectively. Panel D combines the current and past earnings ratios together to compare the performance with the indicators. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment. Regressions only covers the firms with share prices greater than five dollars from July 1969 to December 2020.

but $bm_{i,t-5}$ is still insignificant on forecasting the monthly stock returns. Similar to the full sample cases, the result of multivariate regressions (6) to (10) indicate that only the prior five-year issuance effect may consistently contribute to the robust prediction of expected stock returns. The performance of $r^{i,BE}_{t-5,t}$ are completely absorbed by $r^{i}_{t-5,t}$ and $iss^{i}_{t-5,t}$.

Similar to BE-based regressions, Table 2.13 Panel B depicts the change in RE-related variables. The performance of $rm_{i,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i}_{t-5,t}$ is flattened during the past 17 years (from 2004 to 2020), but the issuance effect remains significant (t-value, -4.04 to -3.95). The insignificance of past RE variables, $rm_{i,t-5}$ and $r^{i,RE}_{t-5,t}$ in univariate regressions is an intriguing trend. Nevertheless, $r^{i,RE}_{t-5,t}$ is statistically significant for measuring monthly returns with all past decomposed components included the multivariate regression (10).

Table 2.13 Panel C indicates that $cm_{i,t}$ only has t-statistic of 0.29 under the data sample for stock prices greater than 5 dollars, compared with 2.38 for full sample estimation. $cm_{i,t-5}$ does not significantly forecast the future stock returns. Like BE and RE returns, CC returns can have a statistically significant influence for monthly stock returns. Intuitively, $r^{i,CC}_{t-5,t}$ is more significant than BE and RE returns, and its t-value is more robust than issuance effect ($|-5.11| > |-3.86|$) even including $r^{i}_{t-5,t}$ in regression (10). Still, ind^{i}_{CC} is unable to distinguish the negative and positive values. Remarkably, $cm_{i,t-5}$ and $r^{i,CC}_{t-5,t}$ can explain the future stock returns at the 5% significant level with an insignificant intangible return ($t = -1.41$) in regression (9), which is far better than BE-based and RE-based regressions.

Finally, Table 2.13 Panel D replicates the joint test to examine the relationship between three price-scaled ratios. The results conclude that: First, regressions (1) to (4) shows that $rm_{i,t}$ dominates other two current value ratios. Second, $bm_{i,t}$ cannot absorb the significance of $cm_{i,t}$ in regression (2). Also, the performance of $cm_{i,t}$ surpasses that of $bm_{i,t}$ for the first time in regression (4) (t-value, $|1.46|$ vs $|-1.85|$). It indicates that $cm_{i,t}$ has marginal trend towards significance in forecasting monthly stock returns, even when the regression contains BE and RE related variables. Third, the performance of lagged ratios is similar to the full sample version in that all variables have negligible impact on predicting monthly stock returns.

2.4.3.3 All-but-micro data sample

Table 2.14 exhibits the results of FM regressions based on All-but-microcaps firms that is trimmed at the 20th percentile of NYSE market capitalisation.

The removal of the small firms further reduces the explanatory power of all right-hand side components in the univariate regressions, regardless of the underlying value variable. In Table 2.14 Panel A and Table 2.14 Panel C, value returns ($r^{i,BE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$), as well as $iss^{i}_{t-5,t}$, have the ability to predict the stock returns, whereas only $rm_{i,t}$ and $iss^{i}_{t-5,t}$ are significant in RE-based regression in Table 2.14 Panel B. In multivariate regressions, Table 2.14 Panel A shows that $r^{i,BE}_{t-5,t}$ has a significant and stable predictive ability for monthly stock returns. It is more robust than $r^{i}_{t-5,t}$ until the addition of the issuance activities $iss^{i}_{t-5,t}$. Likewise, $r^{i,CC}_{t-5,t}$ has a better prediction than $r^{i,BE}_{t-5,t}$. $r^{i,CC}_{t-5,t}$ can exceed the five-year lagged stock returns when issuance effects are included in regression (10) of Table 2.14 Panel C. Nevertheless, RE-related variables generate distinct results from the previous two returns, such that $r^{i,RE}_{t-5,t}$ loses the predictive power in Table 2.14 Panel B. Last, Table 2.13 Panel D shows the same results as the case with only firms' stock prices greater than 5 dollars that $rm_{i,t}$ is better than the other two and all three lagged ratios are inadequate for pricing the monthly stock returns.

According to the analysis presented previously, the reduction of data sam-

Table 2.14: FM regression of monthly return on decomposition of price-scaled variables with all-but-micro database

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.138	0.132					
	(5.34)	(1.93)					
(2)	1.106		0.009				
	(5.25)		(0.17)				
(3)	1.294			-0.159			
	(6.57)			(-3.22)			
(4)	1.171				-0.147		
	(5.82)				(-1.79)		
(5)	1.12					-0.48	
	(5.35)					(-3.59)	
(6)	1.25		-0.045	-0.163			
	(6.25)		(-0.88)	(-3.49)			
(7)	1.204		-0.011	-0.107	-0.082		
	(6.09)		(-0.2)	(-2.39)	(-0.89)		
(8)	1.184		-0.028	-0.099		-0.369	
	(5.69)		(-0.56)	(-2.37)		(-3.25)	
(9)	1.25		-0.045	-0.163			-0.082
	(6.25)		(-0.88)	(-3.49)			(-0.89)
(10)	1.134		0.017	-0.016	-0.126	-0.434	
	(5.59)		(0.31)	(-0.37)	(-1.35)	(-3.97)	

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.258	0.172						-0.379
	(5.99)	(3.45)						(-2.69)
(2)	1.133		0.021					-0.257
	(5.38)		(0.57)					(-1.91)
(3)	1.231			-0.079				-0.36
	(6.47)			(-1.77)				(-2.34)
(4)	1.184				-0.148			-0.28
	(5.96)				(-1.81)			(-2.39)
(5)	1.116					-0.466		-0.096
	(5.4)					(-3.65)		(-0.95)
(6)	1.201		-0.012	-0.086				-0.329
	(6.12)		(-0.32)	(-1.93)				(-2.13)
(7)	1.198		0.04	-0.001	-0.161			-0.321
	(6.11)		(1.03)	(-0.02)	(-1.79)			(-2.19)
(8)	1.131		-0.011	-0.037		-0.422		-0.117
	(5.53)		(-0.3)	(-0.95)		(-3.95)		(-0.92)
(9)	1.201		-0.012	-0.086			-0.161	-0.329
	(6.12)		(-0.32)	(-1.93)			(-1.79)	(-2.13)
(10)	1.122		0.042	0.057	-0.174	-0.465		-0.109
	(5.56)		(1.12)	(1.56)	(-1.97)	(-4.76)		(-0.87)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i}_{t-5,t}$	$iss^{i}_{t-5,t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.106	0.002						0.088
	(5.03)	(0.06)						(1.33)
(2)	1.106		-0.011					0.067
	(5.06)		(-0.41)					(1.09)
(3)	1.284			-0.172				-0.053
	(6.39)			(-4.57)				(-1.35)
(4)	1.167				-0.148			0.05
	(5.74)				(-1.8)			(1.17)
(5)	1.124					-0.484		-0.03
	(5.33)					(-3.63)		(-0.81)
(6)	1.233		-0.045	-0.18				-0.019
	(5.84)		(-1.62)	(-4.76)				(-0.35)
(7)	1.211		-0.037	-0.151	-0.074			-0.023
	(5.87)		(-1.37)	(-4.48)	(-0.9)			(-0.43)
(8)	1.223		-0.017	-0.128		-0.311		-0.07
	(5.72)		(-0.62)	(-4.09)		(-2.65)		(-1.25)
(9)	1.233		-0.045	-0.18			-0.074	-0.019
	(5.84)		(-1.62)	(-4.76)			(-0.9)	(-0.35)
(10)	1.202		-0.005	-0.087	-0.094	-0.37		-0.073
	(5.8)		(-0.2)	(-2.94)	(-1.15)	(-3.37)		(-1.36)

Panel D: Combination information

Reg \Var	Constant	$bm_{i,t}$	$rm_{i,t}$	$cm_{i,t}$	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	ind^{i}_{RE}	ind^{i}_{CC}
(1)	1.243	-0.023	0.18					-0.401	
	(5.89)	(-0.3)	(4.25)					(-3.15)	
(2)	1.025	0.23		-0.093					0.147
	(4.63)	(2.89)		(-3.22)					(1.66)
(3)	1.213		0.169	-0.018				-0.368	0.063
	(5.41)		(3.67)	(-0.68)				(-2.52)	(0.97)
(4)	1.184	0.046	0.145	-0.036				-0.312	0.027
	(5.22)	(0.43)	(2.68)	(-1.17)				(-2.88)	(0.33)
(5)	1.128				-0.029	0.033		-0.263	
	(5.38)				(-0.47)	(0.9)		(-2.09)	
(6)	1.066				0.039		-0.03		0.088
	(4.92)				(0.6)		(-0.98)		(1.09)
(7)	1.109					0.017	-0.008	-0.234	0.063
	(5.03)					(0.49)	(-0.31)	(-1.69)	(1.11)
(8)	1.139				-0.045	0.042	0.007	-0.299	0.033
	(5.1)				(-0.56)	(0.98)	(0.24)	(-2.42)	(0.44)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past annual variables with earning indicator. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$, past five-year earning returns, $r^{i,BE}_{t-5,t}$, $r^{i,RE}_{t-5,t}$ and $r^{i,CC}_{t-5,t}$, five years stock returns $r^{i}_{t-5,t}$ and intangible returns, $r^{i,IB}_{t-5,t}$, $r^{i,IR}_{t-5,t}$ and $r^{i,IC}_{t-5,t}$. Panel A, B and C only include the variables under one measurement, respectively. Panel D combines the current and past earnings ratios together to compare the performance with the indicators. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020.

ples will substantially affect the performance of past variables. Following are a few conclusions:

First, when the data shrinkage level grows, the performance of $bm_{i,t}$ rapidly drops from significant to insignificant, whereas $rm_{i,t}$ is unaltered. Meanwhile, it should not surprise that $cm_{i,t}$ is ineffective in predicting future stock returns.

Second, the performance of the lagged ratios contradicts the findings of DT. According to their assumption, lagged ratios may carry some permanent information on characteristics that can influence stock returns in the future. However, I reach the opposite conclusion about the five-year lagged information. Regardless of data sample restrictions, all three five-year lagged value ratios are insignificant. They can only have substantial explanatory power in univariate regressions with a monthly updated full sample or a sample with share prices greater than 5 dollars.[11]

Third, variables deflated by share prices or market equities, such as $bm_{i,t}$, $bm_{i,t-5}$ and $r^i_{t-5,t}$, can benefit from spanning the data sample at the end of each month. Fewer lags between the estimation and collection periods will increase the significance of these variables, specifically the lagged five-year value ratios, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cc_{i,t-5}$, with t-values greater than 2.

Fourth, the data reducing process improves the predictive power of the five-year lagged CC return, $r^{i,CC}_{t-5,t}$, which outperforms the extraordinary explanatory power of issuance effects in the All-but-micro data sample. In contrast, excluding the small firms can weaken the performance of $r^{i,BE}_{t-5,t}$ and $r^{i,RE}_{t-5,t}$, as well as $r^i_{t-5,t}$. In addition, the insignificance of prior five-year stock returns suggests that, with a few exceptions, the decomposition of value ratios can explain most monthly stock returns.

Finally, the issuance effect, $iss^i_{t-5,t}$, has a higher correlation with expected

[11]This chapter does not include the results of FM regression with a monthly updated data sample comprising share values greater than 5 dollars.

stock returns in large firms than in small firms, as the t-statistics increase as the number of small firms decreases.

2.5 Decomposition of past information into short and long-term components

According to the above subsection, decomposition by DT does not provide consistent and substantial long-term variables for forecasting the monthly stock returns. Prior five-year value ratios are meaningless, and five-year BE and RE returns cannot beat the long-term stock returns. Then, I divide the long period into several intervals to examine if variables from distinct periods can have the same explanatory power on future stock returns. I select the prior year as the time node following Asness (1995). The prior five-year data will be decomposed into two periods, year $t-1$ to year $t-5$ as long-term and medium-term from year $t-1$ to year t. Similarly, the medium-term variables will be segmented into one-year momentum, computed from year $t-1$ to year t skip the last month, and a short-term variable only including the most recent month's information.

2.5.1 Summary statistics

Table 2.15 Panel A and Table 2.15 Panel B presents the details of the decomposition of the past returns based on annual and monthly updated data, respectively.

The difference between annual data and monthly updated data in three five-year lagged price-scaled ratios, $btm_{i,t-5}$, $rtm_{i,t-5}$ and $ctm_{i,t-5}$ are small, and the difference between two data frequency in past return computed is more negligible. The average stock return will increase as the number of cumulative

Table 2.15: Summary statistics of further decomposed data under full sample

Panel A: Annual basis

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
$bm_{i,t-5}$	-0.43	0.96	0.81	11.51	-6.51	-2.76	-1.87	-0.92	-0.40	0.03	0.86	2.68	6.59	2749.51	1964.01
$rm_{i,t-5}$	-0.74	1.06	-0.29	9.16	-7.72	-3.79	-2.53	-1.30	-0.55	-0.09	0.27	2.04	6.01	2749.51	1696.72
$cm_{i,t-5}$	-1.00	1.26	-0.19	6.70	-8.44	-4.45	-3.08	-1.70	-0.91	-0.23	0.81	2.28	5.78	2749.51	2077.21
$iss^i_{t-5,t}$	0.06	0.47	1.74	29.17	-3.68	-0.84	-0.40	-0.17	-0.02	0.19	0.82	1.75	4.93	2749.51	1400.48
$r^{i,BE}_{t-5,t}$	0.73	1.02	-0.12	10.94	-6.09	-2.26	-0.79	0.24	0.67	1.24	2.33	3.42	7.11	2749.51	484.21
$r^{i,RE}_{t-5,t}$	0.73	0.94	0.99	10.96	-4.99	-1.37	-0.25	0.16	0.52	1.19	2.41	3.69	8.10	2749.51	215.88
$r^{i,CC}_{t-5,t}$	0.65	1.02	0.39	12.30	-6.80	-2.01	-0.62	0.09	0.49	1.11	2.38	3.69	8.22	2749.51	387.38
ind^i_{RE}	0.25	0.40	1.57	5.15	0.00	0.00	0.00	0.00	0.00	0.60	0.96	1.00	1.00	2749.51	0.00
ind^i_{CC}	0.07	0.21	4.15	72.86	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.73	0.75	2749.51	0.00
$r^i_{t-5,t}$	0.28	0.91	-0.68	6.59	-5.09	-2.36	-1.31	-0.18	0.37	0.83	1.58	2.26	3.68	2749.51	885.19
$r^i_{t-5,t-1}$	0.24	0.81	-0.56	6.11	-4.31	-2.11	-1.18	-0.19	0.31	0.72	1.43	2.08	3.39	2749.51	916.10
$r^i_{t-1,t}$	0.04	0.41	-0.49	9.73	-2.50	-1.13	-0.64	-0.16	0.06	0.27	0.66	1.07	2.14	2749.51	1110.10
$r^i_{t-1,t-1/12}$	0.04	0.39	-0.45	9.64	-2.37	-1.06	-0.61	-0.16	0.05	0.25	0.61	1.02	2.03	2749.51	1134.74
$r^i_{t-1/12,t}$	0.01	0.12	0.06	13.92	-0.89	-0.31	-0.17	-0.05	0.01	0.06	0.18	0.32	0.93	2749.51	1211.30

Panel B: Monthly basis

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
$bm_{i,t-5}$	-0.52	1.01	0.58	9.83	-6.63	-3.05	-2.10	-1.04	-0.48	-0.02	0.88	2.63	6.34	2749.51	2028.67
$rm_{i,t-5}$	-0.81	1.10	-0.40	8.32	-7.82	-4.05	-2.70	-1.40	-0.62	-0.09	0.25	1.95	5.77	2749.51	1726.90
$cm_{i,t-5}$	-1.09	1.30	-0.22	6.22	-8.56	-4.58	-3.23	-1.82	-1.00	-0.28	0.78	2.20	5.57	2749.51	2104.22
$iss^i_{t-5,t}$	0.05	0.48	1.89	33.47	-3.61	-0.87	-0.41	-0.17	-0.02	0.18	0.81	1.78	5.22	2749.51	1416.94
$r^{i,BE}_{t-5,t}$	0.73	1.02	-0.12	10.94	-6.09	-2.26	-0.79	0.24	0.67	1.24	2.33	3.42	7.11	2749.51	484.21
$r^{i,RE}_{t-5,t}$	0.73	0.94	0.99	10.96	-4.99	-1.37	-0.25	0.16	0.52	1.19	2.41	3.69	8.10	2749.51	215.88
$r^{i,CC}_{t-5,t}$	0.65	1.02	0.39	12.30	-6.80	-2.01	-0.62	0.09	0.49	1.11	2.38	3.69	8.22	2749.51	387.38
ind^i_{RE}	0.25	0.40	1.57	5.15	0.00	0.00	0.00	0.00	0.00	0.60	0.96	1.00	1.00	2749.51	0.00
ind^i_{CC}	0.07	0.21	4.15	72.86	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.73	0.75	2749.51	0.00
$r^i_{t-5,t}$	0.25	0.90	-0.85	7.65	-5.60	-2.48	-1.35	-0.19	0.35	0.79	1.51	2.18	3.58	2749.51	896.64
$r^i_{t-5,t-1}$	0.22	0.79	-0.68	7.13	-4.60	-2.09	-1.15	-0.18	0.29	0.68	1.36	1.99	3.32	2749.51	919.32
$r^i_{t-1,t}$	0.03	0.41	-0.75	10.48	-2.86	-1.22	-0.66	-0.16	0.06	0.25	0.63	1.02	2.05	2749.51	1133.55
$r^i_{t-1,t-1/12}$	0.03	0.39	-0.70	10.28	-2.66	-1.15	-0.63	-0.16	0.05	0.24	0.60	0.97	1.98	2749.51	1143.35
$r^i_{t-1/12,t}$	0.00	0.12	-0.10	15.76	-1.04	-0.33	-0.18	-0.05	0.00	0.06	0.18	0.34	0.96	2749.51	1265.99

Note: This table reports the summary statistics of all long, medium and short term past variables with full samples from July 1969 to December 2020. Panel A shows the results under the annual basis construction, where the accounting information is collected at the end of fiscal year and stock trading information is collected at the end of December. Variables in Panel B span the stock information at the end of each month. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the 60 months lagged earning ratios. $r^{i,BE}_{t-5,t}$, $r^{i,BE}_{t-5,t}$ and $r^{i,BE}_{t-5,t}$ are the past five-year book equity return, retained earnings return and contributed capital return, respectively. $r^i_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^i_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^i_{t-1,t}$ is the prior one-year stock return from the current period. $r^i_{t-1,t-1/12}$, momentum is the prior one-year stock return skip the last one month. $r^i_{t-1/12,t}$ is the prior one-month stock return.

periods increase regardless of the data frequency.

Next, I employ the same data constraints as in the previous section. Table 2.16 Panel A displays the results for firms with only share prices greater than 5 dollars, whereas Table 2.16 Panel B excludes the micro firms whose ME is less than the 20th percentile of NYSE market cap.

The sample distribution has no effect on the average of short-term returns, $r^i_{t-1_{/12},t}$, while the data shrinkage process enhances the medium-term returns, momentum and prior one-year log returns from 0.04 to around 0.10. The mean values of $r^i_{t-5,t}$ increase from 0.28 (full sample) to 0.48 (only prices greater 5 dollars) and 0.56 (All-but-microcaps). Likewise, the data reduction benefits both $r^i_{t-5,t}$, increasing from 0.28 (full sample) to 0.48 (only prices greater 5 dollars) and 0.56 (All-but-microcaps), and $r^i_{t-5,t-1}$, climbing from 0.24 to 0.48 and 0.56. Higher market capitalisation can result in higher average long-term stock returns. The finding is contradicted to Banz (1981) who find that small firms have larger stock returns than large firms over the short term.

Table Table 2.17 describes the correlation between several returns in varying intervals and fundamental earning ratios. The difference between the annual and monthly correlation coefficients is not excessive. Also, Pearson's correlation is analogous to the Spearman ranking correlation. Some findings are shown as follows: The correlation coefficient between five-year past stock return ($r^i_{t-5,t}$) and four-year stock return ($r^i_{t-5,t-1}$) is about 89%. Second, the correlation coefficient between $r^i_{t-1,t}$ and $r^i_{t-5,t-1}$ is only 4.6%. Third, the correlation between prior value returns and long-term stock returns is notably negative, especially if last year's return is excluded from the prior five-year stock return. Fourth, the prior one-month return hardly significantly impacts the past long-term return and has a considerable correlation with prior one-year stock returns (28%). Finally, lagged value ratios ($bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$) have a detectable

Table 2.16: Summary statistics of further decomposed data with data filtering restriction under annual basis

Panel A: share price greater than 5 dollars

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
$bm_{i,t-5}$	-0.41	0.94	0.95	11.71	-6.27	-2.58	-1.78	-0.89	-0.40	0.01	0.84	2.78	6.49	2252.28	1613.62
$rm_{i,t-5}$	-0.81	1.07	-0.14	8.98	-7.64	-3.82	-2.56	-1.37	-0.71	-0.11	0.31	2.21	5.95	2252.28	1556.77
$cm_{i,t-5}$	-1.10	1.26	-0.28	6.49	-8.41	-4.57	-3.18	-1.79	-1.01	-0.31	0.59	2.20	5.51	2252.28	1755.51
$iss^i_{t-5,t}$	0.01	0.41	1.16	28.13	-3.49	-0.85	-0.41	-0.19	-0.04	0.14	0.67	1.44	4.02	2252.28	1263.21
$r^{i,BE}_{t-5,t}$	0.88	0.90	0.48	12.74	-4.70	-1.41	-0.29	0.37	0.76	1.32	2.41	3.48	7.00	2252.28	251.61
$r^{i,RE}_{t-5,t}$	0.85	0.94	1.14	11.12	-4.20	-1.08	-0.18	0.23	0.69	1.31	2.51	3.82	8.10	2252.28	156.22
$r^{i,CC}_{t-5,t}$	0.72	1.02	0.50	12.31	-6.47	-1.82	-0.45	0.11	0.55	1.19	2.45	3.79	8.08	2252.28	280.30
ind^i_{RE}	0.17	0.35	2.39	9.66	0.00	0.00	0.00	0.00	0.00	0.38	0.81	1.00	1.00	2252.28	0.00
ind^i_{CC}	0.08	0.22	3.75	62.96	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.73	0.75	2252.28	0.00
$r^i_{t-5,t}$	0.48	0.73	-0.17	6.33	-3.40	-1.37	-0.69	0.06	0.49	0.90	1.65	2.32	3.66	2252.28	522.13
$r^i_{t-5,t-1}$	0.38	0.68	-0.17	5.45	-3.28	-1.40	-0.72	0.00	0.40	0.77	1.47	2.12	3.35	2252.28	582.81
$r^i_{t-1,t}$	0.10	0.35	0.02	8.80	-1.76	-0.79	-0.45	-0.09	0.09	0.28	0.65	1.05	2.01	2252.28	789.87
$r^i_{t-1,t-1/12}$	0.08	0.33	0.01	8.75	-1.69	-0.77	-0.44	-0.10	0.08	0.26	0.61	0.99	1.92	2252.28	818.55
$r^i_{t-1/12,t}$	0.02	0.10	0.41	13.11	-0.59	-0.23	-0.13	-0.04	0.01	0.06	0.17	0.29	0.75	2252.28	933.04

Panel B: All-but-micro

Variables	Mean	std	skew	kurt	Min	P1	P5	P25	Median	P75	P95	P99	Max	#	Less0
$bm_{i,t-5}$	-0.54	0.86	0.75	10.73	-4.98	-2.60	-1.85	-0.99	-0.51	-0.13	0.61	2.21	5.19	1421.11	1114.06
$rm_{i,t-5}$	-0.93	1.02	-0.28	8.29	-7.09	-3.80	-2.61	-1.47	-0.86	-0.19	0.15	1.71	4.79	1421.11	1067.57
$cm_{i,t-5}$	-1.26	1.23	-0.58	6.09	-8.01	-4.84	-3.37	-1.96	-1.14	-0.38	0.26	1.55	4.24	1421.11	1137.27
$iss^i_{t-5,t}$	0.02	0.39	2.03	23.76	-2.32	-0.72	-0.38	-0.18	-0.05	0.15	0.68	1.42	3.75	1421.11	809.78
$r^{i,BE}_{t-5,t}$	1.02	0.89	1.00	12.71	-3.36	-0.91	-0.09	0.47	0.89	1.46	2.59	3.74	6.94	1421.11	116.04
$r^{i,RE}_{t-5,t}$	0.99	0.96	1.27	12.30	-3.77	-0.85	-0.05	0.30	0.83	1.45	2.69	4.04	7.94	1421.11	76.50
$r^{i,CC}_{t-5,t}$	0.84	1.05	0.75	12.20	-5.41	-1.57	-0.30	0.18	0.68	1.33	2.65	4.07	8.05	1421.11	137.97
ind^i_{RE}	0.14	0.31	3.40	20.59	0.00	0.00	0.00	0.00	0.00	0.28	0.65	0.97	1.00	1421.11	0.00
ind^i_{CC}	0.10	0.24	2.33	7.28	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.73	0.73	1421.11	0.00
$r^i_{t-5,t}$	0.56	0.69	0.00	5.33	-2.63	-1.22	-0.54	0.17	0.55	0.94	1.68	2.37	3.61	1421.11	270.74
$r^i_{t-5,t-1}$	0.45	0.64	-0.01	5.41	-2.58	-1.21	-0.56	0.10	0.44	0.80	1.49	2.14	3.27	1421.11	305.28
$r^i_{t-1,t}$	0.11	0.32	0.03	6.59	-1.42	-0.73	-0.40	-0.06	0.11	0.28	0.61	0.96	1.71	1421.11	464.90
$r^i_{t-1,t-1/12}$	0.09	0.30	0.01	6.78	-1.40	-0.71	-0.39	-0.07	0.09	0.25	0.57	0.90	1.62	1421.11	486.02
$r^i_{t-1/12,t}$	0.02	0.08	0.17	7.85	-0.45	-0.20	-0.11	-0.03	0.02	0.06	0.15	0.25	0.53	1421.11	564.80

Note: This table reports the summary statistics of all long, medium and short term past variables under annual basis with different data sample restriction from July 1969 to December 2020. Accounting information is collected at the end of fiscal year and stock trading information is collected at the end of December each year. Panel A only includes the firms with share prices greater than 5 dollars. Panel B removes the stocks' market capitalisation lower than 20th percentile of NYSE market cap at the end of June each month. $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$ are the 60 months lagged earning ratios. $r^{i,BE}_{t-5,t}$, $r^{i,BE}_{t-5,t}$ and $r^{i,BE}_{t-5,t}$ are the past five-year book equity return, retained earnings return and contributed capital return, respectively. $r^i_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^i_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^i_{t-1,t}$ is the prior one-year stock return from the current period. $r^i_{t-1,t-1/12}$, momentum is the prior one-year stock return skip the last one month. $r^i_{t-1/12,t}$ is the prior one-month stock return.

Table 2.17: Average correlation coefficients under full sample

Panel A: Annual basis

Variables	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	$iss^i_{t-5,t}$	$r^{i,BE}_{t-5,t}$	$r^{i,RE}_{t-5,t}$	$r^{i,CC}_{t-5,t}$	$r^i_{t-5,t}$	$r^i_{t-5,t-1}$	$r^i_{t-1,t}$	$r^i_{t-1,t-1/12}$	$r^i_{t-1,t-1/12}$
$bm_{i,t-5}$	100	54.1	56.8	-7.21	-22.9	-12.29	-14.31	18.73	17.88	6.51	5.98	5.98
$rm_{i,t-5}$	47.81	100	36.35	6.33	-23.09	-47.16	-11.28	-0.55	-0.32	-0.66	-0.6	-0.6
$cm_{i,t-5}$	51.9	36.66	100	15.59	-22.28	-18.9	-30.06	5.78	6.32	0.29	0.7	0.7
$iss^i_{t-5,t}$	-12.05	7.99	8.79	100	12.07	0	22.07	-14	-10.44	-11.25	-10.57	-10.57
$r^{i,BE}_{t-5,t}$	-25.17	-25.09	-22.69	20.78	100	62.68	70.93	55.36	57.24	8.75	7.92	7.92
$r^{i,RE}_{t-5,t}$	-12.23	-50.24	-23.37	-6.29	64.3	100	38.69	42.42	42.8	8.43	7.65	7.65
$r^{i,CC}_{t-5,t}$	-14.76	-11.31	-30.74	33.13	67.86	32	100	30.51	33.89	0.22	-0.18	-0.18
$r^i_{t-5,t}$	18.34	-1.02	6.77	-7.63	50.96	42.76	25.54	100	89.01	42.79	39.97	39.97
$r^i_{t-5,t-1}$	17.63	-0.52	7.16	-3.89	53.06	42.63	28.9	87.84	100	-0.99	-2.15	-2.15
$r^i_{t-1,t}$	5.65	-1.58	0.64	-9.41	7.07	9.36	-0.65	39.62	-0.85	100	95.83	95.83
$r^i_{t-1,t-1/12}$	5.25	-1.54	0.97	-8.96	6.38	8.56	-1.05	37.21	-1.73	95.27	100	100
$r^i_{t-1/12,t}$	1.86	-2.07	-2.1	-4.66	4.91	5.84	1.75	16.42	4	29.49	5.09	5.94

Panel B: Monthly basis

Variables	$bm_{i,t-5}$	$rm_{i,t-5}$	$cm_{i,t-5}$	$iss^i_{t-5,t}$	$r^{i,BE}_{t-5,t}$	$r^{i,RE}_{t-5,t}$	$r^{i,CC}_{t-5,t}$	$r^i_{t-5,t}$	$r^i_{t-5,t-1}$	$r^i_{t-1,t}$	$r^i_{t-1,t-1/12}$	$r^i_{t-1,t-1/12}$
$bm_{i,t-5}$	100	54.34	59.44	-8.06	-36.42	-21.21	-24.23	17.07	16.58	5.73	5.6	5.6
$rm_{i,t-5}$	49.07	100	39.08	5.9	-30.3	-55.19	-17.64	-1.12	-0.98	-1.03	-0.96	-0.96
$cm_{i,t-5}$	55.22	38.99	100	13.04	-32.71	-25.18	-37.73	4.75	5.57	-0.52	-0.43	-0.43
$iss^i_{t-5,t}$	-12.66	7.91	6.36	100	7.75	-1.69	19.86	-17.53	-13.6	-13.24	-12.63	-12.63
$r^{i,BE}_{t-5,t}$	-39.12	-32.23	-33.07	17.57	100	62.68	70.93	39.16	44.35	0.91	0.92	0.92
$r^{i,RE}_{t-5,t}$	-20.93	-57.35	-28.49	-7.76	64.3	100	38.69	32.45	35.04	3.49	3.41	3.41
$r^{i,CC}_{t-5,t}$	-24.58	-16.74	-38.3	31.46	67.86	32	100	17.58	22.32	-4.25	-4.1	-4.1
$r^i_{t-5,t}$	15.82	-2.22	5.59	-10.36	35.21	33.49	13.77	100	88.8	47.04	45.04	45.04
$r^i_{t-5,t-1}$	15.46	-1.85	6.2	-6.47	40.09	35.46	18.12	87.73	100	3.26	3.24	3.24
$r^i_{t-1,t}$	4.64	-1.93	-0.02	-10.43	0.04	4.85	-4.56	42.76	2.42	100	95.44	95.44
$r^i_{t-1,t-1/12}$	4.53	-1.9	0	-9.97	0.17	4.81	-4.36	41	2.52	94.7	100	100
$r^i_{t-1/12,t}$	0.86	-1.68	-0.99	-4.34	1.3	2.65	-0.49	14.13	2.11	28.01	2.25	2.55

Note: This table presents the Pearson average and Spearman rank correlation coefficients for accounting information and returns on stock or accounting measurement only containing share prices greater than five dollars. Panel A shows the information with annual basis. Panel B gives information for the sample constructed by monthly basis. $r^{i,BE}_{t-5,t}$, $r^{i,BE}_{t-5,t}$ and $r^{i,BE}_{t-5,t}$ are the past five-year book equity return, retained earnings return and contributed capital return, respectively. $r^i_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^i_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^i_{t-1,t}$ is the prior one-year stock return from the current period. $r^i_{t-1,t-1/12}$, momentum is the prior one-year stock return skip the last one month. $r^i_{t-1/12,t}$ is the prior one-month stock return. The sample coves the time range from July 1969 to December 2020.

negative correlation with past value returns, although the past value returns are strongly positively correlated with the past stock return. The result illustrates that the lower value ratios can provide higher expected profitability for the firm, producing higher stock returns.

2.5.2 Fama-MacBeth regression

This subsection runs FM regressions of monthly stock returns on past variables further decomposed for several periods. Following the above empirical analysis, I first evaluate annual or monthly updated data performance based on the full data sample. Then I extend the test with two data sample constraints that I filter small firms with share prices below 5 dollars or ME less than the 20th percentile of the NYSE. Prominently, the short-term and medium-term stock returns are calculated at the end of December in year $t-1$ like $r^i_{t-5,t}$ and available from July year t to June year $t+1$.[12] All tables in this subsection have the same structure. Panels A, B and C display the FM regression based on variables relevant to BTM, RTM and, respectively.

2.5.2.1 Full data sample

Table 2.18 contains all accessible firms without restriction. Table 2.18 Panel A displays the results of FM regression based on BE-related variables. Consistent with the finding of the previous section, only $bm_{i,t-5}$ cannot predict monthly returns at the 5% significant level. Both of the five-year skip last year BE return, $r_{i,BE}(t-5,t-1)$, and the medium-term BE return,$r_{i,BE}(t-1,t)$, can significantly estimate the future stock returns (t-value, -4.37 and -3.96). Besides, the decomposed medium-term and long-term stock returns always have

[12] Collection data at the end of December is consistent with DT. A Six-month latency is required between collection and construction date.

Table 2.18: FM regression of monthly returns on further decomposition with full data sample

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IB}_{t-5,t}$
(1)	1.279	0.25										
	(5.43)	(4.06)										
(2)	1.232		0.028									
	(5.32)		(0.53)									
(3)	1.435			-0.235								
	(5.86)			(-4.66)								
(4)	1.39				-0.211							
	(5.78)				(-4.37)							
(5)	1.317					-0.525						
	(5.49)					(-3.96)						
(6)	1.379		-0.035	-0.234								
	(5.76)		(-0.64)	(-4.42)								
(7)	1.293		0.061	-0.059			-0.276					
	(5.67)		(1.21)	(-1.64)			(-3.04)					
(8)	1.273		0.05		-0.049			-0.259				
	(5.62)		(1.01)		(-1.26)			(-2.99)				
(9)	1.241		0.047		-0.055			-0.269	-0.302			
	(5.66)		(0.96)		(-1.48)			(-3.24)	(-2.01)			
(10)	1.261		0.04		-0.065	-0.126		-0.243	-0.268			
	(5.73)		(0.82)		(-1.8)	(-1.65)		(-3)	(-1.8)			
(11)	1.269		0.043		-0.06	-0.123		-0.25		-0.279	-0.282	
	(5.84)		(0.9)		(-1.69)	(-1.63)		(-3.13)		(-1.95)	(-0.88)	
(12)	1.349		-0.051		-0.224	-0.29				-0.027	-0.027	-0.252
	(5.93)		(-0.99)		(-4.51)	(-3.18)				(-0.23)	(-0.09)	(-3.22)

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.346	0.204											-0.048
	(6.05)	(4.35)											(-0.31)
(2)	1.221		0.025										0.077
	(5.54)		(0.7)										(0.47)
(3)	1.332			-0.115									-0.034
	(6.09)			(-2.96)									(-0.22)
(4)	1.318				-0.12								-0.02
	(6.17)				(-3)								(-0.12)
(5)	1.248					-0.176							0.05
	(5.61)					(-1.77)							(0.34)
(6)	1.292		-0.019	-0.129									0.006
	(5.92)		(-0.52)	(-3.2)									(0.04)
(7)	1.285		0.086	0.058			-0.349						-0.13
	(6.01)		(2.4)	(1.68)			(-4.14)						(-0.89)
(8)	1.278		0.062		0.02			-0.297					-0.114
	(6.07)		(1.77)		(0.52)			(-3.73)					(-0.74)
(9)	1.258		0.062		0.018			-0.315	-0.3				-0.162
	(6.1)		(1.78)		(0.48)			(-4.08)	(-2.09)				(-1.12)
(10)	1.249		0.071		0.023	0.171		-0.34	-0.345				-0.153
	(6.04)		(2.07)		(0.64)	(2.73)		(-4.39)	(-2.38)				(-1.07)
(11)	1.261		0.073		0.026	0.172		-0.345		-0.359	-0.357		-0.156
	(6.19)		(2.16)		(0.74)	(2.76)		(-4.5)		(-2.56)	(-1.12)		(-1.09)
(12)	1.263		-0.032		-0.158	-0.014				-0.013	-0.002	-0.348	0
	(6.04)		(-0.94)		(-3.96)	(-0.2)				(-0.11)	(-0.01)	(-4.63)	(0)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.34	0.108											-0.031
	(5.15)	(2.38)											(-0.3)
(2)	1.283		0.022										-0.008
	(5.14)		(0.7)										(-0.09)
(3)	1.415			-0.221									-0.082
	(5.93)			(-5.64)									(-1.06)
(4)	1.403				-0.224								-0.078
	(5.93)				(-5.79)								(-1.04)
(5)	1.316					-0.499							-0.027
	(5.5)					(-6.03)							(-0.36)
(6)	1.393		-0.031	-0.231									-0.071
	(5.61)		(-1.05)	(-6.3)									(-0.79)
(7)	1.326		-0.011	-0.136			-0.251						0.008
	(5.54)		(-0.35)	(-3.78)			(-2.86)						(0.11)
(8)	1.322		-0.015		-0.16			-0.21					-0.019
	(5.5)		(-0.46)		(-4.33)			(-2.48)					(-0.27)
(9)	1.279		-0.025		-0.163			-0.225	-0.304				-0.003
	(5.52)		(-0.82)		(-4.81)			(-2.8)	(-2.05)				(-0.05)
(10)	1.283		-0.024		-0.144	-0.139		-0.217	-0.29				-0.011
	(5.54)		(-0.79)		(-4.1)	(-2.79)		(-2.72)	(-1.96)				(-0.16)
(11)	1.291		-0.022		-0.141	-0.137		-0.222		-0.301	-0.313		-0.012
	(5.65)		(-0.74)		(-4.07)	(-2.76)		(-2.81)		(-2.11)	(-0.98)		(-0.19)
(12)	1.36		-0.038		-0.222	-0.177				-0.071	-0.079	-0.233	-0.086
	(5.75)		(-1.3)		(-6.11)	(-3.42)				(-0.61)	(-0.25)	(-2.95)	(-1)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past decomposed short, medium and long-term variables under full data sample July 1969 to December 2020. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. Panel A, Panel B and Panel C run the FM regression with book-to-market, retained earning-to-marekt and contributed capital-to-market related variables, respectively. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$. $r^{i,*}_{t-5,t}$ represents past five-year earning returns computed by the underlying value. $r^{i,*}_{t-5,t-1}$ and $r^{i,*}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1_{/12}}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1_{/12},t}$ is the prior one-month stock return. ind^{i}_{RE} is the negative indicator to distinguish the negative retained earnings from the positive. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

a statistically significant explanatory power on expected stock returns, except the prior one-year stock return in regression (10). Notably, all short-term and medium-term stock returns lose the predictive power in regression (9) to (12), which contradicts the conventional finding.[13] Lastly, BE returns in all intervals are dominated by prior long-term stock returns, except when only one-year momentum and the prior one-month return are included.

Table 2.18 Panel B shows the results of FM regressions based on RE-related variables. Similarly, $rm_{i,t}$ and decomposed long-term RE returns, $r^{i,RE}_{t-5,t}$ and $r^{i,RE}_{t-5,t-1}$ have a statistical significant performance with t-value 4.35, -2.96 and -3, respectively, while the lagged ratio, $rm_{i,t-5}$, and medium-term RE returns has no explanatory power at the 5% significant level in univariate regressions (2) and (5). However, when decomposed long-term and medium-term RE returns and stock returns are included, the t-statistic of $r^{i,RE}_{t-5,t-1}$ increases to 2.7 in regression (10) and (11). The link between $r^{i,RE}_{t-5,t-1}$ and expected stock returns is changed from negative to positive. Moreover, $rm_{i,t-5}$ has a high explanatory power on monthly stock returns due to the highly negative correlation with the past RE returns. However, the robust performance is lost when the intangible return $r^{i,IR}_{t-5,t}$ is added in regression (12).

In contrast, Table 2.18 Panel C provides different results that CC returns in all intevals can have strong predictive power on the future stock returns, both in univariate and multivariate regressions. Like the earlier analysis, past CC returns outperforms than prior long-term stock returns. Besides, the performance of medium-term CC return, $r^{i,CC}_{t-1,t}$ is superior to that of the long-term CC return, $r^{i,CC}_{t-5,t-1}$ (t-value, -6.03 vs -5.79), but the long-term stock return can reduce the explanatory power of all period CC returns.

[13]Both Novy-Marx (2013)and Ball et al. (2015) have strong one-year stock momentum and prior one-month returns. It is because these variables are constructed at the end of each month with no time latency.

2.5.2.2 Adding data constraints

In this part, I add the data filtering criteria to the full data sample. Table 2.19 presents the results based on the data sample excluding share prices below 5 dollars.

Obviously, significant BE-based variables in full sample analysis are worse after adding the restriction in Table 2.19 Panel A. None of the past decomposed components had a significant capacity for explaining monthly stock returns in multivariate regressions (6) to (11). In regression (12), $r^{i,BE}_{t-5,t-1}$ and $r^{i,BE}_{t-1,t}$ have robust t-statistics, however all other variables are statistically insignificant, indicating that the result cannot be used to infer the performance of the decomposition of $bm_{i,t}$.

Table 2.19 Panel B shows that ind^{i}_{RE} plays a vital role in the regressions, resulting in a significant performance of past long-term stock returns compared with BE- and CC-based regressions. Meanwhile, the performance of other past variables is similarly insignificant to that of the BE-related variables.

Table 2.19 Panel C exhibits radically different results for CC-based FM regressions. First, ind^{i}_{CC} cannot differentiate between negative and positive values. Second, $cm_{i,t}$ and $cm_{i,t-5}$ are unimportant until the medium-term and short-term stock returns are included in regressions, which is the objective that BE- or RE-based regressions fail to achieve. Third, past long-term and medium-term CC returns benefit from the enhancement of $cm_{i,t-5}$, and their estimations are better than those in the full sample case. Fourth, past stock returns lose the potential to predict the monthly stock returns at the 5% significant level.

Table 2.20 displays the outcomes of FM regressions with additional data reduction to exclude micro firms. The data filtering process can diminish the significance of all past decomposition components. Second, $rm_{i,t}$ is the only price-scaled ratio having a statistically explanatory power in predicting the future

Table 2.19: FM regression of further decomposition for only 5 dollars

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IB}_{t-5,t}$
(1)	1.158	0.172										
	(5.39)	(2.81)										
(2)	1.138		0.048									
	(5.3)		(0.95)									
(3)	1.292			-0.16								
	(6.2)			(-3.86)								
(4)	1.249				-0.143							
	(6)				(-3.2)							
(5)	1.205					-0.338						
	(5.63)					(-3.55)						
(6)	1.255		0.002	-0.15								
	(6.02)		(0.05)	(-3.9)								
(7)	1.219		0.044	-0.072			-0.132					
	(6.04)		(0.86)	(-1.89)			(-1.67)					
(8)	1.175		0.042		-0.054			-0.137				
	(5.89)		(0.85)		(-1.38)			(-1.88)				
(9)	1.162		0.033		-0.056			-0.146	-0.078			
	(5.94)		(0.67)		(-1.54)			(-2.04)	(-0.56)			
(10)	1.178		0.026		-0.065	-0.134		-0.114	-0.042			
	(6.01)		(0.53)		(-1.8)	(-1.81)		(-1.58)	(-0.29)			
(11)	1.193		0.029		-0.06	-0.129		-0.124		-0.06	-0.001	
	(6.19)		(0.61)		(-1.7)	(-1.79)		(-1.72)		(-0.45)	(0)	
(12)	1.231		-0.013		-0.134	-0.203				0.066	0.125	-0.127
	(6.17)		(-0.29)		(-3.3)	(-2.77)				(0.54)	(0.36)	(-1.77)

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.279	0.18											-0.308
	(6.04)	(3.87)											(-2.55)
(2)	1.178		0.034										-0.206
	(5.52)		(0.94)										(-1.74)
(3)	1.248			-0.074									-0.277
	(6.21)			(-1.91)									(-2.11)
(4)	1.234				-0.075								-0.262
	(6.18)				(-1.84)								(-2.02)
(5)	1.191					-0.121							-0.22
	(5.69)					(-1.44)							(-1.96)
(6)	1.226		0.008	-0.074									-0.254
	(5.99)		(0.22)	(-2.01)									(-1.92)
(7)	1.228		0.062	0.025			-0.197						-0.255
	(6.15)		(1.7)	(0.81)			(-2.72)						(-2.02)
(8)	1.202		0.052		0.014			-0.182					-0.286
	(6.11)		(1.48)		(0.4)			(-2.69)					(-2.2)
(9)	1.184		0.05		0.014			-0.193	-0.061				-0.29
	(6.13)		(1.4)		(0.41)			(-2.89)	(-0.45)				(-2.32)
(10)	1.183		0.055		0.016	0.075		-0.206	-0.086				-0.281
	(6.12)		(1.57)		(0.49)	(1.26)		(-3.07)	(-0.62)				(-2.27)
(11)	1.201		0.057		0.02	0.077		-0.213		-0.106	-0.042		-0.283
	(6.33)		(1.65)		(0.62)	(1.33)		(-3.21)		(-0.81)	(-0.12)		(-2.3)
(12)	1.195		-0.001		-0.086	-0.029				0.109	0.177	-0.215	-0.261
	(6.1)		(-0.04)		(-2.22)	(-0.46)				(0.9)	(0.5)	(-3.25)	(-2.08)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.14	0.008											0.097
	(5.07)	(0.29)											(1.61)
(2)	1.133		-0.01										0.089
	(5.07)		(-0.37)										(1.6)
(3)	1.283			-0.179									-0.009
	(6.08)			(-5.11)									(-0.22)
(4)	1.273				-0.186								-0.006
	(6.06)				(-5.13)								(-0.15)
(5)	1.191					-0.367							0.038
	(5.55)					(-5.52)							(0.86)
(6)	1.24		-0.05	-0.193									0.028
	(5.65)		(-1.97)	(-5.62)									(0.53)
(7)	1.225		-0.038	-0.157			-0.097						0.03
	(5.77)		(-1.55)	(-4.99)			(-1.41)						(0.59)
(8)	1.19		-0.045		-0.179			-0.074					0.029
	(5.69)		(-1.83)		(-5.78)			(-1.16)					(0.6)
(9)	1.17		-0.05		-0.175			-0.088	-0.071				0.026
	(5.7)		(-2.04)		(-5.93)			(-1.39)	(-0.52)				(0.53)
(10)	1.171		-0.049		-0.156	-0.136		-0.077	-0.05				0.02
	(5.69)		(-2.02)		(-5.04)	(-2.83)		(-1.22)	(-0.36)				(0.4)
(11)	1.183		-0.048		-0.153	-0.13		-0.083		-0.067	-0.041		0.02
	(5.85)		(-1.97)		(-5.05)	(-2.76)		(-1.33)		(-0.51)	(-0.12)		(0.41)
(12)	1.206		-0.057		-0.179	-0.143				0.022	0.047	-0.089	0.018
	(5.73)		(-2.27)		(-5.22)	(-2.92)				(0.18)	(0.13)	(-1.42)	(0.36)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past decomposed short, medium and long-term variables from July 1969 to December 2020 only with firms whose share prices greater than 5 dollars. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2003. Panel A, Panel B and Panel C run the FM regression with book-to-market, retained earnings-to-marekt and contributed capital-to-market related variables, respectively. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$. $r^{i,*}_{t-5,t}$ represents past five-year earning returns computed by the underlying value. $r^{i,*}_{t-5,t-1}$ and $r^{i,*}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1_{/12}}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1_{/12},t}$ is the prior one-month stock return. ind^{i}_{RE} is the negative indicator to distinguish the negative retained earnings from the positive. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

Table 2.20: FM regression for all-but-micro information

Panel A: BTM related variables

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.138	0.132										
	(5.34)	(1.93)										
(2)	1.106		0.009									
	(5.25)		(0.17)									
(3)	1.294			-0.159								
	(6.57)			(-3.22)								
(4)	1.24				-0.138							
	(6.26)				(-2.62)							
(5)	1.198					-0.364						
	(5.83)					(-3.44)						
(6)	1.25		-0.045	-0.163								
	(6.25)		(-0.88)	(-3.49)								
(7)	1.204		-0.011	-0.107			-0.082					
	(6.09)		(-0.2)	(-2.39)			(-0.89)					
(8)	1.166		-0.004		-0.08			-0.112				
	(5.99)		(-0.08)		(-1.67)			(-1.28)				
(9)	1.13		-0.02		-0.082			-0.106	0.012			
	(5.84)		(-0.37)		(-1.85)			(-1.26)	(0.07)			
(10)	1.144		-0.034		-0.093	-0.222		-0.056	0.079			
	(5.91)		(-0.65)		(-2.08)	(-3)		(-0.66)	(0.47)			
(11)	1.155		-0.032		-0.086	-0.217		-0.067		0.09	0.004	
	(6.04)		(-0.63)		(-2)	(-3.03)		(-0.8)		(0.56)	(0.01)	
(12)	1.2		-0.062		-0.131	-0.26				0.163	0.07	-0.077
	(6.14)		(-1.27)		(-2.73)	(-3.47)				(1.1)	(0.15)	(-0.94)

Panel B: RTM related variables

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.258	0.172											-0.379
	(5.99)	(3.45)											(-2.69)
(2)	1.133		0.021										-0.257
	(5.38)		(0.57)										(-1.91)
(3)	1.231			-0.079									-0.36
	(6.47)			(-1.77)									(-2.34)
(4)	1.206				-0.07								-0.335
	(6.35)				(-1.47)								(-2.22)
(5)	1.171					-0.187							-0.295
	(5.77)					(-2)							(-2.33)
(6)	1.201		-0.012	-0.086									-0.329
	(6.12)		(-0.32)	(-1.93)									(-2.13)
(7)	1.198		0.04	-0.001			-0.161						-0.321
	(6.11)		(1.03)	(-0.02)			(-1.79)						(-2.19)
(8)	1.181		0.038		0.006			-0.176					-0.334
	(6.13)		(1)		(0.13)			(-2.05)					(-2.21)
(9)	1.139		0.031		0.004			-0.173	0.026				-0.328
	(5.95)		(0.83)		(0.09)			(-2.09)	(0.15)				(-2.23)
(10)	1.139		0.033		0.007	-0.009		-0.177	0.023				-0.327
	(5.95)		(0.87)		(0.17)	(-0.14)		(-2.12)	(0.14)				(-2.27)
(11)	1.151		0.035		0.014	-0.012		-0.185		0.036	-0.027		-0.328
	(6.08)		(0.95)		(0.33)	(-0.19)		(-2.26)		(0.22)	(-0.06)		(-2.3)
(12)	1.14		-0.022		-0.078	-0.105				0.223	0.164	-0.19	-0.31
	(5.93)		(-0.62)		(-1.68)	(-1.54)				(1.5)	(0.35)	(-2.37)	(-2.12)

Panel C: CTM related variables

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.106	0.002											0.088
	(5.03)	(0.06)											(1.33)
(2)	1.106		-0.011										0.067
	(5.06)		(-0.41)										(1.09)
(3)	1.284			-0.172									-0.053
	(6.39)			(-4.57)									(-1.35)
(4)	1.27				-0.173								-0.046
	(6.32)				(-4.48)								(-1.19)
(5)	1.176					-0.336							0.02
	(5.65)					(-4.87)							(0.48)
(6)	1.233		-0.045	-0.18									-0.019
	(5.84)		(-1.62)	(-4.76)									(-0.35)
(7)	1.211		-0.037	-0.151			-0.074						-0.023
	(5.87)		(-1.37)	(-4.48)			(-0.9)						(-0.43)
(8)	1.183		-0.041		-0.164			-0.069					-0.019
	(5.85)		(-1.54)		(-4.89)			(-0.88)					(-0.37)
(9)	1.142		-0.047		-0.16			-0.071	0.015				-0.022
	(5.69)		(-1.77)		(-5.1)			(-0.93)	(0.09)				(-0.42)
(10)	1.143		-0.046		-0.14	-0.145		-0.058	0.041				-0.028
	(5.69)		(-1.73)		(-4.37)	(-3.1)		(-0.77)	(0.25)				(-0.52)
(11)	1.15		-0.046		-0.136	-0.138		-0.066		0.055	-0.072		-0.025
	(5.79)		(-1.72)		(-4.37)	(-3)		(-0.89)		(0.35)	(-0.15)		(-0.49)
(12)	1.176		-0.051		-0.155	-0.144				0.126	-0.005	-0.073	-0.021
	(5.66)		(-1.85)		(-4.34)	(-3.04)				(0.85)	(-0.01)	(-1)	(-0.4)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on past decomposed short, medium and long-term variables from July 1969 to December 2020 only with firms whose market capitalisation is greater than 20th percentile of NYSE market cap at the end of June each year. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. Panel A, Panel B and Panel C run the FM regression with book-to-market, retained earnings-to-marekt and contributed capital-to-market related variables, respectively. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, $rm_{i,t}$ and $cm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$. $r^{i,*}_{t-5,t}$ represents past five-year earning returns computed by the underlying value. $r^{i,*}_{t-5,t-1}$ and $r^{i,*}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1_{/12}}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1_{/12},t}$ is the prior one-month stock return. ind^{i}_{RE} is the negative indicator to distinguish the negative retained earnings from the positive. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

stock returns with t-statistics of 3.45. In the case of the data sample excluding share prices below 5 dollars, all other ratios are irrelevant at the 5% significant level, of which the current period $bm_{i,t}$ and lagged $cm_{i,t-5}$ can forecast the future stock returns.

Third, the performance of past value returns varies. Past CC returns continue to dominate other variables in the regressions. However, past RE returns cannot provide any incremental explanatory power for forecasting the monthly stock returns. Surprisingly, the explanatory power of past BE returns is statistically better than that in the previous analysis. The significant performance can be explained by the highly negative correlation between past BE returns and btm. The past long-term and medium-term BE returns can considerably forecast the monthly stock returns in All-but-microcaps. Ball et al. (2015) find that $log(BE/ME)$ can significantly explain the expected stock returns in the micro-cap data sample than those in All-but-microcaps (t-value, 8 compared 4), indicating the small firms can have higher value effects than big firms.

2.5.2.3 Monthly updated regressions

Due to the large lags between the data collection and construction, the past stock returns, specifically the medium-term and short-term stock returns, cannot adequately explain monthly stock returns in the above annual data analysis. In this part, I extend the analysis by constructing FM regressions using the monthly updated data sample and build the table differently by including one price-scaled ratio in each table. In each table, Panel A shows the regressions based on a full data sample without data restriction. Then, Panel B and Panel C employ the data reducing constraints, no share prices below 5 dollars or no micro firms.

Table 2.21 displays the results of FM regressions based on the decomposi-

Table 2.21: FM regression of monthly updated further decomposition for BE-based variables

Panel A: Full data sample without restriction

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IB}_{t-5,t}$
(1)	1.32	0.337										
	(5.56)	(4.68)										
(2)	1.289		0.107									
	(5.54)		(2.18)									
(3)	1.435			-0.235								
	(5.86)			(-4.66)								
(4)	1.39				-0.211							
	(5.78)				(-4.37)							
(5)	1.317					-0.525						
	(5.49)					(-3.96)						
(6)	1.403		0.014	-0.221								
	(5.79)		(0.26)	(-4)								
(7)	1.288		0.108	-0.074			-0.305					
	(5.52)		(2.01)	(-1.49)			(-2.75)					
(8)	1.325		0.128		-0.043			-0.34				
	(5.65)		(2.62)		(-1.12)			(-3.37)				
(9)	1.233		0.124		-0.052			-0.337	-0.119			
	(5.44)		(2.52)		(-1.42)			(-3.54)	(-0.62)			
(10)	1.241		0.118		-0.059	-0.11		-0.307	-0.126			
	(5.48)		(2.38)		(-1.62)	(-1.45)		(-3.25)	(-0.66)			
(11)	1.225		0.113		-0.05	-0.131		-0.286		0.482	-6.082	
	(5.23)		(2.25)		(-1.29)	(-1.74)		(-2.98)		(2.52)	(-12.4)	
(12)	1.266		0.016		-0.188	-0.268				0.78	-5.777	-0.296
	(5.28)		(0.3)		(-3.45)	(-3.11)				(4.45)	(-12.43)	(-3.1)

Panel B: Firms with share prices greater than 5 dollars

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IB}_{t-5,t}$
(1)	1.15	0.14										
	(5.35)	(2.1)										
(2)	1.162		0.101									
	(5.5)		(2.07)									
(3)	1.264			-0.149								
	(6.12)			(-3.44)								
(4)	1.225				-0.134							
	(5.94)				(-2.97)							
(5)	1.179					-0.299						
	(5.59)					(-3.07)						
(6)	1.239		0.051	-0.116								
	(5.96)		(1.1)	(-3.02)								
(7)	1.163		0.053	-0.107			-0.028					
	(5.58)		(0.98)	(-2.3)			(-0.28)					
(8)	1.199		0.1		-0.036			-0.141				
	(5.85)		(1.97)		(-0.95)			(-1.71)				
(9)	1.074		0.101		-0.025			-0.152	0.429			
	(5.38)		(2.06)		(-0.73)			(-1.88)	(2.14)			
(10)	1.08		0.097		-0.031	-0.087		-0.129	0.426			
	(5.41)		(1.97)		(-0.88)	(-1.35)		(-1.57)	(2.14)			
(11)	1.074		0.098		-0.023	-0.106		-0.12		0.859	-4.25	
	(5.26)		(1.96)		(-0.63)	(-1.63)		(-1.42)		(4.19)	(-9.42)	
(12)	1.089		0.056		-0.073	-0.152				0.986	-4.115	-0.127
	(5.33)		(1.23)		(-1.92)	(-2.41)				(5.11)	(-9.58)	(-1.52)

Panel C: All-but-micro

Reg \Var	Constant	$bm_{i,t}$	$bm_{i,t-5}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IB}_{t-5,t}$
(1)	1.147	0.131										
	(5.31)	(1.68)										
(2)	1.15		0.075									
	(5.53)		(1.44)									
(3)	1.291			-0.158								
	(6.65)			(-3.27)								
(4)	1.241				-0.141							
	(6.37)				(-2.78)							
(5)	1.19					-0.344						
	(5.86)					(-3.27)						
(6)	1.249		0.012	-0.138								
	(6.32)		(0.24)	(-3.13)								
(7)	1.159		0.026	-0.123			-0.043					
	(5.66)		(0.44)	(-2.45)			(-0.38)					
(8)	1.207		0.066		-0.061			-0.133				
	(6.11)		(1.23)		(-1.35)			(-1.39)				
(9)	1.085		0.074		-0.06			-0.153	0.335			
	(5.53)		(1.4)		(-1.46)			(-1.65)	(1.44)			
(10)	1.092		0.065		-0.066	-0.155		-0.113	0.329			
	(5.58)		(1.24)		(-1.58)	(-2.18)		(-1.19)	(1.42)			
(11)	1.102		0.071		-0.054	-0.168		-0.108		0.681	-3.647	
	(5.55)		(1.34)		(-1.26)	(-2.35)		(-1.12)		(2.85)	(-7.06)	
(12)	1.121		0.026		-0.103	-0.214				0.798	-3.53	-0.114
	(5.7)		(0.53)		(-2.36)	(-3.11)				(3.57)	(-7.07)	(-1.2)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on book equity related past decomposed short, medium and long-term variables from July 1969 to December 2020. All variables are constructed at the end of each month and span monthly. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. Panel A runs the regression with full data sample without restriction. Panel B only includes the share prices greater than 5 dollar. Panel C contains firms whose market capitalisation is greater than 20th percentile of NYSE market cap at the end of June each year. The regressions contain earnings to market ratio in year t, $bm_{i,t}$, lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$. $r^{i,BE}_{t-5,t}$ represents past five-year earning book equity returns. $r^{i,BE}_{t-5,t-1}$ and $r^{i,BE}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1_{/12}}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1_{/12},t}$ is the prior one-month stock return. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

tion of past five-year stock returns and BE-related variables. From univariate regressions (1) to (5) in three panels, the data reducing process reduces the explanatory power of all variables in predicting monthly stock returns. Variables in the sample with share prices above 5 dollars have higher predictive ability than variables in the All-but-microcaps sample. But the performance of the medium-term BE return in All-but-microcaps, $r^{i,BE}_{t-1,t}$, is slightly higher than it in the sample without share prices below 5 dollars (t-value, -3.27 vs -3.07). $bm_{i,t-5}$ can significantly explain the expected stock returns in the multivariate regression with long-term and medium-term BE returns and stock returns included. Moreover, long-term and medium-term BE returns perform similarly to the annual data since BE can only be obtained annually from the COMPUSTAT.

Only medium-term BE return, $r^{i,BE}_{t-1,t}$, can significantly capture expected stock returns in regressions (10) and (11) based on the All-but-micro data. In addition, the prior five-year stock return, $r^{i}_{t-5,t}$, is only useful in the full data sample (t-value, -2.75), as opposed to the other two data samples. The decomposition of the five-year stock return yields three distinct results: $r^{i}_{t-5,t-1}$ is significant in the full data sample, $r^{i}_{t-1,t}$ is significant in the sample with share prices greater than 5 dollars, and both are negligible in the All-but-microcaps. Remarkably, when decomposing the prior one-year return into one-year momentum $r^{i}_{t-1,t-1_{/12}}$ and the prior one-month stock return $r^{i}_{t-1_{/12},t}$, one-year momentum is positively significant and the prior one-month return can negatively predict monthly stock returns. The finding is consistent with Asness (1995) who find a considerable relationship between the average of short- and medium-term past returns and monthly stock returns.

Table 2.22 provides information about RE-based variables. Unlike BE-based regressions, $rm_{i,t}$ and $rm_{i,t-5}$ captures the future stock returns in the majority of regressions, especially in regressions with the decomposed long-term

Table 2.22: FM regression monthly update for RE-based variables

Panel A: Full data sample without restriction

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.373	0.237											-0.075
	(5.97)	(4.26)											(-0.48)
(2)	1.282		0.071										0.016
	(5.82)		(2.03)										(0.09)
(3)	1.332			-0.115									-0.034
	(6.09)			(-2.96)									(-0.22)
(4)	1.318				-0.12								-0.02
	(6.17)				(-3)								(-0.12)
(5)	1.248					-0.176							0.05
	(5.61)					(-1.77)							(0.34)
(6)	1.326		0.026	-0.108									-0.028
	(6.05)		(0.72)	(-2.74)									(-0.17)
(7)	1.289		0.116	0.047			-0.37						-0.178
	(5.85)		(3.01)	(1.09)			(-3.49)						(-1.23)
(8)	1.342		0.108		0.023			-0.367					-0.15
	(6.19)		(3.09)		(0.62)			(-3.87)					(-1)
(9)	1.258		0.111		0.019			-0.37	-0.122				-0.181
	(5.89)		(3.14)		(0.53)			(-4.07)	(-0.65)				(-1.29)
(10)	1.239		0.125		0.028	0.237		-0.413	-0.145				-0.169
	(5.8)		(3.5)		(0.77)	(3.43)		(-4.41)	(-0.77)				(-1.21)
(11)	1.228		0.123		0.034	0.212		-0.394		0.463	-6.157		-0.175
	(5.54)		(3.38)		(0.9)	(3.07)		(-4.14)		(2.43)	(-12.66)		(-1.26)
(12)	1.2		0.017		-0.137	0.045				0.867	-5.742	-0.403	-0.039
	(5.47)		(0.49)		(-3.33)	(0.66)				(4.96)	(-12.47)	(-4.26)	(-0.25)

Panel B: Firms with share prices greater than 5 dollars

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.256	0.158											-0.319
	(5.83)	(3.16)											(-2.52)
(2)	1.212		0.075										-0.275
	(5.79)		(2.1)										(-2.19)
(3)	1.222			-0.07									-0.285
	(6.14)			(-1.78)									(-2.14)
(4)	1.207				-0.071								-0.27
	(6.11)				(-1.7)								(-2.06)
(5)	1.165					-0.102							-0.229
	(5.64)					(-1.2)							(-2.01)
(6)	1.229		0.054	-0.043									-0.292
	(6.08)		(1.6)	(-1.19)									(-2.15)
(7)	1.166		0.063	-0.021			-0.071						-0.28
	(5.66)		(1.63)	(-0.52)			(-0.77)						(-2.21)
(8)	1.234		0.092		0.023			-0.168					-0.28
	(6.17)		(2.62)		(0.67)			(-2.16)					(-2.12)
(9)	1.112		0.094		0.023			-0.174	0.464				-0.3
	(5.67)		(2.7)		(0.71)			(-2.25)	(2.34)				(-2.51)
(10)	1.105		0.101		0.028	0.08		-0.195	0.452				-0.285
	(5.66)		(2.87)		(0.84)	(1.33)		(-2.44)	(2.28)				(-2.39)
(11)	1.101		0.102		0.033	0.062		-0.187		0.88	-4.233		-0.279
	(5.5)		(2.87)		(0.95)	(1.01)		(-2.29)		(4.31)	(-9.43)		(-2.34)
(12)	1.085		0.057		-0.039	-0.007				1.071	-4.034	-0.191	-0.297
	(5.48)		(1.71)		(-1.03)	(-0.11)				(5.53)	(-9.47)	(-2.36)	(-2.41)

Panel C: All-but-micro

Reg \Var	Constant	$rm_{i,t}$	$rm_{i,t-5}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.266	0.177											-0.356
	(5.81)	(3.08)											(-2.33)
(2)	1.186		0.069										-0.277
	(5.72)		(1.84)										(-1.88)
(3)	1.213			-0.072									-0.306
	(6.41)			(-1.6)									(-1.91)
(4)	1.192				-0.066								-0.284
	(6.3)				(-1.38)								(-1.81)
(5)	1.156					-0.149							-0.247
	(5.76)					(-1.58)							(-1.86)
(6)	1.21		0.046	-0.045									-0.303
	(6.2)		(1.24)	(-1.02)									(-1.87)
(7)	1.142		0.064	-0.017			-0.089						-0.315
	(5.55)		(1.46)	(-0.37)			(-0.8)						(-2.15)
(8)	1.223		0.093		0.029			-0.168					-0.288
	(6.25)		(2.46)		(0.66)			(-1.81)					(-1.85)
(9)	1.103		0.1		0.025			-0.187	0.377				-0.317
	(5.65)		(2.62)		(0.62)			(-2.04)	(1.63)				(-2.25)
(10)	1.096		0.105		0.028	0.037		-0.198	0.371				-0.311
	(5.64)		(2.68)		(0.69)	(0.57)		(-2.11)	(1.6)				(-2.23)
(11)	1.112		0.11		0.036	0.019		-0.193		0.716	-3.584		-0.313
	(5.64)		(2.77)		(0.86)	(0.3)		(-2.05)		(2.99)	(-6.96)		(-2.26)
(12)	1.086		0.057		-0.037	-0.05				0.914	-3.382	-0.197	-0.319
	(5.63)		(1.52)		(-0.83)	(-0.79)				(4.09)	(-6.85)	(-2.1)	(-2.18)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on retain earnings related past decomposed short, medium and long-term variables from July 1969 to December 2020. All variables are constructed at the end of each month and span monthly. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. Panel A runs the regression with full data sample without restriction. Panel B only includes the share prices greater than 5 dollar. Panel C contains firms whose market capitalisation is greater than 20th percentile of NYSE market cap at the end of June each year. The regressions contain earnings to market ratio in year t, $rm_{i,t}$, lagged earnings to market ratio in year $t-5$, $rm_{i,t-5}$. $r^{i,RE}_{t-5,t}$ represents past five-year earning book equity returns. $r^{i,RE}_{t-5,t-1}$ and $r^{i,RE}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1_{/12}}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1_{/12},t}$ is the prior one-month stock return. ind^{i}_{RE} is the negative indicator to distinguish the negative retained earnings from the positive. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

and medium-term returns. $r^i_{t-5,t-1}$ only has the minor forecasting capacity for monthly stock returns, but, similar to BE-based regressions, one-year stock momentum and prior one-month return can have substaintial explanatory power. Last, ind^i_{RE} can greatly identify the negative and positive RE in samples with data reducing constraints in Table 2.22 Panel B and Table 2.22 Panel C.

Table 2.23 displays the average FM regression information based on CC-related variables. $cm_{i,t}$ and $cm_{i,t-5}$ are statistically significant in the univariate regressions under the full data sample (t-value, 3.05 and 2.12). In contrast, neither ratios is required for predicting the future stock returns in a constrained data sample. Other past decomposed components do not vary significantly. Past CC returns still hold strong predictive ability under any circumstances. The performance of prior stock returns is in line with BE-based regressions in Table 2.21. Moreover, ind^i_{CC} is still inconsequential. In the underlying data sample, the monthly stock returns can be captured significantly by any period CC returns, $r_{i,CC}(t-5,t)$, $r_{i,CC}(t-5,t-1)$ and $r_{i,CC}(t-1,t)$, and no more than one-year stock returns, $r_i(t-1,t)$, $r_i(t-1_{/12}, t-1_{/1})$ and $r_i(t-1_{/1}, t)$.

Overall, the monthly updated decomposition demonstrates that CC-based regressions leave the least amount of unexplained expected stock returns for intangible returns compared to the other two earning ratios. It means that *ctm* decomposition can exceed *btm* and *rtn* in term of the value factor. The result in this part conflicts with BGLN, who conclude that the *ctm* is the worst of three factors used to estimate the monthly stock returns.

2.5.2.4 Merging with $size_{i,t}$

To explain monthly stock returns, Novy-Marx (2013), Ball et al. (2015) and BGLN apply the size factor, $size_{i,t}(log(ME))$ as a control variable. Table 2.24 presents the results of multivariate FM regressions mixed $size_{i,t}$ with past de-

Table 2.23: FM regression monthly update for CC-based variables

Panel A: Full data sample without restriction

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.379	0.155											-0.063
	(5.23)	(3.05)											(-0.61)
(2)	1.341		0.067										-0.038
	(5.35)		(2.12)										(-0.41)
(3)	1.415			-0.221									-0.082
	(5.93)			(-5.64)									(-1.06)
(4)	1.403				-0.224								-0.078
	(5.93)				(-5.79)								(-1.04)
(5)	1.316					-0.499							-0.027
	(5.5)					(-6.03)							(-0.36)
(6)	1.417		0.004	-0.213									-0.084
	(5.66)		(0.15)	(-6.09)									(-0.92)
(7)	1.325		0.016	-0.154			-0.288						-0.009
	(5.4)		(0.5)	(-4.26)			(-2.75)						(-0.12)
(8)	1.378		0.024		-0.156			-0.299					-0.023
	(5.57)		(0.77)		(-4.37)			(-2.98)					(-0.31)
(9)	1.284		0.018		-0.169			-0.299	-0.113				-0.004
	(5.4)		(0.58)		(-5.33)			(-3.17)	(-0.6)				(-0.05)
(10)	1.288		0.02		-0.149	-0.148		-0.286	-0.118				-0.012
	(5.41)		(0.65)		(-4.58)	(-3.09)		(-3.04)	(-0.62)				(-0.17)
(11)	1.268		0.016		-0.139	-0.157		-0.269		0.488	-6.075		-0.013
	(5.18)		(0.54)		(-4.13)	(-3.27)		(-2.81)		(2.57)	(-12.48)		(-0.19)
(12)	1.286		0.002		-0.192	-0.178				0.77	-5.787	-0.282	-0.07
	(5.24)		(0.08)		(-5.42)	(-3.71)				(4.37)	(-12.33)	(-2.92)	(-0.82)

Panel B: Firms with share prices greater than 5 dollars

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.105	-0.001											0.106
	(4.92)	(-0.03)											(1.77)
(2)	1.156		0.023										0.063
	(5.3)		(0.87)										(1.13)
(3)	1.259			-0.176									-0.007
	(6.05)			(-5.01)									(-0.17)
(4)	1.252				-0.185								-0.005
	(6.04)				(-5.13)								(-0.12)
(5)	1.167					-0.349							0.041
	(5.52)					(-5.25)							(0.9)
(6)	1.231		-0.027	-0.181									0.018
	(5.69)		(-1.11)	(-5.47)									(0.34)
(7)	1.156		-0.027	-0.176			-0.017						0.009
	(5.28)		(-1.06)	(-5.37)			(-0.19)						(0.16)
(8)	1.225		-0.015		-0.165			-0.099					0.025
	(5.74)		(-0.63)		(-5.18)			(-1.28)					(0.5)
(9)	1.1		-0.018		-0.159			-0.106	0.446				0.031
	(5.29)		(-0.77)		(-5.62)			(-1.4)	(2.24)				(0.66)
(10)	1.1		-0.016		-0.141	-0.122		-0.092	0.443				0.026
	(5.28)		(-0.67)		(-4.9)	(-2.91)		(-1.23)	(2.23)				(0.56)
(11)	1.095		-0.013		-0.132	-0.126		-0.086		0.873	-4.201		0.018
	(5.15)		(-0.54)		(-4.5)	(-2.99)		(-1.12)		(4.28)	(-9.26)		(0.38)
(12)	1.095		-0.025		-0.152	-0.133				0.968	-4.104	-0.095	0.021
	(5.19)		(-1.08)		(-4.82)	(-3.14)				(4.96)	(-9.4)	(-1.24)	(0.44)

Panel C: All-but-micro

Reg \Var	Constant	$cm_{i,t}$	$cm_{i,t-5}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1_{/12},t_{/1}}$	$r^{i}_{t-1_{/12},t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.114	0.014											0.069
	(5.01)	(0.41)											(1.02)
(2)	1.168		0.031										0.026
	(5.46)		(1.1)										(0.43)
(3)	1.28			-0.173									-0.051
	(6.44)			(-4.72)									(-1.28)
(4)	1.265				-0.174								-0.042
	(6.36)				(-4.61)								(-1.09)
(5)	1.17					-0.339							0.022
	(5.68)					(-4.97)							(0.51)
(6)	1.259		-0.014	-0.171									-0.039
	(6.01)		(-0.53)	(-4.73)									(-0.69)
(7)	1.176		-0.01	-0.164			-0.044						-0.047
	(5.45)		(-0.37)	(-4.7)			(-0.42)						(-0.84)
(8)	1.248		-0.002		-0.146			-0.103					-0.03
	(6.05)		(-0.06)		(-4.25)			(-1.13)					(-0.54)
(9)	1.118		-0.004		-0.148			-0.119	0.36				-0.027
	(5.49)		(-0.17)		(-4.91)			(-1.36)	(1.55)				(-0.53)
(10)	1.119		-0.002		-0.127	-0.136		-0.103	0.355				-0.032
	(5.5)		(-0.07)		(-4.2)	(-2.95)		(-1.18)	(1.53)				(-0.61)
(11)	1.132		0.003		-0.118	-0.139		-0.1		0.701	-3.576		-0.044
	(5.49)		(0.13)		(-3.85)	(-3.05)		(-1.13)		(2.94)	(-6.92)		(-0.82)
(12)	1.131		-0.01		-0.138	-0.146				0.808	-3.472	-0.106	-0.038
	(5.45)		(-0.36)		(-4.18)	(-3.18)				(3.61)	(-6.89)	(-1.19)	(-0.71)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on contributed capital related past decomposed short, medium and long-term variables from July 1969 to December 2020. All variables are constructed at the end of each month and span monthly. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. Panel A runs the regression with full data sample without restriction. Panel B only includes the share prices greater than 5 dollar. Panel C contains firms whose market capitalisation is greater than 20th percentile of NYSE market cap at the end of June each year. The regressions contain earnings to market ratio in year t, $cc_{i,t}$, lagged earnings to market ratio in year $t-5$, $cc_{i,t-5}$. $r^{i,CC}_{t-5,t}$ represents past five-year earning book equity returns. $r^{i,CC}_{t-5,t-1}$ and $r^{i,CC}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1_{/12}}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1_{/12},t}$ is the prior one-month stock return. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

composition components for the All-but-microcaps.

Except for the past value returns, all variables are rebalanced at the end of each month in regressions. In mixed regressions $size_{i,t}$ has a marginally significant explanatory power near the 5% significant level. In BE-based regressions, $size_{i,t}$ fails to predict monthly stock returns, which is consistent with Fama and French (2015), who find the *size* factor is redundant in the five-factor asset pricing model. Moreover, none of the other variables vary markedly. In Table 2.24 Panel C, the CC-related variables explain the monthly stock returns better than the oterh two price-scaled ratios, even though $cm_{i,t}$ is inconsequential.

2.6 Conclusion

This chapter introduces the retained earnings and contributed capital proposed by Ball et al. (2020) (BGLN) into the decomposition of log book-to-market ratio. I evaluate the performance using three distinct data samples. The first one is a full sample with no restriction. The other two utilise the data filtering criterion of share prices greater than 5 dollars and market capitalisation exceeding the 20th percentile of the NYSE. According to BGLN, I assume that the RE-related variables can outperform BE in terms of prediction performance and estimate a smaller intangible return on the annual data.

The result of annual data indicates that the current period *rtm* will dominate *btm* and *ctm* with the data sample shrinkage and the RE indicator shows a statistically significant trend in distinguishing the values in the regressions with restricted data, which is consistent with BGLN. However, past *rtm* or long-term RE returns cannot provide explanatory power in forecasting the stock returns. Past CC returns can have good predictive power for the future stock returns, possibly due to the considerable correlation between CC returns and long-term

Table 2.24: FM regression of monthly returns on mixed past variables under all-but-micro data samples

Panel A: BTM related variables

Variables	Constant	$bm_{i,t-5}$	$size_{i,t}$	$r^{i,BE}_{t-5,t}$	$r^{i,BE}_{t-5,t-1}$	$r^{i,BE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IB}_{t-5,t}$
(1)	1.479	0.004	-0.053	-0.118			-0.029					
	(4.3)	(0.07)	(-1.62)	(-2.42)			(-0.26)					
(2)	1.532	0.04	-0.054		-0.059			-0.113				
	(4.35)	(0.77)	(-1.63)		(-1.34)			(-1.24)				
(3)	1.427	0.049	-0.055		-0.059			-0.137	0.322			
	(4.25)	(0.95)	(-1.73)		(-1.45)			(-1.53)	(1.4)			
(4)	1.444	0.04	-0.056		-0.065	-0.152		-0.097	0.318			
	(4.33)	(0.79)	(-1.78)		(-1.59)	(-2.21)		(-1.06)	(1.39)			
(5)	1.47	0.046	-0.058		-0.054	-0.165		-0.091		0.671	-3.665	
	(4.36)	(0.89)	(-1.82)		(-1.28)	(-2.4)		(-0.99)		(2.86)	(-7.01)	
(6)	1.482	0.007	-0.057		-0.098	-0.207				0.771	-3.567	-0.097
	(4.42)	(0.14)	(-1.8)		(-2.27)	(-3.04)				(3.53)	(-7.05)	(-1.06)

Panel B: RTM related variables

Variables	Constant	$rm_{i,t-5}$	$size_{i,t}$	$r^{i,RE}_{t-5,t}$	$r^{i,RE}_{t-5,t-1}$	$r^{i,RE}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IR}_{t-5,t}$	ind^{i}_{RE}
(1)	1.473	0.053	-0.055	-0.016			-0.079						-0.35
	(4.27)	(1.26)	(-1.71)	(-0.34)			(-0.72)						(-2.52)
(2)	1.564	0.076	-0.056		0.024			-0.148					-0.318
	(4.44)	(2.12)	(-1.72)		(0.55)			(-1.68)					(-2.12)
(3)	1.466	0.085	-0.058		0.02			-0.171	0.353				-0.35
	(4.38)	(2.32)	(-1.89)		(0.49)			(-1.95)	(1.54)				(-2.6)
(4)	1.461	0.09	-0.058		0.023	0.037		-0.183	0.349				-0.343
	(4.38)	(2.41)	(-1.9)		(0.57)	(0.6)		(-2.04)	(1.53)				(-2.57)
(5)	1.491	0.095	-0.06		0.031	0.02		-0.176		0.692	-3.586		-0.346
	(4.42)	(2.51)	(-1.94)		(0.74)	(0.32)		(-1.96)		(2.95)	(-6.89)		(-2.61)
(6)	1.464	0.047	-0.06		-0.037	-0.043				0.874	-3.401	-0.18	-0.355
	(4.41)	(1.3)	(-1.93)		(-0.83)	(-0.7)				(4)	(-6.8)	(-2.01)	(-2.53)

Panel C: CTM related variables

Variables	Constant	$cm_{i,t-5}$	$size_{i,t}$	$r^{i,CC}_{t-5,t}$	$r^{i,CC}_{t-5,t-1}$	$r^{i,CC}_{t-1,t}$	$r^{i}_{t-5,t}$	$r^{i}_{t-5,t-1}$	$r^{i}_{t-1,t}$	$r^{i}_{t-1/12,t/1}$	$r^{i}_{t-1/12,t}$	$r^{i,IC}_{t-5,t}$	ind^{i}_{CC}
(1)	1.495	-0.029	-0.057	-0.154			-0.027						-0.018
	(4.31)	(-1.07)	(-1.79)	(-4.58)			(-0.27)						(-0.36)
(2)	1.583	-0.021	-0.06		-0.137			-0.082					-0.003
	(4.48)	(-0.8)	(-1.82)		(-4.14)			(-0.96)					(-0.06)
(3)	1.477	-0.023	-0.062		-0.14			-0.102	0.354				0.002
	(4.38)	(-0.95)	(-2)		(-4.75)			(-1.23)	(1.55)				(0.03)
(4)	1.478	-0.021	-0.062		-0.12	-0.128		-0.087	0.349				-0.003
	(4.39)	(-0.83)	(-2)		(-4.06)	(-2.81)		(-1.05)	(1.53)				(-0.06)
(5)	1.507	-0.016	-0.064		-0.112	-0.132		-0.083		0.695	-3.588		-0.011
	(4.43)	(-0.62)	(-2.04)		(-3.72)	(-2.91)		(-0.99)		(2.97)	(-6.86)		(-0.24)
(6)	1.506	-0.026	-0.064		-0.132	-0.138				0.785	-3.501	-0.089	-0.008
	(4.46)	(-0.98)	(-2.03)		(-4.01)	(-3.04)				(3.57)	(-6.85)	(-1.05)	(-0.16)

Note: This table reports the Fama-MacBeth regression of monthly stock returns on size factor and past information mixed with different data construction. All past price-scaled ratios and return on earnings are constructed at the end of June every year while the past stock returns are computed at the end of each month and span monthly. Regressions only covers the firms with ME greater than 20th percentile of NYSE market cap, from July 1969 to December 2020. Panel A, Panel B and Panel C run the FM regression with book-to-market, retained earning-to-marekt and contributed capital-to-market related variables, respectively. The regressions contain lagged earnings to market ratio in year $t-5$, $bm_{i,t-5}$, $rm_{i,t-5}$ and $cm_{i,t-5}$. $r^{i,*}_{t-5,t}$ represents past five-year earning returns computed by the underlying value. $r^{i,*}_{t-5,t-1}$ and $r^{i,*}_{t-1,t}$ are the value returns for past five years skip the last year and the prior one year. $r^{i}_{t-5,t}$ is the prior five-year return from the last trading day of $t-6$ to t. $r^{i}_{t-5,t-1}$ represents the stock return from year $t-5$ to year $t-1$. $r^{i}_{t-1,t}$ is the prior one-year stock return from the current period. $r^{i}_{t-1,t-1/12}$, momentum is the prior one-year stock return skip the last one month. $r^{i}_{t-1/12,t}$ is the prior one-month stock return. ind^{i}_{RE} is the negative indicator to distinguish the negative retained earnings from the positive. All coefficients in this table has been times 100. Fama-MacBeth t-statistics are in round parentheses with Newey and West (1987) adjustment.

issuance effects. Intangible returns in RE-based regressions are more robust than the other two value measurements, indicating that the decomposition of *rtm* has left a greater proportion of expected stock returns unexplained by the past RE-related variables.

Asness (1995) proposes that the prior short-term and long-term stock returns have a different influence on the prediction of expected stock return. I extend the model to a monthly basis that rolls the accounting information at the end of each month rather than collecting it at the end of December each year. Also, I decompose past earnings and stock returns into short- (one month), medium- (no more than one year) and long-term (over one year) variables. In general, monthly construction can increase the significance of price-deflated variables but not the relevance of other variables, particularly annual updated past value returns.

The outcome shows that the stock momentum and short-term stock returns can forecast the monthly returns significantly more than any other variables as long as they can be updated monthly. In contrast, the long-term stock return loses significance in the multivariate regressions. Besides, the BE- and RE-related variables are insufficient to predict the stock returns significantly. Regardless of the data sample size and time intervals, they have the most vital ability to explain the future stock return with any time range information.

Overall, the past information has a significant influence on the performance of earning ratios. BGLN mention that RTM can outperform BTM because retained earnings or free cash flow are more directly associated. Nevertheless, past information will diminish the significance of cumulative cash flow. By incorporating the profitability return and long-term past stock returns, the lag will distort the connection between earnings and future stock return, which causes retained earnings to underperform.

Chapter 3

Asset pricing with managed portfolios and macroeconomic variables

3.1 Introduction

In the asset pricing field, the main challenge is determining the risk-return relationship, which is equivalent to explaining stock returns using some common risk factors (or characteristics). Sharpe (1964), Lintner (1965) and Mossin (1966) develop the capital asset pricing model (CAPM). They depict the risk level of specific assets in relation to *beta*, the representative of the systematic risk factor. Then, Fama and MacBeth (1973) (FM) introduce the famous two-step regression to test CAPM. Their method reduces cross-sectional bias in regressions, and they demonstrate the effectiveness of CAPM and market portfolios. More broadly, Ross (1976) proposes the Arbitrage Pricing Theory (APT) to have a comparable single-factor model from no-arbitrage arguments. The fundamental principle is that the pricing model based on APT can be simply generalised to include multiple sources of risk or factors.

Scholars also investigate common factors that have a substantial impact on asset pricing. size and book-to-market ratio are explanatory variables for capturing expected stock returns, according to Fama and French (1992). The three-factor model (FF3), which accounts for the market, size and value factors, is a groundbreaking innovation introduced by Fama and French (1993). They creatively use long-short position excess returns as the factors to examine the risk exposure of a specific sorted portfolio to the specific factor. Researchers and practitioners join the game of mining new firm characteristics by utilising the FF3 and the FM two-step test methodologies. However, the vast majority of them focus on one or two new factors simultaneously.

Some academics, on the other hand, begin to concentrate on the large dimensional matrix of firm characteristics. Green et al. (2017) document 94 characteristics in order to test the joint significance in a single regression. They discover that 12 characteristics can accurately predict expected stock returns. Kelly et al. (2019) then create adjusted portfolios based on the 36 selected characteristics. Their objective is to determine whether the large-dimensional characteristic adjustment can produce a significant $alpha$ out of the systematic risk and have good prediction. This chapter is motivated by Kelly et al. (2019) and I adjust the stock returns to construct managed portfolios by a large firm characteristic pool, which includes 126 characteristics based on Green et al. (2017).

Above academics look for the common risk factors or anamaly based on equity field, while a group of scholars concentrate how macroeconomic factors affects the prices of assets. Chen et al. (1986) apply the Fama and MacBeth two-step regressions (FM regressions) and identify the significant macroeconomic factors, industrial production, inflation, the bond risk premium, and default, that can influence the expected stock returns. The result demonstrates that macro factors can be exposed to systemic risk with a substantial risk premium . Cochrane (1996) concludes that variation in stock expected returns should be linked to real macroeconomic risk, and he discovers an investment factor, derived from the production side, can adequately capture expected stock returns.

Subsequently, Harvey et al. (2016) list risk factors of large dimensions, including 40 macroeconomic variables. According to their regression, macro variables only have a marginal effect on predicting the expected stock returns, compared to significant equity firm characteristics. McCracken and Ng (2016) establish a monthly updated FRED-MD database with 134 time series based on Stock and Watson (1996) criteria. Using sparse principal component analysis, Rapach and Zhou (2021) recently derive the sparse macroeconomic factors from 120 FRED-MD macro variables (based on McCracken and Ng (2016) dataset). They successfully identify two significant macro factors, *yields* and *housing*, that can perfectly explain the sorted portfolio returns and generate a significant risk premium.

According to the preceding research, macroeconomic variables may have explanatory power on managed portfolios, which is the main point of this chapter, in which I use a methodology to investigate the relationship between a large panel of macroeconomic variables and expected stock returns mixed with a large dimension of stock firm characteristics.

FM regressions and Fama and French (1993) portfolio sorting strategy have been the dominant methods for deriving risk factors and examining the performance of specific factors in explaining expected stock returns over the last few decades. The advantage of FM two-step regression over traditional cross-sectional regression is that it can reduce the cross-sectional correlation of pricing errors. The regressions with a large dimension of right-hand side variables, however, are not fit by it.

On the other hand, it is necessary to have all prior information available without missing data for portfolio sorting, and only information related to the specific factor can be included in the regressions. It is not possible to include all published characteristics in a single multifactor pricing model. When a large number of firm characteristics are included, the portfolio sorting will become complicated to build factors. For this reason, some studies begin by decomposing the variance-covariance matrix of a large panel of sorted portfolio excess returns in order to identify common latent factors. According to Kelly et al. (2019), the advantage of factor analysis is that it reduces the dimension of the factor list and finds common risk factors from a large sample, and all correlated information is included in a multi-factor model.

The conventional Principal Component Analysis (PCA) is the foundational machine learning method for identifying the most significant common risk factors for the underlying assets. Kelly et al. (2019, 2020) propose the Instrumented Principal Component Analysis (IPCA), which has several advantages over the conventional PCA method. First, firm characteristics are introduced as instrumented variables to connect expected stock returns and PCA factors. Second, because firm characteristics change over time, they convert the static pricing model to a dynamic model and use the Alternating Least Square (ALS) method to estimate the factor and loadings. As a result, the IPCA model can measure dynamic loadings that the conventional PCA model cannot. Finally, they use the firm characteristics matrix to calculate managed portfolio returns based on the sorted portfolios. The factors and loadings are then applied to individual firm analysis. Overall, IPCA can combine managed portfolio returns, a large dimension of firm characteristics, and unobservable latent factors in a dynamic model and derive significant latent factors with fewer pricing errors than other factor models, where time-varying firm characteristics contribute significantly to explaining expected stock returns.

This study replicates Kelly et al. (2020) IPCA and further changes the model from dynamic to static in order to keep the right-hand side components constant, implying that the left-hand side returns are a combination of firm characteristics and stock excess returns.[1] I can easily use the principal component to estimate the latent factors and static loadings throughout all periods once the independent variables are static. In addition, I incorporate macroeconomic variables into the modified model to investigate the relationship between macroeconomic variables and expected stock returns adjusted for firm characteristics, whereas the IPCA model cannot do so.

Pukthuanthong et al. (2019) mention that, the level data of non-tradable macroeconomic variables only have mediocre explanatory power in measuring expected stock returns. Vassalou (2003) claims that creating macro factor mimicking portfolios (MPs) is an effective method for eliminating errors and measuring non-tradable macro variables. This conventional method can be used, but Vassalou (2003) emphasises the importance of the mimicking portfolios in relation to the base assets, which means that only the appropriate assets can be used to create mimicking factor portfolios.

Aside from the mimicking portfolios, Giglio and Xiu (2021) (GX) propose a three-pass method (3PM) to derive the latent factors by PCA from underlying excess stock returns and conduct the regression to estimate the risk premium of observed but non-tradable variables based on the estimated latent factors. The key benefit of 3PM is that it helps to lower measurement errors for non-tradable variables and omitted variable bias if some variables are missing. Without building macro factor mimicking portfolios, I use 3PM to identify the unobserved latent factors and calculate the risk premium of a sizable panel of individual

[1] The full procedures will be explained in the later section.

macro variables.

This chapter employs the 3PM in four parts to estimate the risk premium of macroeconomic variables. To create the characteristic-adjusted managed portfolios, I first gather a sizable panel of firm characteristics from the Centre for Research in Security Prices (CRSP) and Compustat between June 1987 and June 2017 and multiply the individual stock excess returns with firm characteristics. Then, using the 3PM, I estimate the latent factors using classic PCA and Risk-premium Principal Component Analysis (RP-PCA from Lettau and Pelger (2020)), and I determine the risk premium associated with the latent factors. To calculate the risk premium of macro variables, I run the regression of 126 observed macroeconomic variables from FRED-MD on the latent factors from the first step of 3PM.

There are multiple contributions and findings in this chapter.

First, I investigate and compare the performance of various data adjustment methods on excess stock returns using a large panel of firm characteristics from Green et al. (2017). In this chapter, I experiment with three different data adjustment methods for the firm characteristics matrix, winsorization from Green et al. (2017) (GHZ), normalised rank transformation by Kozak et al. (2020) (KNS) and centred rank transformation from Kelly et al. (2019) (KPS).[2] I find that KPS centred rank transformation has a more stable sample distribution than the other two methods after comparing the results of R^2 for three different methods.

[2]Green et al. (2017) winsorize the characteristics at 1% and 99% and normalise the winsorized data sample. Then, set the missing data to zero (mean value). Kozak et al. (2020) rank a specific characteristic value over a given time period and divide the rank value by the total number of firms plus one. Then demean the rank result and standardise using the sum of the absolute values of demeaned results. Kelly et al. (2019) modifies the rank value in a unique way. They locate the rank between minus 0.5 to 0.5 by shrinking the characteristic rank between zero and one and minus 0.5 for all values. The values of the characteristics will be modified cross-sectionally rather than across the entire data sample in all three methods.

Second, the main contribution of this chapter is to investigate the relationship between macroeconomic variables and the characteristic-adjusted managed portfolios, where few papers address both aspects. I can estimate the risk premium of macro variables using 3PM without rebuilding. Only one of the fifth macro variables has significant explanatory power in describing managed portfolios, according to the findings. This outcome is consistent with the findings of Giglio and Xiu (2021), who discover that majority of macro variables are insufficient to price equity assets. Through mimicking portfolios (MPs), I also examine the relationship between managed portfolios and macroeconomic variables. However, the 3PM contradicts the findings of MPs. The macroeconomic variables with a significant risk premium from both methods have no similarities and cannot be grouped together.

Finally, I derive the three macro-factor model by using PCA and weak-factor principal component analysis (WF-PCA from Uematsu and Yamagata (2020)) and compare it with four popular multi-factor pricing models. All models fall short of explaining the characteristic-adjusted managed portfolio returns more effectively than the long-short portfolio returns when the pricing errors of multi-factor models are considered. Additionally, Fama and French (1993) three-factor model (FF3) and Fama and French (2015) five-factor model (FF5) are two asset pricing models that the macro-factor model cannot outperform, indicating that the macroeconomic variables have a limited ability to explain equity returns.

This chapter expands on the literature in three different areas. First, document the firm's characteristics in order to estimate expected stock returns and investigate the significance of a large number of characteristics. Scholars have discovered approximately 400 significant characteristics that can capture the exposure to expected stock returns over the last several decades.

Green et al. (2013) summarise a large panel with 330 published characteristics. Hou et al. (2015) build factor mimicking portfolios using characteristics to test the performance of their q-theory multifactor model (HXZ5). McLean and Pontiff (2016) collect 97 characteristics and analyse analyses how those features have evolved over time in the literature. Their conclusion suggests that the explanatory power of the specific characteristic will decline after the publication. Kozak et al. (2020) derive the Stochastic Discount Factor (SDF) from a large panel of firm characteristics. Freyberger et al. (2020) propose a non-linear model for selecting significant variables to explain cross-sectional expected stock returns from 62 characteristics.

The second area is concerned with determining the relationship between macro variables and expected stock returns. Following the discovery of four significant macro factors from ten candidates that can explain the expected stock return by Chen et al. (1986), some scholars approach new factors from different directions. Campbell and Shiller (1988) find that the dividend-price ratio appears to explain long-term expected returns. Lettau and Ludvigson (2001) evidence that the consumption-wealth ratio has strong explanatory power in predicting the excess stock returns and is better than dividend-price ratio to forecast the future returns. Ang and Bekaert (2004) propose a regime-switching strategy to earn higher international equity returns than a static model, implying that regimes have a significant impact on portfolio returns. According to Ang and Bekaert (2007), the dividend yields can predict the excess return in the short-term with short-term interest rates, whereas short-term interest rates have strong negative predictability on excess returns. Evidence of forecasting power can be found in different countries. Rapach et al. (2010) mention that equity premium is forecastable due to the connection with the real economy. They discern that the business cycle has the ability to raise the risk premium when the economy is in a bad state. Cochrane (2011) refers to the significance of discount rate variation for macroeconomics.

Scholars, on the other hand, attempt to combine the variables and conduct a joint analysis. One aspect of this chapter is concerned with the aggregation of macro variables. Stock and Watson (1996) first summarise 76 series macro variables and test the series stability. Ludvigson and Ng (2009) extend the macroeconomic panel to 131 time-series and estimate the common factor to evaluate the risk premium of bonds. Then, McCracken and Ng (2016) adjust the selection criteria and select 134 macroeconomic time series to build the "big data" set. Another aspect is the dimension reduction of a large panel of macroeconomic variables. Ludvigson and Ng (2007) find three latent factors interpreted, as *volatility*, *risk premium* and *real* can have significant predictability to forecast expected excess stock returns. Rapach and Zhou (2021) utilise the sparse principal component analysis (sparse PCA) method to group 120 macroeconomic variables into 10 sparse macro factors and figure out the link between excess stock returns and macro variables in order to construct the multi-factor pricing model. Furthermore, Harvey et al. (2016) document a large number of characteristics from the previously published papers, including 40 macroeconomic variables. They suggest a new criteria in which only variables' t-statistics above the threshold of *3.0* are deemed significant.

This study also builds on the dimension reduction method to derive the common unobservable factor that captures expected stock returns. Stock and Watson (2002) and Bai and Ng (2002) introduce the PCA method and develop the criteria to determine the number of factors in the pricing model. Fan et al. (2016) propose the Projected Principal Component Analysis (Projected-PCA). They find that projected PCA outperforms conventional PCA in terms of accuracy, and the derived factors can be related to firm characteristics. Freyberger et al. (2020) employ the least absolute shrinkage and selection operator (group LASSO) to examine which firm characteristic has incremental explanatory power on expected returns. They find only 13 anomalies for the period from 1965 to 2014. Lettau and Pelger (2020) introduce risk-premium PCA (RP-PCA). They apply the first-moment penalty to the regression objective function and use the principal component to generate five significant latent factors. They test latent factors using portfolio excess returns from 37 firm characteristics. The result shows that RP-PCA can outperform conventional PCA in both in-sample and out-of-sample estimation. Kozak et al. (2020) refer to the PCA method with L^1 and L^2 penalties to add sparsity and economic shrinkage restriction. They estimate the latent factors with 50 firm characteristics adjustment and derive the SDF based on the estimated latent factors. They reveal evidence that SDF calculated by combination PCA method can have a better out-of-sample performance than built by the characteristics directly.

The remaining sections of this chapter are structured as follows: Section 3.2 demonstrates the core methodology, how to construct the managed portfolio and how to utilise the three-pass method. Section 3.3 describles the data collection and data washing processes in detail. Section 3.5 reports empirical construction and main finding, and Section 3.6 concludes with remarks.

3.2 Methodology

This section revisits the original Instrumented Principal Component Analysis (IPCA) and three-pass method (3PM). Then, I show how to use the modified 3PM to estimate the risk premium of non-tradable observed macroeconomic variables based on the excess returns of characteristics-adjusted managed portfolios. I also go over the in-sample and out-of-sample tests used to assess performance. Finally, I mention two different PCA methods for determining latent factors.

I start with some fundamental information in this chapter:

- T is the number of time periods;
- L is the number of firm characteristics;
- N or N_t is the number of firms in a specific period t;
- M is the number of macroeconomic variables.
- γ_g is the risk premium of macroeconomic variables.
- γ is the first-moment penalty in the risk-premium PCA (RP-PCA) method.
- $\bar{F}$ is the risk premium of latent factors in the first step of 3PM.

3.2.1 Instrumented Principal Component Analysis

I start with the procedure how to link the large panel of firm characteristics with individual excess stock returns. The idea is motivated by Kelly et al. (2019) Instrumented Principal Component analysis (IPCA) model. They construct the multifactor model for individual stock excess returns, $r_{i,t+1}$:

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}. \tag{3.1}$$

where $\alpha_{i,t} = z'_{i,t}\Gamma_\alpha$, $\beta_{i,t} = z'_{i,t}\Gamma_\beta$ and $z_{i,t}$ is the $N \times L$ instrumented firm characteristic matrix containing all financial information for each individual stock during the period t.

In the vector notation, I denote the regression of individual excess stock return vector r_{t+1} in period $t+1$ is:

$$\underbrace{r_{t+1}}_{N\times 1} = \underbrace{Z_t}_{N\times L}\underbrace{\Gamma_\alpha}_{L\times 1} + \underbrace{Z_t}_{N\times L}\underbrace{\Gamma_\beta}_{L\times K}\underbrace{f_{t+1}}_{K\times 1} + \underbrace{\epsilon_{t+1}}_{N\times 1}, \tag{3.2}$$

The key pattern of this model is the variation of the characteristics matrix Z_t varies with time t. The combination of latent factor information and firm characteristics can improve the estimation for the excess stock returns (Kelly et al. (2019)). The dynamic right-hand side can make estimation more complicated than static regression, which can simply apply the ordinary least squares (OLS).

As a result, they refer to an important step in introducing managed portfolios,

$$x_{t+1} = \frac{Z_t' r_{t+1}}{N_{t+1}}, \tag{3.3}$$

where r_{t+1} is $N \times 1$ excess stock returns vector for period $t+1$ and x_{t+1} is $L \times 1$ managed portfolio vector for L firm characteristics.

The model can be rewritten as the managed portfolio pattern:

$$x_{t+1} = W_t \Gamma_\alpha + W_t \Gamma_\beta f_{t+1} + \epsilon_{t+1} \tag{3.4}$$

where $W_t = \frac{Z_t' Z_t}{N_{t+1}}$.

Managed portfolios can aid in reducing the sensitivity of returns in order to produce a stable estimator. As an example, Kelly et al. (2019) use the alternating least squares (ALS) on characteristics-adjusted managed portfolio returns, x_{t+1}, to estimate latent factors and loadings following principal component analysis (PCA).

This chapter converts the model from dynamic to static for estimating the factor in a simple structure by premultiplying the equation with $(Z_t' Z_t)^{-1}$.[3] The cross-sectional model for period $t+1$ will then be:

$$\begin{aligned}(Z_t' Z_t)^{-1} Z_t' r_{t+1} &= \Gamma_\alpha + \Gamma_\beta f_{t+1} + (Z_t' Z_t)^{-1} Z_t \epsilon_{t+1} \\ r_{z,t+1} &= \Gamma_\alpha + \Gamma_\beta f_{t+1} + \epsilon_{z,t+1}\end{aligned} \tag{3.5}$$

where $r_{z,t+1}$ is the $L \times 1$ managed portfolio return vector for period $t+1$.

When all time periods are considered, the time-series model for characteristics-adjusted managed portfolio excess return, R_z, will be:

$$\underbrace{R_z}_{T\times L} = \underbrace{\mathbb{1}}_{T\times 1} \underbrace{\Gamma_\alpha'}_{1\times L} + \underbrace{F}_{T\times K} \underbrace{\Gamma_\beta'}_{K\times L} + \underbrace{\epsilon_z}_{T\times L} \tag{3.6}$$

where $\mathbb{1}$ is $T \times 1$ vector of ones.

Kelly et al. (2019) set the restricted model with $\Gamma_\alpha = 0$ and apply the PCA and ALS to the managed portfolio regressions to obtain the factors and loadings. The restricted model's estimation goal is to minimise the sum of squared errors:

$$\min_{\Gamma_\beta, F} \frac{1}{T} \sum_{t=1}^{T-1} \left(r_{z,t+1} - \Gamma_\beta f_{t+1}\right)' \left(r_{z,t+1} - \Gamma_\beta f_{t+1}\right) \tag{3.7}$$

Following Kelly et al. (2019) restricted model, the right-hand-side factors can be decomposed into two parts, risk premium of factors ($\bar{F}$) and demeaned factor matrix ($\tilde{F}$). The model will be as follows:

$$R_z = \mathbb{1}\, \bar{F}\, \Gamma_\beta' + \tilde{F}\, \Gamma_\beta' + \epsilon_z \tag{3.8}$$

This equation is similar to Giglio and Xiu (2021) original three-pass regres-

[3] I would like to thank my supervisor, Dr. Adam Golinski, for suggesting the system rotation technique for adjusting the dynamic IPCA model to the static version.

sion. In order to estimate the rik premium of macroeconomic variables, I can therefore introduce the characteristics-adjusted managed portfolio to 3PM.

3.2.2 Three-pass Method

Next, I describe how the 3PM calculates the risk premium of observed non-tradable variables. The method's main goal is to identify some latent factors that can act as an intermediary between objectively observed but non-tradable variables. They use the PCA method to identify the latent factor, which has the robust explanatory power of test assets, and they assume that the latent factor is correlated with observed variables.

In three-pass method, the first step is to create a multifactor model to estimate the latent factor:

$$r_t = \beta\gamma_v + \beta v_t + u_t, \tag{3.9}$$

where r_t is $N \times 1$ excess return matrix, β is $N \times K$ factor loadings, v_t is the $K \times 1$ innovation factor matrix, with zero mean for each factor, and γ_v is the $K \times 1$ risk premium vector for k factors.

The 3PM estimates the risk premium of non-tradable factors using v_t. The non-tradable factors g_t are assumed to have relations with PC factors v_t in Giglio and Xiu (2021). The regression between derived PC factor, v_t, and the non-tradable factor, g_t, is as follows:

$$g_t = \delta + \eta v_t + e_t, \tag{3.10}$$

where η denotes the level of explanation of latent factors on observed variables.

Furthermore, Giglio and Xiu (2021) define the risk premium of observed non-tradable variables, γ_g, as η, the coefficient relation between g_t and v_t times the risk premium of derived factors v_t,

$$\gamma_g = \eta \gamma_v. \tag{3.11}$$

3.2.3 Modified three-pass method for firm characteristics

Next, I combine the concept of managed portfolio analysis with 3PM from Giglio and Xiu (2021) to determine the relationship between macroeconomic variables and managed portfolio excess returns derived adjusted by firm characteristics.

Following Giglio and Xiu (2021), I begin by adjusting the regression in the first step of 3PM. Let R_z denote the characteristic-adjusted managed portfolio excess returns:

$$\underbrace{R_z}_{T\times L} = \underbrace{\mathbb{1}}_{T\times 1} \underbrace{\bar{F}}_{1\times K} \underbrace{\Gamma'_\beta}_{K\times L} + \underbrace{\tilde{F}}_{T\times K} \underbrace{\Gamma'_\beta}_{K\times L} + \underbrace{U}_{T\times L}, \tag{3.12}$$

where $\tilde{F}$ is the $T \times K$ demeaned latent factor matrix, Γ_β is the $L \times K$ factor loading matrix and $\bar{F}$ is the $K \times 1$ risk premium vector.

Let the $T \times M$ matrix, G, denote M observable macroeconomic variables.

$$\underbrace{G}_{T\times M} = \underbrace{\tilde{F}}_{T\times K} \underbrace{\eta'}_{K\times M} + \underbrace{E}_{T\times M}, \tag{3.13}$$

where η is a $M \times K$ matrix representing the latent factor's exposure to observed macroeconomic variables.

The steps of 3PM involved in this study are outlined here:

1. First, I utilise dimensional reduction method to derive latent factors and factor loadings, denoting $r_{z,t+1}$ as the characteristics-adjusted portfolios excess returns, f_{t+1} as the PC derived factors, Γ_β as the PC factor loadings.[4]

[4] In the first step, I use conventional PCA to decompose the variance of excess returns. How-

The equation is:

$$\underbrace{r_{z,t+1}}_{L\times 1} = \underbrace{\Gamma_\beta}_{L\times K}\underbrace{f_{t+1}}_{K\times 1} + \underbrace{\epsilon_{z,t+1}}_{L\times 1}. \tag{3.14}$$

where Γ_β is $L \times K$ factor loading matrix and f_{t+1} is $K \times 1$ latent factor vector.

The equation is written as a matrix expression as follows:

$$\underbrace{R_z}_{T\times L} = \underbrace{F}_{T\times K}\underbrace{\Gamma'_\beta}_{K\times L} + \underbrace{U}_{T\times L}. \tag{3.15}$$

The objective function of conventional PCA is to minimise the sum squared errors:

$$\min_{\Gamma_\beta, f} \frac{1}{T}\sum_{t=1}^{T}(r_{z,t+1} - f_{t+1}\Gamma'_\beta)'(r_{z,t+1} - f_{t+1}\Gamma'_\beta). \tag{3.16}$$

It is equivalent to apply the decomposition to the variance-covariance matrix of R_z:

$$\frac{1}{T}(R_z - \bar{R}_z)'(R_z - \bar{R}_z) \tag{3.17}$$

As a result, the PCA estimator of factors and factor loadings are:[5]

$$\hat{F} = T^{1/2}\left(\xi_1 : \xi_2 : \ldots : \xi_{\hat{K}}\right)', \quad \text{and} \quad \hat{\Gamma}_\beta = \frac{1}{T}\tilde{R}'_z\hat{F}$$

where $\xi_1 : \xi_2 : \ldots : \xi_{\hat{K}}$ are the first K largest eigenvalues of the decomposition of covariance matrix of excess returns, and $\tilde{R}_z$ is demeaned characteristics-adjusted managed portfolios excess returns.

2. I perform a cross-sectional OLS regression of average excess portfolio returns, $\bar{R}_z$, on estimated latent factor loadings $\hat{\Gamma}_\beta$ from the first step through all periods to calculate the risk premium of PC factors,

ever, I also apply the RP-PCA by Lettau and Pelger (2020) as a comparison. The RP-PCA will be discuss on 3.2.5.

[5] The mean values of PC factor $\hat{F}$ are zero.

barF:[6]

$$\bar{F} = (\hat{\Gamma}'_{\beta}\hat{\Gamma}_{\beta})^{-1}\hat{\Gamma}'_{\beta}\bar{R}_z.$$

3. Finally, I assume that the macroeconomic variables are tied to PC factors. They can explain the movements of managed portfolios excess returns as a result of interaction. Run the time-series regression of macro variables, G_t, on PC factors, $\tilde{F}$ to estimate the coefficient η:

$$\hat{\eta} = G'\hat{F}(\hat{F}'\hat{F})^{-1}.$$

The risk premium of macro variables is then equal to

$$\hat{\gamma}_g = \hat{\eta}\bar{F}.$$

3.2.4 Testing

This section will go over how to create tests to evaluate the model's performance. Following Kelly et al. (2019) and Giglio and Xiu (2021), this study uses the goodness of fit, R^2 for different regressions.

First, I apply in-sample total-R^2 from Kelly et al. (2019) to assess how well the estimated PC factors, F, explain the characteristic-adjusted managed portfolio returns:

$$R^2_{tot} = 1 - \frac{\sum_{z,t}(r_{z,t+1} - \Gamma_{\beta} f_{t+1})^2}{\sum_{z,t} r^2_{z,t+1}}. \tag{3.18}$$

In addition, the in-sample predictive R^2 use the conditional static information of the expected value of factors,$\bar{f}$ through all periods.

$$R^2_{pred} = 1 - \frac{\sum_{z,t}(r_{z,t+1} - \Gamma_{\beta}\bar{f})^2}{\sum_{z,t} r^2_{z,t+1}}. \tag{3.19}$$

[6]The risk premium and demeaned factor will construct the model $\bar{R}_z = \bar{F}\hat{\Gamma}'_{\beta} + \hat{F}\hat{\Gamma}'_{\beta} + \epsilon_z$.

Second, I also investigate the effectiveness of PC factors derived from managed portfolios in explaning individual stock returns. Total-R^2 and predictive-R^2 are defined as:

$$
\begin{aligned}
R^2_{tot} &= 1 - \frac{\sum_{i=1}^{N}\sum_{t}^{T-1}(r_{i,t+1} - z_t\Gamma_\beta f_{k,t+1})^2}{\sum_{i=1}^{N}\sum_{t}^{T-1} r^2_{i,t+1}}, \\
R^2_{pred} &= 1 - \frac{\sum_{i=1}^{N}\sum_{t}^{T-1}(r_{i,t+1} - z_t\Gamma_\beta \bar{f}_k)^2}{\sum_{i=1}^{N}\sum_{t}^{T-1} r^2_{i,t+1}}.
\end{aligned}
\tag{3.20}
$$

In this study, the number of managed portfolios is determined by the number of firm characteristics. To compute the R^2, I will sum all portfolios together covering all sample periods to acquire a global result rather than the individual R^2 for each characteristic-adjusted managed portfolio.

Furthermore, I look at the out-of-sample (OOS) fit. I use all past information from the beginning up to period t to estimate the factors and loadings in period t and then apply the result for period $t+1$ to compute the predictive R^2. The out-of-sample window begins at 120. The OOS total-R^2 and predictive-R^2 will only be collected based on the information from June 1997 ($t = 121$) until the end since the first 120 periods do not have available results. The forecast OOS factor, f_{t+1}, in period $t+1$ is

$$f_{t+1} = (\Gamma'_\beta\Gamma_\beta)^{-1}\Gamma'_\beta r_{z,t+1}.$$

Finally, I check the explanatory power of latent factors on unobservable macroeconomic variables g_t. The goodness-of-fit are defined as follows:

$$R^2_g = 1 - \frac{\sum_{t=121}^{T-1}(g_{m,t+1} - \hat{f}_{t+1}\hat{\eta}'_m)^2}{\sum_{t=121}^{T-1} g^2_{m,t+1}} \tag{3.21}$$

The predictability of all latent factors on managed portfolio excess returns is considered in the first step of 3PM. However, in the third step, I want to determine the relationship between each macroeconomic variable and the test managed portfolios. So that I can calculate R_g^2 sequentially for all macroeconomic variables.

I also test the total explanatory power of all factors, so I have the R^2:

$$R_{sum}^2 = \frac{\sum_{m,t} \tilde{f}\eta'}{\sum_{m,t} g_{m,t}^2},\ for\ m = 1 \ldots M, \tag{3.22}$$

3.2.5 Additional Principal Component Analysis

In this chapter, I also applied two principal component approaches for different usage. The first one is the risk-premium PCA (RP-PCA) introduced by Lettau and Pelger (2020) and the second method is weak factor sparse orthogonal factor regression (WF-PCA) by Uematsu and Yamagata (2020).

3.2.5.1 Risk-premium Principal Component Analysis

I utilise the RP-PCA to replace the conventional PCA in the first stage of the 3PM. Lettau and Pelger (2020) build the RP-PCA based on the factor model:

$$R_z = F\,\Gamma_\beta' + U \quad l = 1, \ldots, L, t = 1, \ldots, T. \tag{3.23}$$

where X is the $T \times L$matrix of return for portfolios, F is $T \times K$ latent factor matrix and Γ_β is $L \times K$ factor loadings matrix. Then apply PCA to the covaraince matrix,

$$\Sigma_{RP} = \frac{1}{T} R_z' R_z + \gamma \bar{R}_z \bar{R}_z' \tag{3.24}$$

where γ is the first-moment penalty.

Some benefits of RP-PCA are highlighted by Lettau and Pelger (2020): First, it incorporates mean-related data into the objective function via the first-moment penalty. The Sharpe ratio can be increased by a large margin when using RP-PCA instead of the standard PCA method. Second, it is able to single out the weak factors, something that the conventional PCA fails to do. This chapter's objective is to examine the relationship between stock returns and macroeconomic variables, which usually has a weak relationship. So I add the RP-PCA as an alternative method in the first stage of 3PM.

3.2.5.2 Weak Factor Sparse Orthogonal Factor Regression

Uematsu and Yamagata (2020) introduce the WF-PCA based on the factor model:

$$R_z = F\,\Gamma'_\beta + U, \tag{3.25}$$

where X is $T \times L$ zero-mean stationary matrix, Γ_β is $L \times K$ factor loading matrix and F is $T \times K$ latent factors.

Using the weak factor sparse orthogonal factor regression (WF-SOFAR), the objective function is:

$$(\widehat{F}, \widehat{\Gamma}_\beta) = \underset{(F,\Gamma_\beta)}{\arg\min} \left\{ \frac{1}{2} \left\| R_z - F\Gamma'_\beta \right\|_{\mathrm{F}}^2 + \sigma_n \|\Gamma_\beta\|_1 \right\}, \tag{3.26}$$

where $\sigma_n \|\Gamma_\beta\|_1$ is a sparsity-inducing penalty term.

Rapach and Zhou (2021) prove that sparse information has a significant impact on the estimation of macroeconomic related factors. WF-PCA also introduces the sparsity information to improve the estimation of latent factors. In addition, like the RP-PCA, WF-PCA can have a consistent estimation even when the latent factors are weak. So I add the WF-PCA to produce the macroeconomic factors in the third stage of 3PM.

3.3 Data

This section illustrates the data process preceding the execution of empirical analysis. Data collection, data cleaning, firm characteristic adjustment and factor selection for the multi-factor model are all part of the process.

3.3.1 Firm characteristics

I first construct the characteristic matrix by following several previous papers. Hou et al. (2015) derive 74 anomalies from the Center for Research in Security Prices (CRSP), Compustat and I/B/E/S to examine their four-factor q-model. They categorise the characteristics into six groups based on their familiarity (Momentum, Values-versus-growth, Investment, Profitability, Intangibles and Trading frictions). Green et al. (2017) then test 93 characteristics from the Hou et al. (2015) database. They do not group the characteristics, but instead perform FM regressions for each one. Kozak et al. (2020) recently use 51 characteristics to construct the decile portfolios. Jensen et al. (2021) build a massive dataset with 153 characteristics from 93 countries. All of these papers share some characteristics.

In this case, I use the CRSP and Compustat databases to extract 126 characteristics. The characteristics that I construct are shown in Table 3.1.[7]

[7] The complete list and definition of characteristics can be found in Appendix Table A.1.

Table 3.1: List of characteristics

abbr.	name	abbr.	name	abbr.	name
prctmaxprc	Current Price To High Price Over Last Year	empch	Employee Growth Rate	stinvch	Change In Short-Term Investment
cash	Cash Holdings Scaled By Total Assets	gp	Gross Profitability	txtsup	Tax Expenses Surprise
taxch	Change In Tax Expense	gplag	Gross Profitability Lagged Assets	tacc	Total Accruals
cinvest	Change In Investment Over Average Of Prior Three Years Investment	capxch1	Change In Capital Expenditures 1 Year	mom1m	1-Month Momentum (Short-Term Reversal)
earnvol	Earning Volatility	capxch2	Change In Capital Expenditures 2 Years	mom12m	12-Month Momentum (Prior One Year Return Skip Last One)
roavol	Return On Assets Volatility	capxch3	Change In Capital Expenditures 3 Years	mom6m	6-Month Momentum 6-1
roaq	Return On Assets	capxpch	Percent Change In Capital Expenditure Over Prior 2 Years	mom36m	36-Month Momentum Skip Last Year

Table 3.1: continued from previous page

abbr.	name	abbr.	name	abbr.	name
roeq	Return On Book Equity	cepch	Growth In Common Shareholders Equity	mret60	Prior Five Years Return
rsup	Revenue Surprise	ltdpch	Growth In Long-Term Debt	chmom	Change In 6-Month Momentum
stdacc	Accrual Volatility	ltnoach	Growth In Long-Term Net Operating Assets Scaled By Average Total Assets	iss	Composite Issuance
cfvol	Cash Flow Volatility	invpch	Percent Change In Inventory 1 Year	dolvolm	Dollar Trading Volumn
abcinv	Abnormal Corporate Investment	invest	Property Investment Change Scaled By Assets	onep	1/Share Price
oacc	Operating Accruals	lgr	Percent Change In Liability	turn	Share Turnover
absoacc	Absolute Accruals	saleemppch	Labour Force Efficiency	maxret	Maximum Daily Return
pctacc	Percent Accruals	liqat	Liquidity Of Book Assets	retvol	Return Volatility
zscore	Altman Z-Score	ndtp	Net Debt To Price	baspread	High-Low Bid-Ask Spread

Table 3.1: continued from previous page

abbr.	name	abbr.	name	abbr.	name
atpch	Asset Growth 1 Year	nceqiss	Net Equity Issuance	zerotrade	Zero Trading Days
atan	Asset Tangibility	noa	Net Operating Assets	dolvol	Dollar Trading Volumn
ato	Asset Turnover	age	Number Of Year	beta	Market Beta
bleve	Book Leverage	oscore	Ohlson O-Score	betasq	Square Of Market Beta
cashdebt	Cash Flow To Debt	oleve	Operating Leverage	ill	Illiquidity
cboplag	Cash-Based Lagged Operating Profitability	op	Operating Profitability	stddolvol	Volatility Of Liquidity (Dollar Trading Volume)
cbop	Cash-Based Operating Profitability	oplag	Operating Profitability Lagged Book Assets	stdturn	Volatility Of Liquidity (Share Turnover)
cta	Cash-To-Assets	oaccni	Percent Accruals	beta_capm	Beta From Capm
ocfta	Operating Cash Flow -To-Assets	taccni	Percent Total Accruals	beta_down	Downside Beta From Capm
dtpch3	Change In Total Debt 3 Year	fscore	Pitroski F-Score	ivolcapm	Idiosyncratic Volatility Capm
flch3	Change In Financial Liabilities	qr	Quick Ratio	ivolff3	Idiosyncratic Volatility Ff3

Table 3.1: continued from previous page

abbr.	name	abbr.	name	abbr.	name
coach	Change In Current Operating Assets	rdctba	R&D Capital To Book Assets	ivolq	Idiosyncratic Volatility Qmodel
colch	Change In Current Operating Liabilities	rdat	R&D To Assets	ivoldown	Idiosyncratic Volatility Downside Beta From Capm
cowcch	Change In Current Operating Working Capital	rdind	R&D Increase	iskewcapm	Idiosyncratic Skewness Capm
invch	Change In Inventory	rdsale	R&D To Sales	iskewff3	Idiosyncratic Skewness Ff3
nfach	Change In Net Financial Assets	roic	Return On Invested Capital	iskewq	Idiosyncratic Skewness Qmodel
nncoach	Change In Net Noncurrent Operating Assets	ronoa	Return On Net Operating Assets	iskewdown	Idiosyncratic Skewness Downside
ncoach	Change In Noncurrent Operating Assets	salepch3	Sales Growth 3	me	Size
ncolch	Change In Noncurrent Operating Liabilities	salecash	Sales To Cash	atme	Assets-To-Market Capitalization

Table 3.1: continued from previous page

abbr.	name	abbr.	name	abbr.	name
dcv	Convertible Debt	pchgmpchsale	% Change In Gross Margin - % Change In Sales	btm	Book-To-Market Ratio
dcvind	Convertible Debt Indicator	pchsalepchxsga	% Change In Sales - % Change In Sg&A	cashpr	Cash Productivity
cr	Current Ratio	saleinv	Sales To Inventory	cfp	Cash-Flow-To-Price Ratio
crpch	% Change In Current Ratio	pchsaleinv	% Change In Sales-To-Inventory Ratio	divy	Dividend Yield
depr	Deprecitation Percentage	salerec	Sales To Receivables	ep	Earnings To Price Ratio
dppch	% Change In Depreciation	secured	Secured Debt Scaled By Total Liabilities	rdme	R&D To Market Capitalization
earnvari	Earnings Variability	securedind	Secured Debt Indicator	stp	Sales To Price

Note: This table contains summary information for all 126 characteristics. The left column of the group displays the abbreviation for each characteristic, while the right column describes the variable in detail.

The first category includes all available US market data from CRSP that is related to stock prices, returns, or trading activities. [8] All CRSP data are collected at the end of the month t and are available for the following month $t+1$.

Compustat-obtained accounting-based data comprise a separate group. All balance sheet data is obtained at fiscal year t for calendar year $t-1$ and will be available throughout the calendar year t. If the annual information is reported at the end of December in year $t-1$, it will be available from January to December in year t. If the data is collected on a quarterly basis, such as quarterly total assets (ATQ), I will do so at the end of the fiscal quarter. It will be available for the reminder of the quarter. Unlike previous empirical work, this chapter will not require a six-month lag between data collection and data establishment. Up-to-date data can improve the time-sensitivity of a large panel of characteristics matrix, allowing for the construction of adjusted managed portfolios with delay.

The critical step in managed portfolio construction is to apply $(Z_t'Z_t)^{-1}$ to excess stock returns, which necessitates the availability of all characteristic information. Table 3.2 showcases the availability of all variables from 1962 to 2018. Five characteristics ($roavol$, $stdacc$, $cfvol$, $cboplag$ and $earnvari$) have less than 60% availability, with $stdacc$ and $cfvol$ having less than 42% data available for use in managed portfolio construction. According to the full sample database, all data for a specific characteristic is missing for some periods, so I cannot use the data adjustment method to eliminate missing data. As a result, I build the non-missing managed portfolio from June 1987 to June 2017 to ensure that firms have available data for each period and characteristic. Table 3.3 counts the percentage of available sub-samples. All firm characteristics have 50% available

[8] Using Shumway (1997) method, I apply the common criteria to handle missing and delisting data of stock returns.

data at least. More importantly, I can use the data adjustment method to create a non-missing managed portfolio.[9]

[9]The data adjustment methods are explained in Section 3.4.

Table 3.2: Missing counting for all firm characteristics in full sample from 1962-2018

Row	Miss #	AVBL %	Row	Miss #	AVBL %	Row	Miss #	AVBL %
prctmaxprc	1492	99.95	emppch	146964	95.22	stinvch	153712	95.00
cash	460837	85.01	gp	22545	99.27	txtsup	208434	93.22
taxch	715148	76.74	gplag	205859	93.31	tacc	1225511	60.15
cinvest	847683	72.43	capxch1	426809	86.12	mom1m	1935	99.94
earnvol	1132826	63.16	capxch2	672551	78.13	mom12m	184419	94.00
roavol	1438951	53.20	capxch3	892875	70.96	mom6m	54768	98.22
roaq	516367	83.21	capxpch	672551	78.13	mom36m	670100	78.21
roeq	368474	88.02	cepch	216505	92.96	mret60	1090841	64.53
rsup	604815	80.33	ltdpch	146964	95.22	chmom	184419	94.00
stdacc	1786436	41.90	ltnoach	375324	87.79	iss	1090841	64.53
cfvol	1787729	41.86	invpch	285836	90.70	dolvolm	5677	99.82
abcinv	908485	70.46	invest	543599	82.32	onep	0	100.00
oacc	480174	84.38	lgr	207518	93.25	turn	1589	99.95
absoacc	480174	84.38	saleemppch	212216	93.10	maxret	101	100.00
pctacc	388008	87.38	liqat	507455	83.50	retvol	101	100.00
zscore	611780	80.10	ndtp	8252	99.73	baspread	62	100.00
atpch	196283	93.62	nceqiss	522630	83.00	zerotrade	0	100.00
atan	423684	86.22	noa	515397	83.24	dolvol	199581	93.51
ato	205331	93.32	age	0	100.00	beta	614494	80.02
bleve	14539	99.53	oscore	675017	78.05	betasq	614494	80.02
cashdebt	527702	82.84	oleve	22471	99.27	ill	199642	93.51
cboplag	1287207	58.14	op	22545	99.27	stddolvol	199581	93.51
cbop	1197572	61.05	oplag	205859	93.31	stdturn	196657	93.60
cta	196018	93.63	oaccni	1140634	62.91	beta_capm	4750	99.85
ocfta	1140613	62.91	taccni	1225511	60.15	beta_down	44001	98.57
dtpch3	609238	80.19	fscore	0	100.00	ivolcapm	4750	99.85
flch3	613739	80.04	qr	539776	82.45	ivolff3	4752	99.85
coach	663160	78.43	rdctba	613739	80.04	ivolq	100964	96.72
colch	627614	79.59	rdat	8252	99.73	ivoldown	44001	98.57
cowcch	510015	83.41	rdind	0	100.00	iskewcapm	4750	99.85
invch	285836	90.70	rdsale	21744	99.29	iskewff3	4752	99.85
nfach	153712	95.00	roic	132710	95.68	iskewq	100964	96.72
nncoach	515397	83.24	ronoa	130005	95.77	iskewdown	44001	98.57
ncoach	663220	78.43	salepch3	698442	77.29	me	54522	98.23
ncolch	635091	79.35	salecash	21744	99.29	atme	54522	98.23
dcv	364677	88.14	pchgmpchsale	213391	93.06	btm	54522	98.23
dcvind	0	100.00	pchsalepchxsga	285679	90.71	cashpr	54522	98.23
cr	510006	83.41	saleinv	100200	96.74	cfp	1145166	62.76
crpch	665703	78.35	pchsaleinv	294926	90.41	divy	67696	97.80
depr	167584	94.55	salerec	97215	96.84	ep	58859	98.09
dppch	343150	88.84	secured	1063631	65.41	rdme	54522	98.23
earnvari	1401058	54.44	securedind	0	100.00	stp	65983	97.85

Note: This table displays the statistics for missing data for all 126 characteristics. The first column displays the characteristic's name, the second column displays the number of missing data (Miss #), and the third column displays the percentage of available data (AVBL %). The total sample size is *3,065,497*. The sample includes data from 1962 to 2018.

Table 3.3: Missing counting for all characteristics subsample from 1987 to 2017

Row	Miss #	AVBL %	Row	Miss #	AVBL %	Row	Miss #	AVBL %
prctmaxprc	654	99.97	emppch	98289	95.39	stinvch	101890	95.22
cash	64194	96.99	gp	13294	99.38	txtsup	130899	93.86
taxch	237835	88.85	gplag	126359	94.08	tacc	328230	84.61
cinvest	345765	83.79	capxch1	271718	87.26	mom1m	888	99.96
earnvol	664544	68.84	capxch2	435129	79.60	mom12m	118670	94.44
roavol	737255	65.43	capxch3	581663	72.73	mom6m	32302	98.49
roaq	112975	94.70	capxpch	435129	79.60	mom36m	443205	79.22
roeq	65519	96.93	cepch	121329	94.31	mret60	711170	66.66
rsup	263226	87.66	ltdpch	98289	95.39	chmom	118670	94.44
stdacc	1030125	51.70	ltnoach	235816	88.94	iss	711170	66.66
cfvol	1031310	51.65	invpch	167663	92.14	dolvolm	2122	99.90
abcinv	592432	72.22	invest	399708	81.26	onep	0	100.00
oacc	239048	88.79	lgr	126559	94.07	turn	151	99.99
absoacc	239048	88.79	saleemppch	132134	93.80	maxret	42	100.00
pctacc	148492	93.04	liqat	415734	80.51	retvol	42	100.00
zscore	474877	77.74	ndtp	3845	99.82	baspread	0	100.00
atpch	120808	94.34	nceqiss	304295	85.73	zerotrade	0	100.00
atan	315912	85.19	noa	420484	80.29	dolvol	2424	99.89
ato	126287	94.08	age	0	100.00	beta	418903	80.36
bleve	3883	99.82	oscore	514011	75.90	betasq	418903	80.36
cashdebt	379716	82.20	oleve	13187	99.38	ill	2465	99.88
cboplag	388069	81.81	op	13294	99.38	stddolvol	2424	99.89
cbop	299991	85.93	oplag	126359	94.08	stdturn	9	100.00
cta	120602	94.35	oaccni	244845	88.52	beta_capm	2599	99.88
ocfta	244824	88.52	taccni	328230	84.61	beta_down	18279	99.14
dtpch3	410273	80.76	fscore	0	100.00	ivolcapm	2599	99.88
flch3	413249	80.62	qr	430007	79.84	ivolff3	2599	99.88
coach	507178	76.22	rdctba	413249	80.62	ivolq	82982	96.11
colch	489212	77.06	rdat	3845	99.82	ivoldown	18279	99.14
cowcch	417114	80.44	rdind	0	100.00	iskewcapm	2599	99.88
invch	167663	92.14	rdsale	13146	99.38	iskewff3	2599	99.88
nfach	101890	95.22	roic	71874	96.63	iskewq	82982	96.11
nncoach	420484	80.29	ronoa	69313	96.75	iskewdown	18279	99.14
ncoach	507217	76.22	salepch3	456885	78.58	me	28039	98.69
ncolch	493792	76.85	salecash	13146	99.38	atme	28039	98.69
dcv	218091	89.77	pchgmpchsale	132471	93.79	btm	28039	98.69
dcvind	0	100.00	pchsalepchxsga	159185	92.54	cashpr	28039	98.69
cr	417114	80.44	saleinv	51934	97.57	cfp	248955	88.33
crpch	508533	76.16	pchsaleinv	174690	91.81	divy	39524	98.15
depr	113785	94.67	salerec	38035	98.22	ep	30932	98.55
dppch	221604	89.61	secured	400865	81.21	rdme	28039	98.69
earnvari	922595	56.74	securedind	0	100.00	stp	35483	98.34

Note: This table displays the statistics for missing data for all 126 characteristics. The first column displays the characteristic's name, the second column displays the number of missing data (Miss #), and the third column displays the percentage of available data (AVBL %). The full data sample size is *1,629,000*. The sample covers the data sub-sample from 1987 to 2017.

3.3.2 Macroeconomic variables

McCracken and Ng (2016) introduce the large panel database with 130 macro variables. I follow them and collect 127 individual macro variables from the FRED-MD database as observed variables in the Giglio and Xiu three-pass method to estimate the risk premium.[10]

To correspond with managed portfolio returns, the macro variables begin in June 1987 to June 2017. I use data transformation to adjust the raw macro variables to stationary, as suggested by McCracken and Ng (2016). The FRED-MD database provides a time-series transformation code that can be used to differentiate between different transformation cases, such as the first difference or second difference.[11] The missing values are replaced by the unconditional mean for each variable after the transformation. Finally, all macro variables are normalised, and outliers are eliminated.[12]

I also give the reason for not choosing the other two methods. First, although KNS has a good forecasting capability on characteristics-adjusted managed portfolio returns and OOS-SR, I cannot apply it to the 3PM since it provides less significant characteristics and has bad in-sample and out-of-sample performance. Also, managed portfolio returns have much larger volatility than the other two, which will considerably decrease the utility of excess returns. Second, GHZ's winsorization generates fewer of significant firm characteristics

[10]FRED-MD:
`https://research.stlouisfed.org/econ/mccracken/fred-databases`, the Federal Reserve Bank of St. Louis' monthly database for macroeconomic research.

[11]McCracken and Ng (2016) describe how to adjust the data emplying transformation code. The database contains six scenarios, each with its specific code. (1) no adjustment; (2) change in variable; (3) square of change in variable; (4) natural logarithm of variable; (5) change in natural logarithm of variable; (6) square of change in natural logarithm of variable; (7) change in percent change of variable.

[12]The outlier is defined as the absolute difference between the underlying macroeconomic value and mean that is greater than 10 times the difference between the 75% percentile and 25% percentile of underlying macroeconomic variable (McCracken and Ng (2016)).

most of the time, and in-sample performance is also worse than KPS. Generally, the out-of-sample R^2 and OSS-SR cannot dominate KPS's results.

3.4 Comparision of data transform adjustment method

Before conducting empirical regression analysis, I compare three different data adjustment methods for optimising the firm characteristics matrix: the rank transformation from Kelly et al. (2019) (KPS) and Kozak et al. (2020) (KNS), as well as data winsorization from Green et al. (2017) (GHZ). Overall, KPS outperforms the other two approaches. This chapter also establishes the 3PM based on KPS with five latent factors. When apply Lettau and Pelger (2020) RP-PCA, I set the first-moment penalty γ equal to 5.

3.4.1 Characteristics data transform adjustment

This chapter uses three different methods to adjust raw data and handle missing data to improve the availability and stability of firm characteristics. The first method is used by Green et al. (2017) (GHZ), who winsorize the accounting information at 1% and 99% and normalise the data. The missing data value will be set to zero. Kelly et al. (2019) (KPS) and Kozak et al. (2020) (KNS) employ two rank transformation methods for the data. The former ranks the firm characteristics in each period and shrinks the results between -0.5 to 0.5, whereas the latter concentrates the data between 0 and 1 and normalises the results by mean.

Both methods will require the rank of variables, $Rank^k_{n,t}$, for stock n, characteristic k and time period t. N_t is the total firm number of period t. Kelly et al.

(2019) normalise the rank by

$$z^k_{n,t} = \frac{Rank^k_{n,t} - 1}{N_t - 1} - 0.5. \tag{3.27}$$

Kozak et al. (2020) calculate the transformed chracterisitc in two step. The first step weights the rank

$$RW^k_{n,t} = \frac{Rank^k_{n,t}}{N_t + 1}. \tag{3.28}$$

The second step normalises the weighted rank by deviation

$$d^k_{n,t} = RW^k_{n,t} - \frac{1}{n_t}\sum_{n=1}^{N_t} RW^k_{n,t}. \tag{3.29}$$

So the rank-transformed firm characteristics denote as:

$$z^k_{n,t} = \frac{d^k_{n,t}}{\sum_{n=1}^{N_t} d^k_{n,t}}. \tag{3.30}$$

3.4.2 Characteristics results

3.4.2.1 Fama-MacBeth regression of individual excess stock returns

Before constructing PC factors, I perform FM regressions of excess returns on individual characteristics to investigate the significance of each firm characteristic.

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
abcinv	-0.11	0.00	-4.10	-0.50	0.00	-3.62	-567.80	1.90	-3.64
absoacc	-0.02	0.00	-0.21	0.16	0.00	0.59	227.44	4.19	0.61

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
age	0.08	0.00	1.18	0.51	0.00	2.22	817.61	4.32	2.11
atan	0.20	0.00	4.15	0.74	0.00	4.40	1031.95	3.27	4.35
atme	0.12	0.00	1.39	0.67	0.00	1.82	1027.81	7.07	1.72
ato	0.12	0.00	3.18	0.50	0.00	3.22	635.89	2.33	2.91
atpch	-0.32	0.00	-6.55	-1.05	0.00	-4.89	-1548.90	3.77	-4.93
baspread	0.21	0.00	1.25	0.12	0.01	0.22	176.66	9.26	0.20
beta	-0.07	0.00	-0.43	-0.21	0.01	-0.42	-245.52	7.22	-0.40
beta_capm	-0.13	0.00	-1.15	-0.36	0.00	-1.00	-568.79	4.92	-1.01
beta_down	-0.01	0.00	-0.27	0.02	0.00	0.12	27.46	2.28	0.10
betasq	0.45	0.00	3.55	0.83	0.00	2.34	1048.72	5.33	2.31
bleve	-0.01	0.00	-0.25	0.03	0.00	0.13	72.10	5.40	0.16
btm	0.29	0.00	3.35	0.88	0.00	2.89	1344.65	5.58	2.68
capxch1	-0.20	0.00	-4.64	-0.77	0.00	-5.05	-1038.80	2.30	-5.07
capxch2	-0.24	0.00	-5.30	-0.88	0.00	-5.64	-1075.60	2.35	-5.67
capxch3	-0.22	0.00	-4.63	-0.76	0.00	-4.97	-846.31	1.85	-4.98
capxpch	-0.19	0.00	-5.54	-0.93	0.00	-5.86	-1148.39	2.39	-5.92
cash	0.06	0.00	0.59	0.30	0.00	0.92	460.41	6.32	0.87
cashdebt	0.11	0.00	1.22	0.54	0.00	1.51	715.31	5.20	1.51
cashpr	0.00	0.00	-0.08	-0.43	0.00	-1.53	-699.41	5.33	-1.52
cbop	0.40	0.00	6.96	1.51	0.00	7.88	1786.28	2.87	7.09
cboplag	0.29	0.00	4.81	1.07	0.00	5.29	1123.01	2.39	4.61
cepch	-0.10	0.00	-2.81	-0.34	0.00	-1.61	-529.78	3.34	-1.71
cfp	0.26	0.00	2.89	1.30	0.00	3.64	1528.50	5.82	2.85
cfvol	-0.09	0.00	-1.63	-0.32	0.00	-1.01	-236.72	2.87	-0.95
chmom	-0.10	0.00	-1.43	-0.37	0.00	-1.59	-477.41	3.43	-1.37

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
cinvest	0.10	0.00	3.38	0.57	0.00	5.41	748.17	1.51	5.57
coach	-0.23	0.00	-4.82	-0.75	0.00	-4.65	-918.12	2.18	-4.82
colch	-0.10	0.00	-2.83	-0.27	0.00	-2.01	-339.79	1.66	-2.11
cowcch	0.01	0.00	0.25	0.12	0.00	0.64	131.93	2.82	0.55
cr	-0.05	0.00	-0.88	0.05	0.00	0.25	55.66	2.94	0.20
crpch	-0.09	0.00	-2.75	-0.05	0.00	-0.44	-76.37	1.38	-0.51
cta	-0.03	0.00	-0.36	0.18	0.00	0.60	286.87	5.61	0.61
dcv	-0.09	0.00	-2.76	-0.74	0.00	-4.84	-455.52	0.96	-5.21
dcvind	-0.11	0.00	-4.11	-0.69	0.00	-4.16	-458.46	1.19	-4.48
depr	0.05	0.00	0.84	0.52	0.00	1.79	758.13	5.11	1.63
divy	-0.04	0.00	-0.76	-0.25	0.00	-0.81	-345.80	5.49	-0.70
dolvol	-0.21	0.00	-2.49	-0.67	0.00	-2.26	-1017.43	4.00	-2.18
dolvolm	-0.26	0.00	-2.66	-0.79	0.00	-2.32	-1219.30	5.38	-2.25
dppch	-0.30	0.00	-6.18	-1.02	0.00	-6.49	-1412.65	2.32	-6.16
dtpch3	-0.02	0.00	-1.64	-0.71	0.00	-8.12	-865.77	1.41	-7.84
earnvari	0.00	0.00	0.06	0.07	0.00	0.35	68.94	2.08	0.39
earnvol	-0.06	0.00	-1.57	-0.33	0.00	-1.53	-338.63	2.00	-1.50
emppch	-0.18	0.00	-5.10	-0.66	0.00	-4.12	-987.42	2.37	-4.19
ep	0.05	0.00	0.39	0.73	0.00	1.85	1081.22	7.19	1.67
flch3	-0.25	0.00	-8.94	-0.90	0.00	-9.47	-1102.67	1.53	-9.16
fscore	0.32	0.00	5.48	1.11	0.00	5.25	1649.89	3.41	4.88
gp	0.23	0.00	5.12	0.84	0.00	5.40	1260.02	2.31	5.37
gplag	0.09	0.00	2.04	0.45	0.00	2.65	598.48	2.28	2.45
ill	0.18	0.00	1.39	0.02	0.01	0.03	22.80	7.92	0.03
invch	-0.25	0.00	-6.94	-0.83	0.00	-7.12	-1166.52	1.99	-7.17

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
invest	-0.30	0.00	-6.26	-1.06	0.00	-6.76	-1335.47	2.30	-6.76
invpch	-0.13	0.00	-7.09	-0.87	0.00	-7.60	-1222.28	1.95	-7.65
iskewcapm	-0.02	0.00	-0.42	-0.09	0.00	-0.57	-145.72	3.38	-0.52
iskewdown	-0.01	0.00	-0.37	-0.02	0.00	-0.25	-40.56	1.27	-0.36
iskewff3	0.00	0.00	0.02	-0.02	0.00	-0.14	-40.53	3.38	-0.14
iskewq	-0.03	0.00	-0.59	-0.13	0.00	-0.77	-193.90	3.31	-0.71
iss	0.04	0.00	0.58	-0.05	0.00	-0.21	-3.99	3.08	-0.02
ivolcapm	0.07	0.00	0.44	-0.04	0.01	-0.08	-76.40	9.10	-0.09
ivoldown	0.10	0.00	0.74	0.05	0.00	0.12	76.96	7.71	0.10
ivolff3	0.07	0.00	0.47	-0.03	0.01	-0.06	-60.16	9.11	-0.07
ivolq	0.07	0.00	0.46	-0.03	0.01	-0.05	-55.24	9.02	-0.06
lgr	-0.22	0.00	-6.55	-0.76	0.00	-6.40	-1115.67	1.95	-6.30
liqat	0.13	0.00	1.69	0.46	0.00	1.68	611.27	4.54	1.65
ltdpch	-0.05	0.00	-3.52	-0.55	0.00	-7.08	-802.25	1.56	-6.73
ltnoach	-0.33	0.00	-7.30	-1.16	0.00	-7.12	-1617.44	2.66	-7.14
maxret	-0.07	0.00	-0.61	-0.33	0.00	-0.73	-516.46	7.84	-0.70
me	-0.20	0.00	-1.82	-0.61	0.00	-1.60	-956.91	5.17	-1.55
mom12m	0.14	0.00	1.20	0.62	0.00	1.45	939.06	6.12	1.47
mom1m	-0.53	0.00	-4.69	-1.42	0.00	-3.93	-2219.70	5.10	-3.67
mom36m	-0.30	0.00	-3.97	-1.07	0.00	-3.44	-1365.93	4.97	-3.56
mom6m	0.06	0.00	0.52	0.28	0.00	0.68	503.52	5.20	0.77
mret60	-0.18	0.00	-2.25	-0.83	0.00	-2.21	-863.06	4.23	-2.31
nceqiss	-0.33	0.00	-4.35	-1.09	0.00	-3.88	-1439.00	4.03	-3.75
ncoach	-0.33	0.00	-6.21	-1.25	0.00	-6.77	-1494.51	2.70	-6.85
ncolch	-0.05	0.00	-1.69	-0.12	0.00	-0.91	-151.35	1.76	-0.96

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
ndtp	-0.11	0.00	-1.31	-0.42	0.00	-1.56	-688.45	5.43	-1.51
nfach	0.09	0.00	2.94	0.53	0.00	6.09	761.53	1.52	5.83
nncoach	-0.15	0.00	-2.20	-0.52	0.00	-2.14	-637.14	3.75	-1.99
noa	-0.14	0.00	-1.62	-0.50	0.00	-1.78	-625.71	4.57	-1.67
oacc	-0.13	0.00	-1.89	-0.64	0.00	-3.39	-867.58	2.94	-3.35
oaccni	-0.16	0.00	-4.56	-0.76	0.00	-4.74	-934.77	2.33	-4.52
ocfta	0.28	0.00	2.52	1.31	0.00	3.69	1483.78	5.48	2.86
oleve	0.14	0.00	2.66	0.49	0.00	2.31	684.75	3.79	1.97
onep	0.76	0.00	5.35	0.76	0.01	1.64	1143.22	7.62	1.54
op	0.26	0.00	3.03	0.93	0.00	3.63	1428.03	4.07	3.46
oplag	0.20	0.00	2.60	0.59	0.00	2.28	810.01	3.68	2.11
oscore	0.06	0.00	1.00	0.27	0.00	0.67	347.93	5.42	0.69
pchgmpchsale	0.12	0.00	3.32	0.58	0.00	5.10	781.19	1.49	4.81
pchsaleinv	-0.02	0.00	-0.96	0.28	0.00	3.46	375.26	1.22	3.22
pchsalepchxsga	0.08	0.00	2.59	-0.16	0.00	-0.62	-213.99	4.41	-0.52
pctacc	-0.07	0.00	-2.70	-0.38	0.00	-1.68	-571.82	3.57	-1.68
prctmaxprc	-0.04	0.00	-0.29	0.10	0.00	0.19	245.57	7.18	0.31
qr	-0.04	0.00	-0.69	0.14	0.00	0.56	190.39	3.85	0.57
rdat	0.13	0.00	1.17	0.56	0.01	1.32	804.96	8.14	1.19
rdctba	0.19	0.00	1.67	0.70	0.00	1.71	863.60	6.51	1.62
rdind	0.14	0.00	1.31	0.64	0.01	1.38	736.94	7.51	1.19
rdme	0.30	0.00	2.87	0.72	0.00	1.83	1020.78	7.22	1.64
rdsale	0.00	0.00	0.00	0.44	0.01	1.03	631.12	8.22	0.92
retvol	0.04	0.00	0.26	-0.08	0.01	-0.16	-132.91	8.63	-0.16
roaq	0.14	0.00	1.27	0.70	0.00	2.01	1002.87	5.73	1.87

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
roavol	-0.02	0.00	-0.23	0.17	0.00	0.43	193.60	4.67	0.47
roeq	0.10	0.00	1.43	0.70	0.00	2.18	1043.09	5.52	2.03
roic	0.04	0.00	0.40	0.55	0.00	1.63	818.52	5.86	1.51
ronoa	0.07	0.00	1.44	0.67	0.00	2.17	979.32	5.29	1.97
rsup	-0.02	0.00	-0.49	0.24	0.00	1.49	285.39	2.43	1.31
salecash	-0.02	0.00	-0.87	0.14	0.00	0.59	148.09	4.76	0.38
saleemppch	-0.08	0.00	-3.13	-0.05	0.00	-0.61	-71.96	1.34	-0.64
saleinv	0.02	0.00	0.77	0.23	0.00	1.97	360.13	2.00	1.99
salepch3	-0.12	0.00	-3.37	-0.57	0.00	-4.12	-685.41	1.65	-4.16
salerec	0.00	0.00	0.14	0.19	0.00	1.69	198.59	2.17	1.22
secured	-0.04	0.00	-0.94	-0.19	0.00	-1.49	-264.22	1.99	-1.70
securedind	-0.04	0.00	-1.22	-0.13	0.00	-1.18	-349.12	2.63	-1.54
stdacc	-0.09	0.00	-1.81	-0.32	0.00	-1.04	-233.83	2.74	-0.96
stddolvol	0.15	0.00	2.52	0.49	0.00	2.46	751.47	2.90	2.44
stdturn	-0.04	0.00	-0.54	0.40	0.00	1.17	676.27	5.54	1.16
stinvch	-0.11	0.00	-3.13	-0.25	0.00	-2.97	-295.91	1.20	-2.94
stp	0.24	0.00	2.87	0.98	0.00	3.14	1410.01	5.97	2.82
tacc	0.03	0.00	0.79	0.21	0.00	1.27	359.11	2.45	1.63
taccni	0.01	0.00	0.44	-0.01	0.00	-0.09	98.39	1.96	0.60
taxch	0.13	0.00	4.94	0.65	0.00	5.66	835.63	1.48	5.19
turn	-0.30	0.00	-3.00	-0.85	0.00	-2.38	-1285.35	5.57	-2.17
txtsup	0.08	0.00	2.96	0.44	0.00	4.02	603.92	1.52	3.72
zerotrade	0.09	0.00	1.53	0.57	0.00	1.69	542.69	3.03	1.52
zscore	-0.17	0.00	-2.25	-0.34	0.00	-1.08	-426.00	3.73	-1.09

Table 3.4: FM regression of excess returns on individual characteristics

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW

Note: This table displays the results of Fama-MacBeth regressions on the excess returns on 126 characteristics of individual stocks. The regressions are run from June 1987 to June 2017 (362 periods in total). Each characteristic is abbreviated in the first column. The mean value (Mean), standard deviation (Std), and t-statistics after Newey and West (1987) adjustment (t_NW) under Green et al. (2017) sample transformation are shown in the second to fourth columns. The fifth to seventh columns are similar, but with Kelly et al. (2019) rank transformation. The final three columns show the effects of Kozak et al. (2020) data adjustment.

Table 3.4 displays the regression results using three different data adjustment methods. The first group (column 2-4) shows the mean value, standard deviation and t-statistics for GHZ after HAC adjustment(Newey and West (1987)). The second group (column 5-7) is for KPS and the third group (column 8-10) is for KNS. With $t > 3$, GHZ counts 38 firm characteristics. KPS and KNS have 42 and 38 characteristics, respectively. If I lower the threshold to $t > 2$, GHZ, KPS and KNS have 60, 59, and 55 firm characteristics that can capture the movement of individual excess returns, respectively. When the results are compared, KNS has the worst performance in all three adjustments. The rank transformation with characteristics adjustment could be the cause. The normalised rank will be divided by the sum of deviations when I use the Kozak et al. (2020) rank transformation method. This calculation will result in small values for all variables. However, I use $(Z_t'Z_t)^{-1}$ to adjust the excess returns, which produces very large returns. The procedure causes more fluctuation than the other two adjustment methods. When I run FM regressions with single characteristics for all stocks, the result is is slightly better than Green et al. (2017), who have 30 significant characteristics.

I also run FM regressions on all-but-micro stocks, and modify the database with 20th percentile of market equity from NYSE.[13] According to Table 3.5, GHZ and KNS only have 22 characteristics with t-statistics greater than 3. While KPS have one more significant characteristic. Under $t > 2$ constraint, KNS has the most variables (44), followed by GHZ (41) and KPS (43). When I compare Table 3.4 and Table 3.5, I get the same result as Green et al. (2017). Only including All-but-micro stocks reduces the explanatory power for capturing excess stock returns for the majority of characteristics. However, a few of them have distinctive patterns that will be important to describe stock returns once the micro firms have been eliminated.[14] These variables have little impact on small firms. However, most characteristic will be affected by the firm's scale, like *size* (Banz (1981)). Fama and French (1995) interpret that small firms will have higher expected stock returns than large firms and size effect is derive by the low profits. In a nutshell, when I adjust the full data sample to All-but-micro, the characteristics related to profit and earnings show significant performance changes. As a result, Hou et al. (2015) introcude investment and profitability factors from the q-theory and they claim that the q-theory model can absorb most anomalies which cannot be explained by Fama-French three-factor model.

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
absoacc	0.02	0.00	0.23	0.20	0.00	0.99	103.76	1.46	0.77
age	0.05	0.00	0.74	0.29	0.00	1.10	248.47	2.21	1.18

[13]The breakpoints of market equity are collect from Ken. French website. `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[14]They are *cashdebt, mom12m, ptcacc, roaq and roic*. Some other characteristics, *ocfta, roeq* and *ronoa*, are already significant in predicting excess stock returns without subtraction, but they still benefit from All-but-micro sample adjustment.

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
atan	0.11	0.00	2.16	0.39	0.00	2.22	241.16	1.38	2.05
atme	0.02	0.00	0.31	0.04	0.00	0.11	3.35	3.37	0.01
ato	0.09	0.00	2.06	0.36	0.00	2.17	212.15	1.02	1.99
atpch	-0.18	0.00	-2.94	-0.50	0.00	-2.35	-357.37	1.65	-2.33
baspread	-0.12	0.00	-0.74	-0.25	0.00	-0.52	-217.53	3.64	-0.58
beta	-0.05	0.00	-0.42	-0.19	0.00	-0.50	-108.48	2.60	-0.48
beta_capm	-0.01	0.00	-0.06	0.01	0.00	0.03	-53.08	2.93	-0.17
beta_down	0.03	0.00	0.49	0.15	0.00	0.79	83.00	1.25	0.59
betasq	0.09	0.00	0.91	0.18	0.00	0.65	89.84	1.99	0.52
bleve	0.04	0.00	0.68	0.19	0.00	0.64	119.55	2.42	0.52
btm	0.02	0.00	0.23	0.02	0.00	0.06	-10.35	2.96	-0.04
capxch1	-0.15	0.00	-3.33	-0.46	0.00	-2.87	-294.79	1.12	-2.90
capxch2	-0.14	0.00	-3.15	-0.41	0.00	-2.78	-222.62	0.85	-2.64
capxch3	-0.11	0.00	-2.52	-0.32	0.00	-2.18	-153.13	0.81	-1.96
capxpch	-0.13	0.00	-2.92	-0.47	0.00	-2.99	-259.40	0.87	-2.90
cash	0.05	0.00	0.45	0.21	0.00	0.60	135.70	2.86	0.51
cashdebt	0.13	0.00	1.81	0.50	0.00	2.23	317.05	1.57	2.24
cashpr	0.00	0.00	0.13	-0.04	0.00	-0.16	-16.96	2.25	-0.09
cbop	0.36	0.00	4.81	1.26	0.00	6.98	735.18	1.47	7.28
cboplag	0.24	0.00	3.30	0.94	0.00	5.32	509.20	0.96	5.93
cepch	-0.05	0.00	-1.33	-0.08	0.00	-0.50	-71.16	1.01	-0.63
cfp	0.25	0.00	2.78	0.98	0.00	2.58	549.21	3.02	2.01
cfvol	-0.07	0.00	-1.52	-0.27	0.00	-1.12	-102.59	1.04	-1.10
chmom	-0.12	0.00	-1.52	-0.40	0.00	-1.59	-200.60	1.70	-1.15

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
cinvest	0.10	0.00	3.18	0.28	0.00	2.88	153.59	0.63	2.75
coach	-0.11	0.00	-2.12	-0.33	0.00	-1.81	-181.50	0.91	-1.83
colch	0.03	0.00	0.53	0.08	0.00	0.42	13.66	1.14	0.13
cowcch	-0.04	0.00	-0.82	-0.06	0.00	-0.30	-24.95	1.24	-0.23
cr	-0.07	0.00	-0.96	-0.08	0.00	-0.25	-46.61	1.85	-0.24
crpch	-0.07	0.00	-1.95	-0.08	0.00	-0.96	-51.05	0.48	-1.13
cta	0.02	0.00	0.19	0.17	0.00	0.48	112.80	2.84	0.44
dcv	-0.04	0.00	-1.01	-0.34	0.00	-1.99	-123.51	0.56	-2.22
dcvind	-0.06	0.00	-1.98	-0.33	0.00	-1.98	-127.88	0.59	-2.19
depr	-0.02	0.00	-0.33	0.24	0.00	0.94	153.79	1.87	0.84
divy	-0.02	0.00	-0.28	-0.02	0.00	-0.05	7.65	2.61	0.03
dolvol	-0.03	0.00	-0.57	-0.13	0.00	-0.71	-75.46	1.35	-0.55
dolvolm	-0.04	0.00	-0.74	-0.16	0.00	-0.82	-97.01	1.40	-0.68
dppch	-0.18	0.00	-3.24	-0.45	0.00	-2.16	-289.24	1.49	-2.07
dtpch3	-0.02	0.00	-0.94	-0.47	0.00	-6.06	-265.18	0.56	-6.01
earnvari	-0.04	0.00	-0.81	-0.06	0.00	-0.39	-23.72	0.79	-0.33
earnvol	-0.03	0.00	-0.79	-0.11	0.00	-0.55	-55.92	0.94	-0.58
emppch	-0.09	0.00	-2.16	-0.31	0.00	-1.58	-216.71	1.34	-1.56
ep	0.17	0.00	1.88	0.65	0.00	1.86	440.51	3.14	1.59
flch3	-0.16	0.00	-5.86	-0.54	0.00	-5.52	-306.04	0.70	-5.19
fscore	0.21	0.00	3.74	0.75	0.00	3.83	506.24	1.47	3.42
gp	0.16	0.00	3.46	0.63	0.00	3.66	426.44	1.14	3.75
gplag	0.07	0.00	1.30	0.33	0.00	1.68	195.19	1.17	1.54
ill	-0.17	0.00	-1.17	-0.41	0.00	-0.87	-329.59	3.50	-0.92

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
invch	-0.13	0.00	-3.52	-0.40	0.00	-3.40	-245.82	0.67	-3.33
invest	-0.16	0.00	-3.95	-0.51	0.00	-3.75	-282.98	0.77	-3.62
invpch	-0.12	0.00	-4.32	-0.44	0.00	-3.83	-273.74	0.66	-3.74
iskewcapm	0.00	0.00	0.06	-0.02	0.00	-0.14	-18.79	1.12	-0.18
iskewdown	-0.02	0.00	-0.74	-0.05	0.00	-0.73	-26.35	0.48	-0.51
iskewff3	0.01	0.00	0.42	0.03	0.00	0.20	8.04	1.16	0.08
iskewq	0.00	0.00	-0.02	-0.05	0.00	-0.36	-43.36	1.12	-0.43
iss	0.02	0.00	0.39	-0.08	0.00	-0.39	-46.87	1.07	-0.47
ivolcapm	-0.15	0.00	-1.06	-0.38	0.00	-0.78	-298.26	3.65	-0.81
ivoldown	-0.13	0.00	-1.09	-0.29	0.00	-0.75	-237.90	2.98	-0.79
ivolff3	-0.15	0.00	-1.05	-0.37	0.00	-0.77	-292.14	3.66	-0.80
ivolq	-0.14	0.00	-0.91	-0.32	0.01	-0.64	-262.02	3.71	-0.70
lgr	-0.17	0.00	-3.66	-0.42	0.00	-2.76	-288.33	1.16	-2.63
liqat	0.09	0.00	1.09	0.31	0.00	1.01	180.62	2.03	0.92
ltdpch	-0.03	0.00	-2.02	-0.46	0.00	-6.94	-295.83	0.52	-6.64
ltnoach	-0.22	0.00	-4.52	-0.65	0.00	-4.01	-408.28	1.23	-3.84
maxret	-0.17	0.00	-1.43	-0.51	0.00	-1.21	-380.69	3.24	-1.19
me	-0.01	0.00	-0.23	-0.01	0.00	-0.04	14.63	1.14	0.11
mom12m	0.23	0.00	1.99	0.86	0.00	2.16	562.94	2.69	2.05
mom1m	-0.10	0.00	-0.99	-0.41	0.00	-1.36	-225.03	1.84	-0.97
mom36m	-0.12	0.00	-1.78	-0.32	0.00	-1.37	-199.75	1.39	-1.46
mom6m	0.15	0.00	1.27	0.41	0.00	1.07	328.12	2.51	1.17
mret60	-0.07	0.00	-1.05	-0.22	0.00	-0.83	-98.22	1.40	-0.73
nceqiss	-0.20	0.00	-2.46	-0.64	0.00	-2.30	-377.56	1.83	-2.10

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
ncoach	-0.23	0.00	-5.01	-0.73	0.00	-5.21	-400.15	0.97	-5.11
ncolch	-0.01	0.00	-0.59	-0.01	0.00	-0.13	-11.89	0.54	-0.25
ndtp	-0.07	0.00	-0.70	-0.26	0.00	-0.86	-191.22	2.55	-0.79
nfach	0.09	0.00	2.73	0.41	0.00	4.21	256.63	0.85	3.68
nncoach	-0.11	0.00	-1.33	-0.37	0.00	-1.32	-204.83	1.70	-1.18
noa	-0.14	0.00	-1.58	-0.47	0.00	-1.65	-255.29	1.95	-1.40
oacc	-0.15	0.00	-2.78	-0.57	0.00	-3.79	-325.00	1.11	-3.36
oaccni	-0.08	0.00	-1.34	-0.55	0.00	-3.46	-287.62	0.94	-3.12
ocfta	0.36	0.00	3.68	1.24	0.00	4.68	695.03	1.96	3.84
oleve	0.09	0.00	2.32	0.35	0.00	2.19	215.45	0.95	2.06
onep	0.03	0.00	0.29	0.07	0.00	0.22	19.77	2.14	0.09
op	0.26	0.00	4.55	0.85	0.00	4.56	608.11	1.71	4.57
oplag	0.18	0.00	3.24	0.58	0.00	3.11	388.59	1.28	3.23
oscore	0.03	0.00	0.61	0.02	0.00	0.05	1.13	2.13	0.01
pchgmpchsale	0.12	0.00	3.83	0.33	0.00	3.38	192.84	0.62	3.09
pchsaleinv	-0.02	0.00	-0.75	0.27	0.00	2.72	140.91	0.80	2.00
pchsalepchxsga	0.01	0.00	0.50	-0.12	0.00	-0.50	-65.72	1.83	-0.41
pctacc	-0.08	0.00	-2.94	-0.59	0.00	-2.66	-385.29	1.57	-2.43
prctmaxprc	0.10	0.00	0.76	0.31	0.00	0.72	296.76	3.04	0.93
qr	-0.06	0.00	-0.80	-0.04	0.00	-0.11	-17.16	2.22	-0.08
rdat	0.14	0.00	1.32	0.45	0.00	1.15	304.51	3.23	1.03
rdctba	0.15	0.00	1.45	0.50	0.00	1.39	295.87	2.47	1.35
rdind	0.12	0.00	1.18	0.62	0.01	1.31	285.87	2.93	1.07
rdme	0.13	0.00	1.84	0.46	0.00	1.37	303.56	2.66	1.24

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
rdsale	0.02	0.00	0.24	0.37	0.00	0.90	244.83	3.33	0.80
retvol	-0.16	0.00	-1.14	-0.39	0.00	-0.80	-307.16	3.57	-0.84
roaq	0.17	0.00	2.05	0.65	0.00	2.69	433.70	1.85	2.58
roavol	-0.03	0.00	-0.35	-0.08	0.00	-0.27	-43.06	1.62	-0.28
roeq	0.13	0.00	2.13	0.72	0.00	2.88	474.19	1.95	2.60
roic	0.16	0.00	2.06	0.75	0.00	3.10	500.57	1.96	2.76
ronoa	0.12	0.00	3.24	0.82	0.00	3.82	543.22	1.76	3.42
rsup	0.02	0.00	0.49	0.13	0.00	0.94	54.41	0.91	0.63
salecash	0.00	0.00	-0.18	0.08	0.00	0.28	46.32	2.59	0.19
saleemppch	-0.06	0.00	-3.03	0.05	0.00	0.61	25.85	0.55	0.44
saleinv	0.05	0.00	1.92	0.15	0.00	1.24	95.44	0.83	1.19
salepch3	-0.12	0.00	-2.35	-0.16	0.00	-0.95	-80.91	0.85	-0.85
salerec	0.01	0.00	0.59	0.26	0.00	1.97	168.82	1.06	1.76
secured	-0.03	0.00	-0.97	-0.21	0.00	-1.81	-142.60	0.73	-2.20
securedind	-0.05	0.00	-1.32	-0.14	0.00	-1.32	-172.31	1.09	-1.72
stdacc	-0.08	0.00	-1.82	-0.29	0.00	-1.24	-107.92	0.97	-1.22
stddolvol	0.05	0.00	1.30	0.18	0.00	1.31	106.41	0.84	1.11
stdturn	-0.02	0.00	-0.25	0.15	0.00	0.43	84.07	2.65	0.32
stinvch	-0.01	0.00	-0.33	-0.10	0.00	-1.03	-71.02	0.76	-1.12
stp	0.07	0.00	1.08	0.43	0.00	1.22	238.85	3.23	0.86
tacc	-0.01	0.00	-0.08	0.06	0.00	0.34	71.46	1.14	0.64
taccni	0.01	0.00	0.11	0.04	0.00	0.26	59.98	0.98	0.68
taxch	0.06	0.00	1.78	0.33	0.00	2.65	171.48	0.69	2.31
turn	-0.13	0.00	-1.19	-0.29	0.00	-0.79	-217.80	2.75	-0.79

Table 3.5: Fama-MacBeth regression of excess returns on individual characteristics with All-but-Microcaps

	Green et al. (2017)			Kelly et al. (2019)			Kozak et al. (2020)		
Characteristic	Mean	Std	t_NW	Mean	Std	t_NW	Mean	Std	t_NW
txtsup	0.09	0.00	2.69	0.43	0.00	3.39	253.97	0.71	3.13
zerotrade	-0.03	0.00	-1.00	0.02	0.00	0.10	-59.45	0.37	-1.50
zscore	-0.04	0.00	-0.53	0.14	0.00	0.46	84.04	1.98	0.44

Note: This table displays the results of Fama-MacBeth regressions on the excess returns on 126 characteristics of individual stocks under All-but-micro data sample (Market capitalisation above the 20th percentile of NYSE). The regressions are run from June 1987 to June 2017 (362 periods in total). Each characteristic is abbreviated in the first column. The mean value (Mean), standard deviation (Std), and t-statistics after Newey and West (1987) adjustment (t_NW) under Green et al. (2017) sample transformation are shown in the second to fourth columns. The fifth to seventh columns are similar, but with Kelly et al. (2019) rank transformation. The final three columns show the effects of Kozak et al. (2020) data adjustment.

3.4.2.2 Characteristic-adjusted managed portfolio

This part compares the excess returns on managed portfolios after applying the proposed adjustment methods. I multiply the individual stock excess returns with different transformed firm characteristics.

Figure 3.1 displays the key summary statistics, *mean*, *5th percentile* and *95th percentile*. Figure 3.1a represents managed portfolio built with GHZ. Figure 3.1b shows the results for KPS. KNS excess portfolio returns can be found in Figure 3.1c. In comparison, GHZ has the most centralised distribution of the three methods, with the 5% and 95% returns not fluctruating across all characteristics, but the KPS and KNS rank transformations have magnified a problem in which some characteristics have a high proportion of zero values in the sample.[15] Following rank transformation, zero will have a "real" value, which will significantly alter the tail values, 5 and 95 percentiles, in comparison to GHZ winsorization. Nevertheless, the results of Fama-Macbeth regression from the previous discussion explain that the rank transformation can still provide significant predictive power even if the distribution is not centralised. Following data adjustment, the KPS method can have a more focused value for each characteristic among the three adjustment methods. KNS portfolio returns will have more extreme values than the other two methods.

3.4.3 Number of PC factor selection

One critical question is how many PC factors should be used to optimise the estimation. I use two techniques to select the number of PC factors in the first step of the three-pass method: Sharpe ratio (SR) and goodness-of-fit R^2.

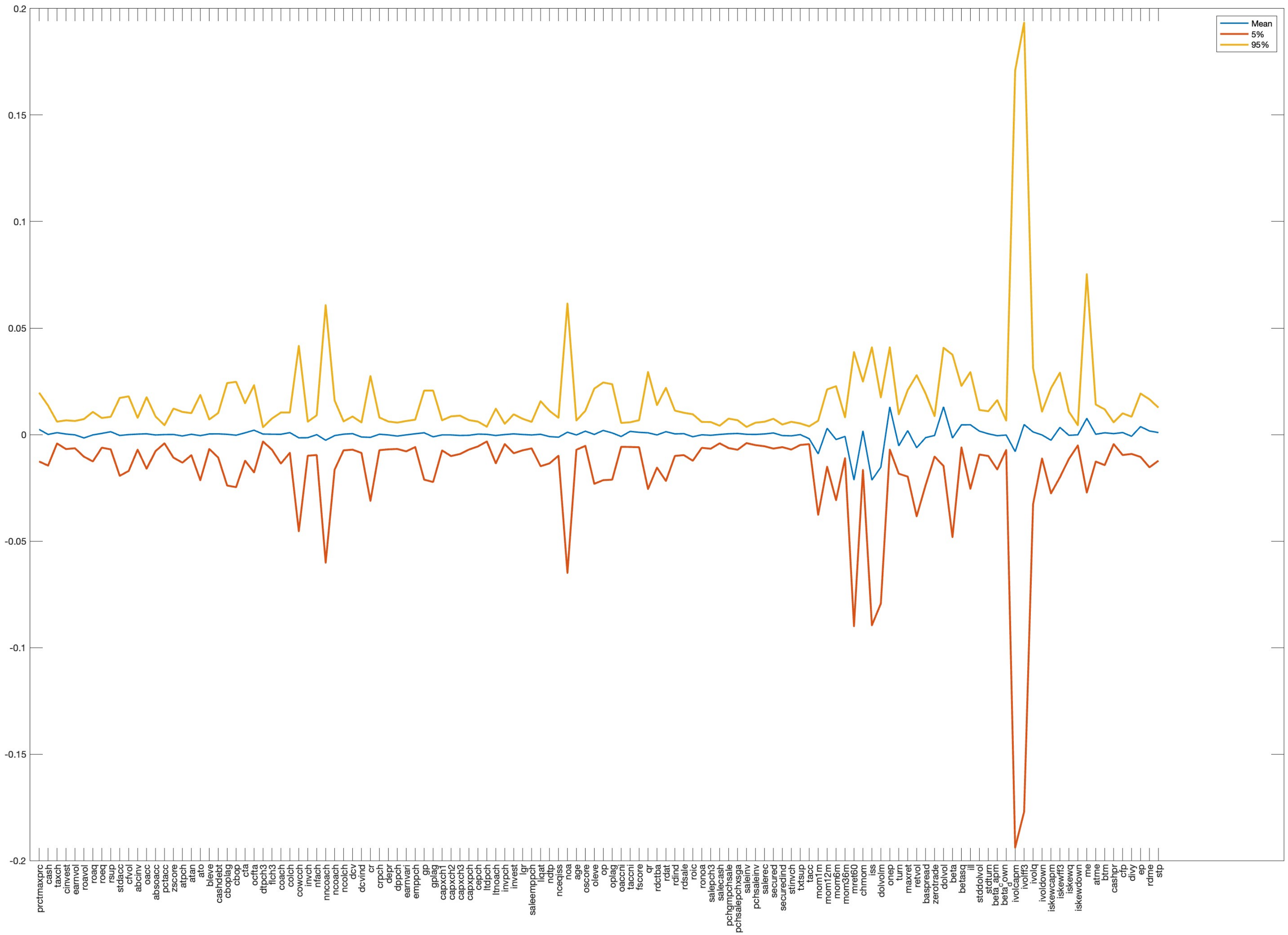
The first method, cross-validation, is applied by Lettau and Pelger (2020) and Kozak et al. (2020). The 362 months in this study were divided into three subsamples using 3-fold cross-validation. Following that, I divide three subsamples into three estimating samples and three forecasting samples. For example, if I combine the first and third subsamples to form the estimating sample, the second subsample will be transformed into the forecasting sample. Once the samples have been established, I utilise PCA or Risk-premium PCA (RP-PCA) to estimate the factor $\hat{F}$ and loadings $\hat{\Gamma_{\beta}}$. Then, for forecasting sample, I compute the out-of-sample SR (OOS-SR) . The OOS-SR is defined as follows:

$$\text{OOS-SR} = \frac{\hat{f}_{oos} w}{(w' \Sigma_{\hat{f}_{oos}} w)^{1/2}}, \tag{3.31}$$

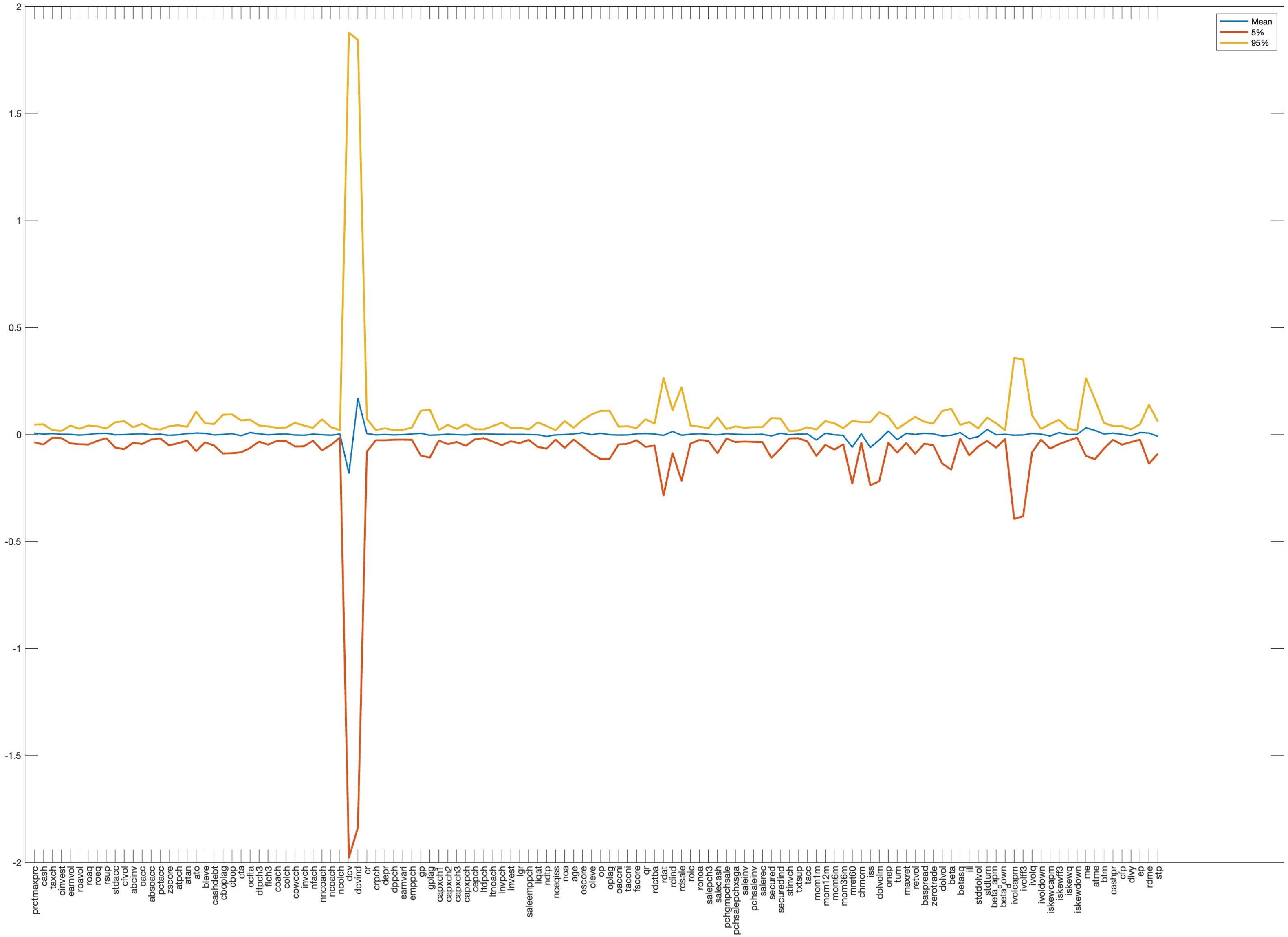
[15]Some characteristics, such as *dcvt*, convertible debt, and *dcvind*, issue indicator of convertible debt, have zero when I collect them from CRSP or Compustat. If the firm does not hold convertible debt for an extended period of time, both variables will be zero.

Figure 3.1: Summary statistics of 126 characteristic adjusted managed portfolios

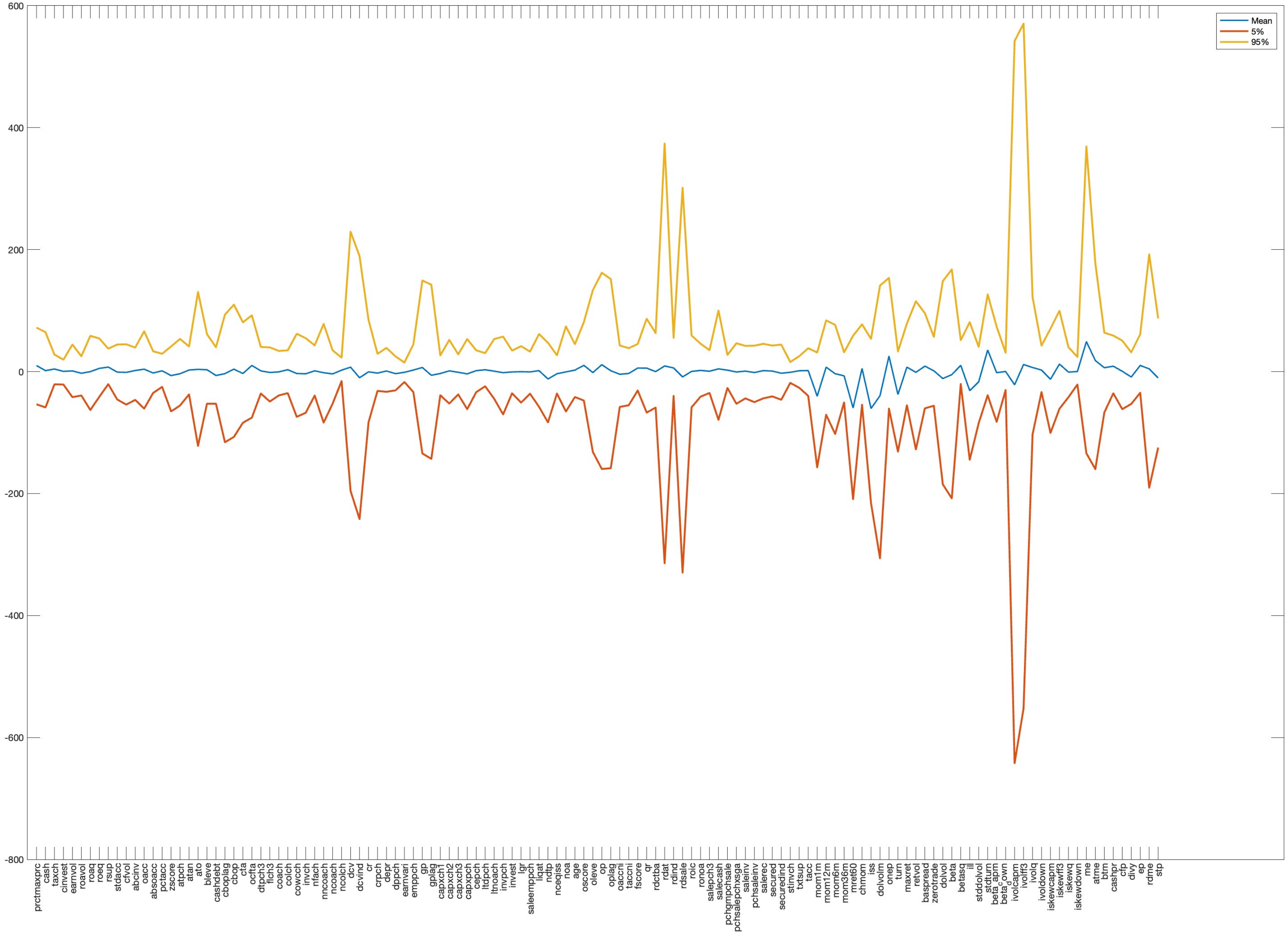
(a) Green et al. (2017)



(b) Kelly et al. (2019)



(c) Kozak et al. (2020)



Note: This figure represents the summary statistics for three different transformations of 126 characteristic-adjusted managed portfolios. The data sample spans the years June 1987 to June 2017. It includes the mean value (Mean), 5% percentile value and 95% percentile value. The first (a) is about the outcomes of using Green et al. (2017) winsorization. The second (b) is derived from Kelly et al. (2019) rank transformation. The final (c) displays summary statistics of normalised transformation by Kozak et al. (2020).

where $w = \frac{\Sigma_{\hat{F}}^{-1}\bar{F}'}{\mathbb{1}_{1\times L}\Sigma_{\hat{F}}^{-1}\bar{F}'}$ is to find the optimal weights for all factors and $\hat{f}_{oos} = R_z\hat{\Gamma_\beta}(\hat{\Gamma_\beta}'\hat{\Gamma_\beta})^{-1}$ is the OOS estimated factors.

The entire estimation procedures will be repeated three times and the global OOS-SR will be the mean of three-time results. This method is renewcommanded by Lettau and Pelger (2020) for risk-premium principal component analysis (RP-PCA) to select a better number of factors and the penalty value, γ.[16]

Figure 3.2 shows OOS-SR from 3-fold cross-validation for maximum 10 factors and γ up to 50. The details for GHZ are provided by Figure 3.2a. When γ is greater than 5, the OOS-SR is not significantly improved. Similarly, the OOS-SR remains stable when 6 or more factors are included. KPS in Figure 3.2b performs similar to the larger γ and number of factors better OOS-SR. However, KPS, which only includes 5 factors, can produce better OOS-SR (around 0.9) than GHZ (0.77). Increasing γ has no effect on the outcome, implying that γ equals 4 or 5 is sufficient to achieve a 0.88 OOS-SR. Surprisingly, under the same conditions ($\gamma = 5$, 5 factors), KNS in Figure 3.2c has the best OOS-SR under the same condition (0.93). The reason KNS outperforms the competition is the large value of firm characteristics. Overall, RP-PCA can provide better out-of-sample estimation than conventional PCA methods, and the rank transformation outperforms traditional winsorization.

The second method is to compare the R^2 for various settings. Kelly et al. (2019) employ total-R^2 and predictive-R^2 to examine the performance with different numbers of factors. The outcome will always be better if I include more factors to explain the underlying assets, but this approach contradicts our goal of maximising performance with fewer factors. Like the first method, I should strike a balance between having more factors and having a higher R^2.

[16] $\gamma = -1$ is equivalent to apply the conventional PCA to excess stock returns.

Figure 3.2: Out-of-sample Sharpe ratio of RP-PCA 3-fold Cross Validation

(a) Green et al. (2017)

Gamma \ Number of factor	1	2	3	4	5	6	7	8	9	10
-1	0.07	0.17	0.32	0.52	0.71	0.74	0.75	0.82	0.86	0.84
0	0.45	0.61	0.66	0.68	0.79	0.8	0.91	0.9	0.95	1
1	0.49	0.56	0.67	0.69	0.78	0.8	0.92	0.92	0.96	1.01
2	0.5	0.55	0.67	0.7	0.78	0.8	0.93	0.92	0.96	1.01
3	0.51	0.55	0.67	0.7	0.78	0.8	0.93	0.92	0.96	1.01
4	0.51	0.55	0.67	0.7	0.77	0.81	0.93	0.93	0.96	1.01
5	0.51	0.55	0.67	0.7	0.77	0.81	0.93	0.93	0.96	1.01
6	0.51	0.55	0.67	0.7	0.77	0.81	0.93	0.93	0.96	1.01
7	0.51	0.55	0.67	0.7	0.77	0.81	0.93	0.93	0.96	1.01
8	0.51	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
9	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
10	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
11	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
12	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
13	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
14	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
15	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
16	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
17	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
18	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
19	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
20	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
21	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
22	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
23	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
24	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
25	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
26	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
27	0.52	0.55	0.67	0.71	0.77	0.81	0.93	0.93	0.96	1.01
28	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
29	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
30	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
31	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
32	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
33	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
34	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
35	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
36	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
37	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
38	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
39	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
40	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
41	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
42	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
43	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
44	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
45	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
46	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
47	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
48	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
49	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01
50	0.52	0.55	0.68	0.71	0.77	0.81	0.93	0.93	0.96	1.01

Number of factor

(b) Kelly et al. (2019)

Gamma \ Number of factor	1	2	3	4	5	6	7	8	9	10
-1	0.15	0.21	0.48	0.48	0.57	0.58	0.6	0.59	0.7	0.72
0	0.15	0.48	0.66	0.66	0.78	0.84	0.83	0.84	0.9	0.88
1	0.16	0.68	0.72	0.73	0.85	0.87	0.87	0.88	0.92	0.92
2	0.16	0.72	0.75	0.75	0.87	0.89	0.89	0.89	0.93	0.93
3	0.16	0.74	0.76	0.77	0.88	0.89	0.89	0.9	0.93	0.93
4	0.17	0.75	0.77	0.77	0.89	0.9	0.9	0.9	0.93	0.93
5	0.17	0.75	0.77	0.78	0.89	0.9	0.9	0.9	0.93	0.93
6	0.17	0.76	0.78	0.78	0.89	0.9	0.9	0.9	0.93	0.93
7	0.17	0.76	0.78	0.78	0.9	0.9	0.9	0.91	0.93	0.93
8	0.17	0.76	0.78	0.78	0.9	0.9	0.9	0.91	0.93	0.93
9	0.17	0.77	0.78	0.79	0.9	0.9	0.9	0.91	0.93	0.93
10	0.17	0.77	0.78	0.79	0.9	0.9	0.9	0.91	0.93	0.93
11	0.17	0.77	0.79	0.79	0.9	0.9	0.9	0.91	0.93	0.93
12	0.17	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.93	0.93
13	0.17	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.93	0.93
14	0.17	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.93	0.93
15	0.17	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.93	0.93
16	0.17	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.93	0.93
17	0.18	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
18	0.18	0.77	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
19	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
20	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
21	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
22	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
23	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
24	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
25	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
26	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
27	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
28	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
29	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.93
30	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.94
31	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.94
32	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.94
33	0.18	0.78	0.79	0.79	0.9	0.91	0.91	0.91	0.94	0.94
34	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
35	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
36	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
37	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
38	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
39	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
40	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
41	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
42	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
43	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
44	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
45	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
46	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
47	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
48	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
49	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94
50	0.18	0.78	0.79	0.8	0.9	0.91	0.91	0.91	0.94	0.94

Number of factor

(c) Kozak et al. (2020)

Gamma \ Number of factor	1	2	3	4	5	6	7	8	9	10
-1	0.17	0.45	0.45	0.52	0.53	0.65	0.65	0.68	0.69	0.74
0	0.39	0.66	0.67	0.73	0.78	0.85	0.86	0.87	0.91	0.95
1	0.62	0.71	0.72	0.79	0.85	0.89	0.92	0.93	0.97	0.98
2	0.65	0.73	0.75	0.82	0.87	0.91	0.93	0.95	0.99	1
3	0.66	0.74	0.76	0.83	0.88	0.92	0.94	0.96	1	1
4	0.68	0.75	0.77	0.83	0.89	0.92	0.94	0.97	1	1.01
5	0.68	0.75	0.77	0.84	0.89	0.92	0.94	0.97	1	1.01
6	0.69	0.75	0.78	0.84	0.89	0.93	0.95	0.97	1	1.01
7	0.69	0.75	0.78	0.84	0.9	0.93	0.95	0.97	1.01	1.01
8	0.7	0.75	0.78	0.84	0.9	0.93	0.95	0.97	1.01	1.01
9	0.7	0.75	0.78	0.84	0.9	0.93	0.95	0.98	1.01	1.01
10	0.7	0.76	0.79	0.84	0.9	0.93	0.95	0.98	1.01	1.01
11	0.7	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
12	0.7	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
13	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
14	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
15	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
16	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
17	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
18	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
19	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
20	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
21	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
22	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
23	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
24	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
25	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
26	0.71	0.76	0.79	0.85	0.9	0.93	0.95	0.98	1.01	1.01
27	0.71	0.76	0.79	0.85	0.9	0.93	0.96	0.98	1.01	1.01
28	0.71	0.76	0.79	0.85	0.9	0.93	0.96	0.98	1.01	1.01
29	0.71	0.76	0.8	0.85	0.9	0.93	0.96	0.98	1.01	1.01
30	0.71	0.76	0.8	0.85	0.9	0.93	0.96	0.98	1.01	1.01
31	0.71	0.76	0.8	0.85	0.9	0.93	0.96	0.98	1.01	1.01
32	0.71	0.76	0.8	0.85	0.9	0.93	0.96	0.98	1.01	1.01
33	0.72	0.76	0.8	0.85	0.9	0.93	0.96	0.98	1.01	1.01
34	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
35	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
36	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
37	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
38	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
39	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
40	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
41	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
42	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
43	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
44	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
45	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
46	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
47	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
48	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
49	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01
50	0.72	0.76	0.8	0.85	0.91	0.93	0.96	0.98	1.01	1.01

Note: This figure displays the 3-fold cross-validation Risk-premium Principal Component Analysis (RP-PCA) by Lettau and Pelger (2020) based on 126 characteristics-adjusted managed portfolios from June 1987 to June 2017. The x-axis is the number of latent factors (ascending from left to right), and y-axis is the first-moment penalty (ascending from top to bottom). Numbers in the chart is the out-of-sample Sharpe ratio (OOS-SR). The 362 periods are seperated into three folds. Two sub-samples are used to estimate the factors and the remaining one is used to forecast.

Table 3.6: Total and predictive R^2 of managed portfolios estimation for three data transformation

Panel A: portfolio analysis under conventional PCA

Method	K1	K2	K3	K4	K5	K6
GHZ rr2	0.344	0.67	0.757	0.81	0.844	0.857
GHZ rpr2	0.001	0.002	0.022	0.022	0.024	0.024
GHZ ur2	0.371	0.696	0.765	0.819	0.85	0.863
GHZ upr2	0.028	0.028	0.028	0.028	0.028	0.028
KPS rr2	0.889	0.921	0.933	0.944	0.951	0.957
KPS rpr2	0.019	0.019	0.019	0.02	0.02	0.021
KPS ur2	0.892	0.924	0.937	0.946	0.953	0.959
KPS upr2	0.022	0.022	0.022	0.022	0.022	0.022
KNS rr2	0.34	0.474	0.55	0.6	0.647	0.683
KNS rpr2	8E-04	0.006	0.01	0.01	0.014	0.017
KNS ur2	0.369	0.498	0.571	0.621	0.664	0.699
KNS upr2	0.03	0.03	0.03	0.03	0.03	0.03

Panel B: Portfolio analaysis under RPPCA with $\gamma = 5$

Method	K1	K2	K3	K4	K5	K6
GHZ rr2	0.341	0.666	0.756	0.809	0.844	0.858
GHZ rpr2	0.002	0.003	0.024	0.024	0.025	0.028
GHZ ur2	0.368	0.694	0.763	0.816	0.849	0.859
GHZ upr2	0.028	0.028	0.028	0.028	0.028	0.028
KPS rr2	0.889	0.92	0.931	0.943	0.951	0.957
KPS rpr2	0.019	0.019	0.021	0.022	0.022	0.022
KPS ur2	0.892	0.924	0.933	0.944	0.952	0.958
KPS upr2	0.022	0.022	0.022	0.022	0.022	0.022
KNS rr2	0.338	0.455	0.539	0.599	0.649	0.688
KNS rpr2	0.001	0.016	0.022	0.026	0.026	0.027
KNS ur2	0.367	0.476	0.55	0.604	0.654	0.692
KNS upr2	0.03	0.03	0.03	0.03	0.03	0.03

Note: This table displays the regression R^2 of characteristics-adjusted managed portfolios on conventional PCA (Panel A) and RP-PCA with $\gamma = 5$ (Panel B) from one factor (K1) to six factors (K6). "GHZ", "KPS" and "KNS" represent three data transformation methods, winsorisation from Green et al. (2017), rank tranksformation from Kelly et al. (2019) and rank transformation with normalisation from Kozak et al. (2020), respectively. For each group (four rows), the first two rows ("rr2" and "rpr2") show total R^2 and predictive R^2 of the restricted model with no intercept. While the next two rows ("ur2" and "upr2") report R^2 of unrestricted model with intercept. The regression covers the time periods from June 1987 to June 2017.

Table 3.6 shows the in-sample R^2 for three characteristic-adjusted managed portfolios with restricted and unrestricted models based on Equation 3.18 and Equation 3.19. Panel A employs the conventional PCA method. Panel B is based on RP-PCA with $\gamma = 5$. From the results, I cannot find significant difference between PCA and RP-PCA under the same condition regardless of whether apply the resricted or unrestricted model. γ in RP-PCA can improve the explanatory power on predictive-R^2. However, the shift is insufficient.

Similarly, total-R^2 for unrestricted model is not noticeably higher than the restricted model. Also, whether the number of factor is 1 or 6, the predictive-R^2 is always the same, and this issue occurs in all three data adjustment methods. It is because intercept exists in the unrestricted model. The estimated factors lose explanatory power for managed portfolio returns, whereas the intercept is significant for measuring returns. KPS has the best in-sample R^2 in three adjustments, and even when only one factor is included, the R^2 remains around 89% (34% to 37% for GHZ and KNS). The total-R^2 for six PC factors in the regression is around around 96%, indicating that these factors can almost perfectly capture the characteristic-adjusted managed portfolio returns. However, KPS does not have the best predicting power of the three, with the predictive-R^2 of only 2.1% for the unrestricted model compared to GHZ 2.4% and KNS 3%. Overall, KPS demonstrates a comparable R^2 in three in-sample managed portfolio analysis.

Table 3.7 returns the total-R^2 and predictive-R^2 from Equation 3.20. The result changes when I apply the calculation to individual stock with estimated factors and loadings. Panel A describes conventional PCA estimation. GHZ and KNS do not perform well that total-R^2 is never higher than 3.5%, which is far lower than KPS's 10% to 12 %. Only GHZ's predictive-R^2 with more than three factors can outperform KPS for the restricted model. Similar to managed portfolio analysis, the intercept will boost the right-hand side of the regression's

Table 3.7: Total and predictive R^2 of individual stock estimation for three data transformation

Panel A: Individual stock under conventional PCA

Method	K1	K2	K3	K4	K5	K6
GHZ rr2	-0.0241	0.002	0.0058	0.0072	0.016	0.034
GHZ rpr2	-0.0001	2E-04	0.0029	0.0029	0.003	0.003
GHZ ur2	-0.0175	0.009	0.0109	0.0123	0.022	0.039
GHZ upr2	0.0067	0.007	0.0067	0.0067	0.007	0.007
KPS rr2	0.0962	0.098	0.1019	0.102	0.106	0.112
KPS rpr2	0.0012	0.001	0.0012	0.0009	0.001	0.003
KPS ur2	0.1012	0.103	0.1071	0.1073	0.111	0.116
KPS upr2	0.0059	0.006	0.0059	0.0059	0.006	0.006
KNS rr2	0.002	0.008	0.0113	0.0142	0.026	0.03
KNS rpr2	0.0001	0	-0.0001	-0.0001	6E-04	7E-04
KNS ur2	0.0034	0.01	0.0143	0.0171	0.029	0.031
KNS upr2	0.0023	0.002	0.0023	0.0023	0.002	0.002

Panel B: Individual stock under RPPCA $\gamma = 5$

Method	K1	K2	K3	K4	K5	K6
GHZ rr2	0.0014	0.0029	0.008	0.009	0.016	0.028
GHZ rpr2	0.0002	0.0003	0.004	0.004	0.004	0.005
GHZ ur2	0.0076	0.0092	0.011	0.013	0.021	0.031
GHZ upr2	0.0067	0.0067	0.007	0.007	0.007	0.007
KPS rr2	0.0962	0.0982	0.103	0.106	0.104	0.114
KPS rpr2	0.0012	0.0013	0.003	0.003	0.004	0.004
KPS ur2	0.1012	0.103	0.107	0.11	0.107	0.117
KPS upr2	0.0059	0.0059	0.006	0.006	0.006	0.006
KNS rr2	0.002	0.0081	0.013	0.024	0.027	0.031
KNS rpr2	0.0001	0.001	0.002	0.002	0.002	0.002
KNS ur2	0.0033	0.0108	0.015	0.026	0.028	0.032
KNS upr2	0.0023	0.0023	0.002	0.002	0.002	0.002

Note: This table displays the regression R^2 of individual stock excess returns on conventional PCA (Panel A) and RP-PCA with $\gamma = 5$ (Panel B) from one factor (K1) to six factors (K6). "GHZ", "KPS" and "KNS" represent three data transformation methods, winsorisation from Green et al. (2017), rank tranksformation from Kelly et al. (2019) and rank transformation with normalisation from Kozak et al. (2020), respectively. For each group (four rows), the first two rows ("rr2" and "rpr2") show total R^2 and predictive R^2 of the restricted model with no intercept. While the next two rows ("ur2" and "upr2") report R^2 of unrestricted model with intercept. The regression covers the time periods from June 1987 to June 2017.

fitness, with GHZ benefiting more than the other two methods. The outcome is consistent with the findings of FM regressions in 3.4.2, which show that the rank transformation of KNS has less significance than GHZ and KPS in the regression of individual stock excess returns. There's still a chance that this is due to characteristics' low numerical value.

Next, I test the out-of-sample performance following Section 3.2.4. The results is exhibited in Table 3.8. Table 3.8 Panel A exhibits the total- and predictive-R^2 for the restricted model of managed portfolios. Only KPS can have a good fitness with PCA in Block I and RP-PCA with $\gamma = 5$ in Block II. Two PC techniques can produce nearly identical results (about 95% for total and 1.4% for prediction). GHZ and KNS, on the other hand, cannot be significant with just one or two PC factors.[17] Even if I increase the number of factors to six, the total R^2 is still less than 75% .

The OOS R^2 of the regression of individual stock returns in Table 3.8 Panel B has almost the same interpretation as the OOS R^2 of managed portfolios. GHZ and KNS have no explanatory power for OOS prediction. KPS can still pass the test with around 12% total-R^2 and 0.13% predictive-R^2 for 5-factor model. RP-PCA can do even better with 0.265% predictive-R^2 in Block IV. When I examine the model with four factors derived by conventional PCA in Block III, KPS loses predicting power noticeably.

Further, I calculate the OOS-SR for three methods shown in Table 3.9. According the results, when deriving more than two PC factors, RP-PCA with $\gamma = 5$ can have a better OOS-SR than conventional PCA method. KPS OOS-SR consistently ranks in the top two of the three data adjustment techniques.

Overall, I thoroughly assess the performance of in-sample and OOS tests, and I prefer the KPS rank transformation method to modify the firm charac-

[17]The predictive-R^2 is negative when include one or two factors, which is unacceptable.

190

Table 3.8: Out-of-sample Fits

Panel A: Out-of-sample fits of estimation for three data transformation

Block I:

Row	K1	K2	K3	K4	K5	K6
Green r2	3.2755	48.512	59.159	67.522	72.805	73.989
Green pr2	-0.0602	-1.7182	0.7105	0.7097	0.9422	0.9011
Kelly r2	91.297	93.692	94.275	94.92	95.557	95.964
Kelly pr2	1.3935	1.3603	1.3684	1.3915	1.4185	1.4444
Kozak r2	30.821	43.598	50.566	54.183	58.223	62.377
Kozak pr2	-0.2963	0.0934	0.2394	0.4889	0.5038	0.8905

Block II:

Row	K1	K2	K3	K4	K5	K6
Green r2	15.132	47.2001	59.3036	67.1947	72.4614	74.5353
Green pr2	-2.1717	-4.1391	0.4607	0.6114	0.6619	0.8667
Kelly r2	91.287	93.5693	94.3567	95.0012	95.5409	96.0166
Kelly pr2	1.3951	1.1972	1.4088	1.4173	1.439	1.4941
Kozak r2	30.0759	41.1103	49.4343	53.9701	57.9445	61.9504
Kozak pr2	-1.3038	-0.3342	0.246	0.8846	0.9484	1.0495

Panel B: Out-of-sample fits of individual stock analysis

Block III:

Row	K1	K2	K3	K4	K5	K6
Green r2	-1.0032	0.0715	0.2796	0.1023	0.8584	1.5379
Green pr2	-0.0093	0.0685	0.2408	0.2319	0.1786	0.1425
Kelly r2	11.243	11.262	10.925	11.375	12	12.372
Kelly pr2	0.076	0.1009	0.0725	-0.0226	0.1312	0.189
Kozak r2	0.0207	0.6017	0.7748	1.64	1.8578	2.4124
Kozak pr2	0.023	-0.0182	0.0154	0.0385	0.0327	0.0236

Block IV:

Row	K1	K2	K3	K4	K5	K6
Green r2	-0.4757	0.0096	0.2612	0.3386	0.729	1.8321
Green pr2	-0.0929	0.1565	0.2654	0.2487	0.1275	0.0721
Kelly r2	11.242	11.2208	11.4246	11.3284	11.9704	12.6318
Kelly pr2	0.0744	0.1985	0.071	0.1551	0.2654	0.3486
Kozak r2	-0.0197	0.6236	0.7883	2.0038	2.2986	2.3485
Kozak pr2	0.0849	-0.0637	0.0096	-0.0288	0.0184	0.0586

Note: This table shows out-of-sample R^2 of regression of on conventional PCA (Block I and III) and RPPCA with $\gamma = 5$ (Block II and IV) from one factor (K1) to six factors (K6). Panel A shows the results of regression of characteristics-adjusted managed portfolio excess returns, and Panel B displays the regression of individual excess stock returns. The data sample spans from June 1987 to June 2017. "GHZ", "KPS" and "KNS" represent three data transformation methods, winsorization from Green et al. (2017), rank tranksformation from Kelly et al. (2019) and rank transformation with normalization from Kozak et al. (2020), respectively. For each group (two rows), the first row ("r2") show total R^2 and the second row displays predictive R^2.

Table 3.9: Out-of-sample Sharpe Ratio

	K1	K2	K3	K4	K5	K6
Panel A: PCA						
GHZ	0.032	0.03	0.58	0.57	0.59	0.57
KPS	0.145	0.15	0.2	0.4	0.37	0.6
KNS	-0.001	0.2	0.24	0.15	0.31	0.46
Panel B: RPPCA ($\gamma = 5$)						
GHZ 5	0.019	0.02	0.6	0.6	0.64	0.94
KPS 5	0.145	0.15	0.56	0.61	0.67	0.74
KNS 5	0.012	0.34	0.52	0.72	0.7	0.73

Note: This table shows out-of-sample Share-ratio (OOS-SR) of regression of individual excess stock returns on conventional PCA (Panel A) and RPPCA with $\gamma = 5$ (Panel B) from one factor (K1) to six factors (K6). "GHZ", "KPS" and "KNS" represent three data transformation methods, winsorization from Green et al. (2017), rank tranksformation from Kelly et al. (2019) and rank transformation with normalization from Kozak et al. (2020), respectively. The data sample spans from June 1987 to June 2017.

teristics and construct characteristic-adjusted managed portfolios. In general, I also find five PC latent factors is a better choice. If RP-PCA is used, the first moment penalty, γ should be set to 5. I did not select KNS to adjust the firm's characteristics for several reasons. First, KNS has less significant characteristics and poor in-sample and out-of-sample estimation compared to KPS or GHZ. Second, using the KNS to build managed portfolios will result in much larger outliers than the other two methods, making the managed portfolio returns distribution much larger than the other two, which will significantly reduce the usefulness of excess returns. On the other hand, GHZ's winsorization produces fewer significant firm characteristics than KPS most of the time, and its in-sample performance is also lower. Besides, the out-of-sample R^2 and OSS-SR cannot dominant KPS's results.

3.5 Empirical analysis

The empirical analysis of the 3PM is presented in this section. First, I examine the macrofactor-mimicking portfolio. The results of the three-pass method are then summarised.

3.5.1 Macro factor-mimicking portfolio analysis

contends that mimicking portfolios can aid in estimating risk premiums and investigating whether the underlying macroeconomic variable has significant explanatory power in measuring expected returns. Furthermore, the macroeconomic variables are non-tradable, we cannot calculate the risk premium through regressions directly.

Following Vassalou (2003), I construct the mimicking portfolio in two different approaches. The first one is to choose Fama-French 6 bivariate portfolios sorted by size (small and big) and book-to-market (value, neutral and growth) as the base assets. Another strategy is to use 126 firm characteristics-adjusted managed portfolios developed in this chapter as the base assets. I construct the mimicking portfolio in the following steps: First, run the simple regression of macroeconomic variable on two groups of base assets one at a time,

$$MP_m = \alpha + B\,C + e_m, \quad \text{for} \quad m = 1, 2, \cdots, M, \tag{3.32}$$

where M_m is the $T \times 1$ matrix of macro factor m, B is the $T \times N$ base assets matrix and C is the $N \times 1$ corresponding coefficient vector of mimicking portfolios on base assets. N is 6 if the base assets are bivariate sorted portfolios. If apply the characteristics-adjusted managed portfolios, N is 126.

Table 3.10: **Summary statistics of macroeconomic factor mimicking portfolios**

		FF sorted portfolios			126 firm characteristics		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Seq.	Macro varaibles	*risk premium*	*std*	*t-stats*	*risk premium*	*std*	*t-stats*
1	RPI	0.011	0.007	1.501	0.068	0.033	**2.059**
2	W875RX1	0.004	0.007	0.496	-0.040	0.043	-0.918
3	DPCERA3M086SBEA	0.019	0.006	**2.931**	0.126	0.021	**5.932**
4	CMRMTSPLx	0.024	0.010	**2.460**	0.122	0.024	**5.005**
5	RETAILx	0.016	0.008	1.913	0.004	0.027	0.146
6	INDPRO	-0.006	0.005	-1.207	0.037	0.034	1.084
7	IPFPNSS	-0.012	0.005	**-2.588**	0.007	0.034	0.190
8	IPFINAL	-0.014	0.004	**-3.362**	0.010	0.030	0.320
9	IPCONGD	-0.025	0.005	**-5.337**	0.023	0.024	0.950
10	IPDCONGD	0.003	0.006	0.500	-0.050	0.027	-1.876
11	IPNCONGD	-0.045	0.009	**-5.200**	0.093	0.028	**3.347**
12	IPBUSEQ	0.010	0.005	1.777	-0.004	0.039	-0.112
13	IPMAT	0.002	0.004	0.445	0.061	0.031	1.991
14	IPDMAT	0.005	0.007	0.749	-0.045	0.040	-1.120
15	IPNMAT	0.014	0.006	**2.448**	0.015	0.035	0.429
16	IPMANSICS	0.000	0.005	-0.025	-0.028	0.036	-0.784
17	IPB51222S	-0.007	0.008	-0.873	0.218	0.030	**7.391**
18	IPFUELS	-0.026	0.008	**-3.361**	-0.002	0.020	-0.098
19	CUMFNS	-0.009	0.004	**-2.145**	-0.124	0.033	**-3.764**
20	HWI	0.005	0.008	0.619	-0.319	0.034	**-9.308**
21	HWIURATIO	0.019	0.006	**3.217**	-0.211	0.032	**-6.527**
22	CLF16OV	-0.028	0.006	**-4.512**	0.018	0.024	0.731
23	CE16OV	-0.013	0.004	**-3.020**	0.001	0.031	0.037
24	UNRATE	-0.018	0.006	**-2.958**	0.005	0.037	0.130

Table 3.10: Summary statistics of macroeconomic factor mimicking portfolios

		FF sorted portfolios			126 firm characteristics		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Seq.	Macro varaibles	γ	*std*	*t-stats*	γ	*std*	*t-stats*
25	UEMPMEAN	0.006	0.008	0.683	0.085	0.030	**2.865**
26	UEMPLT5	-0.012	0.006	-1.952	-0.160	0.023	**-6.874**
27	UEMP5TO14	-0.033	0.007	**-5.089**	0.003	0.023	0.138
28	UEMP15OV	0.005	0.005	1.210	0.087	0.043	**2.008**
29	UEMP15T26	-0.004	0.002	-1.825	0.161	0.028	**5.811**
30	UEMP27OV	0.009	0.005	1.700	0.012	0.040	0.292
31	CLAIMSx	-0.015	0.005	**-2.857**	-0.169	0.026	**-6.567**
32	PAYEMS	0.001	0.006	0.143	0.017	0.044	0.379
33	USGOOD	-0.004	0.006	-0.734	-0.120	0.053	**-2.278**
34	CES1021000001	-0.016	0.003	**-4.795**	-0.206	0.041	**-5.068**
35	USCONS	-0.016	0.005	**-3.242**	-0.021	0.041	-0.512
36	MANEMP	0.008	0.006	1.401	-0.130	0.050	**-2.603**
37	DMANEMP	0.011	0.005	**2.200**	-0.132	0.046	**-2.854**
38	NDMANEMP	-0.006	0.007	-0.845	-0.070	0.050	-1.381
39	SRVPRD	0.004	0.007	0.629	0.096	0.043	**2.242**
40	USTPU	0.006	0.004	1.429	0.006	0.053	0.112
41	USWTRADE	0.001	0.008	0.091	-0.077	0.065	-1.184
42	USTRADE	0.004	0.004	1.023	0.065	0.039	1.688
43	USFIRE	0.037	0.009	**4.029**	0.147	0.061	**2.413**
44	USGOVT	-0.012	0.008	-1.533	0.140	0.028	**5.024**
45	CES0600000007	-0.010	0.007	-1.384	-0.225	0.063	**-3.594**
46	AWOTMAN	-0.001	0.005	-0.220	0.027	0.030	0.901
47	AWHMAN	-0.005	0.007	-0.725	-0.255	0.061	**-4.160**
48	HOUST	0.008	0.007	1.151	0.435	0.073	**5.946**

Table 3.10: Summary statistics of macroeconomic factor mimicking portfolios

		FF sorted portfolios			126 firm characteristics		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Seq.	Macro varaibles	γ	*std*	*t-stats*	γ	*std*	*t-stats*
49	HOUSTNE	0.001	0.007	0.135	0.327	0.052	**6.297**
50	HOUSTMW	0.030	0.010	**3.092**	0.597	0.076	**7.830**
51	HOUSTS	0.003	0.006	0.521	0.232	0.069	**3.387**
52	HOUSTW	0.000	0.006	-0.081	0.451	0.069	**6.517**
53	PERMIT	0.007	0.007	0.966	0.348	0.072	**4.820**
54	PERMITNE	0.004	0.005	0.761	0.367	0.054	**6.749**
55	PERMITMW	0.016	0.009	1.763	0.653	0.079	**8.217**
56	PERMITS	0.000	0.006	0.074	0.061	0.060	1.024
57	PERMITW	0.009	0.006	1.449	0.390	0.069	**5.618**
58	ACOGNO	0.018	0.007	**2.599**	-0.005	0.042	-0.114
59	AMDMNOx	0.009	0.005	1.800	-0.128	0.032	**-4.040**
60	ANDENOx	-0.012	0.006	-1.916	-0.135	0.024	**-5.589**
61	AMDMUOx	-0.003	0.006	-0.544	-0.376	0.051	**-7.333**
62	BUSINVx	-0.002	0.009	-0.284	-0.128	0.065	-1.956
63	ISRATIOx	-0.025	0.010	**-2.454**	-0.031	0.027	-1.130
64	M1SL	-0.006	0.013	-0.500	0.020	0.034	0.578
65	M2SL	0.026	0.010	**2.552**	-0.201	0.032	**-6.350**
66	M2REAL	-0.013	0.007	-1.819	-0.095	0.048	-1.980
67	BOGMBASE	0.033	0.015	**2.130**	-0.131	0.037	**-3.563**
68	TOTRESNS	0.011	0.012	0.891	-0.037	0.040	-0.914
69	NONBORRES	0.002	0.009	0.261	-0.096	0.032	**-2.996**
70	BUSLOANS	0.028	0.006	**4.484**	-0.035	0.029	-1.178
71	REALLN	0.005	0.012	0.390	-0.179	0.037	**-4.832**
72	NONREVSL	0.054	0.008	**6.978**	-0.176	0.028	**-6.185**

Table 3.10: **Summary statistics of macroeconomic factor mimicking portfolios**

		FF sorted portfolios			126 firm characteristics		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Seq.	Macro varaibles	γ	*std*	*t-stats*	γ	*std*	*t-stats*
73	CONSPI	0.014	0.008	1.762	0.003	0.039	0.080
74	S&P 500	0.126	0.039	**3.187**	-0.008	0.059	-0.137
75	S&P: indust	0.120	0.038	**3.202**	-0.054	0.058	-0.930
76	S&P div yield	-0.102	0.028	**-3.634**	-0.064	0.044	-1.458
77	S&P PE ratio	0.107	0.029	**3.725**	0.160	0.061	**2.633**
78	FEDFUNDS	0.005	0.002	**2.158**	-0.094	0.031	**-2.999**
79	CP3Mx	-0.006	0.005	-1.343	-0.086	0.030	**-2.858**
80	TB3MS	0.007	0.003	**2.711**	-0.073	0.027	**-2.678**
81	TB6MS	-0.004	0.001	**-3.269**	-0.037	0.027	-1.395
82	GS1	-0.004	0.002	-1.625	-0.023	0.027	-0.852
83	GS5	-0.020	0.006	**-3.176**	0.029	0.032	0.910
84	GS10	-0.024	0.006	**-4.074**	0.081	0.034	**2.371**
85	AAA	-0.030	0.006	**-5.057**	-0.002	0.032	-0.071
86	BAA	-0.043	0.012	**-3.507**	0.078	0.040	1.968
87	COMPAPFFx	-0.018	0.009	**-2.133**	-0.017	0.039	-0.433
88	TB3SMFFM	0.001	0.006	0.207	-0.190	0.032	**-5.948**
89	TB6SMFFM	-0.003	0.005	-0.574	-0.165	0.031	**-5.387**
90	T1YFFM	0.007	0.006	1.189	0.043	0.029	1.443
91	T5YFFM	0.008	0.011	0.656	0.107	0.047	**2.292**
92	T10YFFM	0.001	0.011	0.052	-0.007	0.052	-0.139
93	AAAFFM	-0.005	0.011	-0.505	-0.052	0.053	-0.994
94	BAAFFM	-0.007	0.011	-0.598	-0.048	0.059	-0.809
95	TWEXAFEGSMTHx	-0.017	0.010	-1.615	0.211	0.043	**4.926**
96	EXSZUSx	0.002	0.007	0.330	0.260	0.039	**6.719**

Table 3.10: Summary statistics of macroeconomic factor mimicking portfolios

		FF sorted portfolios			126 firm characteristics		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Seq.	Macro varaibles	γ	*std*	*t-stats*	γ	*std*	*t-stats*
97	EXJPUSx	0.025	0.014	1.872	0.270	0.048	**5.673**
98	EXUSUKx	-0.001	0.008	-0.125	-0.151	0.044	**-3.403**
99	EXCAUSx	-0.046	0.021	**-2.149**	0.046	0.058	0.801
100	WPSFD49207	-0.032	0.007	**-4.621**	-0.065	0.028	**-2.296**
101	WPSFD49502	-0.028	0.007	**-4.107**	-0.043	0.029	-1.491
102	WPSID61	-0.007	0.004	-1.915	-0.131	0.031	**-4.274**
103	WPSID62	0.000	0.002	0.223	-0.061	0.030	**-2.026**
104	OILPRICEx	-0.024	0.005	**-5.095**	0.216	0.036	**5.997**
105	PPICMM	-0.016	0.010	-1.542	0.123	0.030	**4.136**
106	CPIAUCSL	0.007	0.008	0.839	-0.075	0.028	**-2.692**
107	CPIAPPSL	0.026	0.007	**3.579**	-0.027	0.027	-0.977
108	CPITRNSL	0.003	0.008	0.357	-0.120	0.033	**-3.678**
109	CPIMEDSL	-0.005	0.003	**-2.059**	-0.039	0.015	**-2.571**
110	CUSR0000SAC	0.005	0.009	0.535	-0.115	0.032	**-3.597**
111	CUSR0000SAD	-0.002	0.008	-0.224	0.107	0.016	**6.817**
112	CUSR0000SAS	0.019	0.006	**3.285**	0.027	0.015	1.818
113	CPIULFSL	0.011	0.008	1.373	-0.077	0.028	**-2.815**
114	CUSR0000SA0L2	0.005	0.008	0.595	-0.089	0.030	**-2.972**
115	CUSR0000SA0L5	0.006	0.008	0.725	-0.075	0.026	**-2.833**
116	PCEPI	0.007	0.008	0.918	0.092	0.030	**3.045**
117	DDURRG3M086SBEA	-0.026	0.006	**-4.560**	0.051	0.024	**2.154**
118	DNDGRG3M086SBEA	0.006	0.008	0.815	-0.145	0.034	**-4.311**
119	DSERRG3M086SBEA	0.016	0.011	1.446	0.340	0.025	**13.351**
120	CES0600000008	0.005	0.005	1.046	-0.098	0.017	**-5.864**

Table 3.10: Summary statistics of macroeconomic factor mimicking portfolios

		FF sorted portfolios			126 firm characteristics		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Seq.	Macro varaibles	γ	*std*	*t-stats*	γ	*std*	*t-stats*
121	CES2000000008	0.010	0.005	1.973	-0.006	0.018	-0.338
122	CES3000000008	0.004	0.003	1.068	-0.062	0.015	**-4.027**
123	UMCSENTx	0.036	0.021	1.724	0.164	0.043	**3.796**
124	DTCOLNVHFNM	0.005	0.010	0.531	0.008	0.030	0.249
125	DTCTHFNM	0.038	0.005	**7.782**	-0.122	0.025	**-4.962**
126	INVEST	0.001	0.005	0.303	-0.110	0.025	**-4.464**
127	VXOCLSx	-0.047	0.023	**-2.012**	0.562	0.104	**5.409**
Number of variables ($t > 2$)				**48**			**74**
Number of variables Both cases with($t > 2$)							**23**

Note: This table presents the summary statistics results for 127 macroeconomic factor mimicking portfolios from June 1987 to June 2017. The first column displays the individual macroeconomic variable's sequence number. The FRED-MD abbreviated name of the macroeconomic variables is shown in the second column. The third to fifth columns show the results for the mimicking portfolios estimated by Fama-French portfolios sorted by size and book-to-market. Sixth to eighth columns present the mimicking portfolio computed from 126 managed portfolios, adjusted by Kelly et al. (2019) rank transformation. γ is the risk premium of the macroeconomic factor mimicking portfolio. *t-stats* is the t-statistics corrected for the heteroskedasticity and serial correlation with six lags by the Newey and West (1987) estimator. **Bold** number of *t-stats* indicates that the mimicking portfolios are significant at least 5% significant level. The table's final row counts the number of significant variables with ($t > 2$).

The macro factor-mimicking portfolio is then equal to the base assets multiplied by the corresponding coefficients. The summary statistics of macro factor-mimicking portfolios are displayed in Table 3.10. Column (3) to (5) indicate that 48 out of 127 macroeconomic variables can construct the significant size and B/M sorted portfolio-based mimicking portfolios (FFMP). Summary statistics of characteristics-based factor mimicking portfolios (CBMP) reports in column (6) to (8). 74 of CBMP are significant. Comparing the results, the latter method produces more usable and significant factor-mimicking portfolios. Remarkably, only *HOUSTMW* can have significant FFMP, while the rest of the housing-related macroeconomic FFMP cannot measure the stock returns because they lack a strong correlation with equity base assets. Except for *PERMITS*, all housing-related variables can have a significant correlation with stock returns by CBMP. On the contrary, IP-related macroeconomic variables perform so differently that FFMP can derive six macroeconomic variables with t-values greater than 2, whereas CBMP can only have two. The Findings suggest that IP related information does not have a strong correlation with characteristic-adjusted managed portfolio excess returns. It could be because the managed portfolios include a large amount of firm information in various aspects. The information of over-diversified firms cannot be described by IP-related macroeconomic variables. They may have a stronger correlation with FFMP than CBMP. The final row of Table 3.10 shows 23 mimicking portfolios, which is significant in both FFMP and CBMP.[18]

The result of mimicking portfolio analysis is apparently different from FM analysis. Mimicking portfolio analysis can provide more significant macroeconomic variables to measure the excess stock returns. One reason is that mimicking portfolios are linearly correlated with base assets, which means that their returns contain more stock information or less noise than the underlying macroe-

[18]The list of the significant mimicking portfolios is: *DPCERA3M086SBEA, CMRMTSPLx, IPNCONGD, CUMFNS, HWIURATIO, CLAIMSx, CES1021000001, DMANEMP, USFIRE, HOUSTMW, M2SL, BOGMBASE, NONREVSL, S&P PE ratio, FEDFUNDS, TB3MS, GS10, WPSFD49207, OILPRICEx, CPIMEDSL, DDURRG3M086SBEA, DTCTHFNM, VXOCLSx.*

conomic variable. Significant macro factor-mimicking portfolios, on the other hand, cannot be simply classified into specific areas such as FM analysis. FM analysis, for example, can derive three significant macroeconomic variables related to IP and five related to housing, whereas mimicking portfolio analysis can only derive one significant factor related to IP and one housing-based factor.

I also compare the correlations for macroeconomic variables and mimicking portfolios. The majority of macroeconomic variables have a low level of correlation, and only a few variables can have a significant correlation with others. However, for the majority of mimicking portfolios, both FFMP and CBMP have a high correlation relationship. FFMP has a stronger correlation relationship between the various mimicking portfolios than CBMP, which has a smaller but stable correlation. This means that the sorted portfolios, which serve as the base assets, can add to the explanatory power rather than cancel it out.

3.5.2 Three-pass method

In this section, I will demonstrate the analysis of the results for the three-pass method (3PM) and also extend the test from individual macroeconomic variables to macro factors derived from two PCA methods.

3.5.2.1 Choosing the number of PC factors

The important question before running the regression is determining the number of latent factors in the first step of the 3PM. I use a number of criteria, including out-of-sample Sharpe ratio (OOS-SR) and Onatski (2010) criteria, as well as total-R^2 and predictive-R^2 from Kelly et al. (2019). According to the comparison, five latent factors are the best choice for the first step of the 3PM for KPS's characteristics-adjusted managed portfolios. The optimal first-moment penalty of the object function is $\gamma = 5$ when the RP-PCA method is used to replace the conventional PCA method in the first step of the 3PM

3.5.2.2 Estimation of risk premium

I first derive five latent factors by PCA or RP-PCA (with $\gamma = 5$) from 126 characteristics-adjusted managed portfolios using the methodology described in Section 3.2.3. The risk premiums for 127 transformed macroeconomic variables are then estimated using PC factors from the first step of the 3PM.

Table 3.11: Summary statistics of macroeconomic variables' risk premium

	A: PCA			B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Macro varaibles	γ_g	$R_g^2(\%)$	*wald(p-value)*	γ_g	$R_g^2(\%)$	*wald(p-value)*
RPI	0.098	**5.655**	**0.030**	0.049	2.429	0.200
W875RX1	0.077	3.532	0.258	0.027	1.647	0.534
DPCERA3M086SBEA	0.040	4.833	**0.015**	0.023	4.041	0.063
CMRMTSPLx	0.013	2.275	0.094	0.012	2.573	0.348
RETAILx	0.015	**6.701**	**0.010**	0.024	**7.576**	**0.003**
INDPRO	-0.025	1.195	0.735	-0.021	1.300	0.825
IPFPNSS	-0.022	0.972	0.726	-0.022	1.113	0.813
IPFINAL	-0.018	1.037	0.654	-0.023	1.140	0.688
IPCONGD	-0.018	1.085	0.530	-0.020	1.023	0.596
IPDCONGD	-0.035	2.351	0.355	-0.016	1.683	0.566
IPNCONGD	0.009	2.113	0.059	-0.017	1.907	0.066
IPBUSEQ	-0.012	0.439	0.783	-0.018	0.937	0.838
IPMAT	-0.024	1.154	0.662	-0.018	1.111	0.776
IPDMAT	-0.030	2.936	0.541	-0.016	2.272	0.894
IPNMAT	-0.031	1.676	0.272	-0.015	1.510	0.405

Summary statistics of macroeconomic variables' risk premium

	A: PCA			B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Macro varaibles	γ_g	$R_g^2(\%)$	*wald(p-value)*	γ_g	$R_g^2(\%)$	*wald(p-value)*
IPMANSICS	-0.029	1.941	0.583	-0.019	1.824	0.838
IPB51222S	0.024	1.288	0.316	-0.004	1.115	0.321
IPFUELS	-0.016	1.397	0.550	-0.006	1.650	0.599
CUMFNS	-0.033	3.114	0.066	-0.021	2.824	0.213
HWI	0.038	2.444	0.143	0.032	2.068	0.222
HWIURATIO	0.016	1.162	0.331	0.013	1.111	0.491
CLF16OV	0.008	3.030	**0.044**	0.010	3.004	**0.024**
CE16OV	-0.009	1.889	**0.038**	-0.002	1.907	**0.044**
UNRATE	0.025	1.001	0.572	0.015	0.860	0.793
UEMPMEAN	0.013	0.370	0.137	0.010	0.431	0.664
UEMPLT5	0.022	1.706	0.054	0.021	1.442	0.350
UEMP5TO14	-0.017	1.414	0.449	-0.012	1.172	0.664
UEMP15OV	0.021	1.071	0.255	0.006	1.059	0.327
UEMP15T26	0.012	0.472	0.464	-0.004	0.520	0.237
UEMP27OV	0.019	0.694	0.486	0.014	0.785	0.733
CLAIMSx	0.004	3.987	**0.004**	-0.015	3.520	**0.006**
PAYEMS	-0.009	0.425	0.869	-0.013	0.800	0.922
USGOOD	-0.019	0.955	0.551	-0.020	1.503	0.840
CES1021000001	-0.016	0.904	0.528	-0.024	1.212	0.707
USCONS	-0.006	0.303	0.889	-0.004	0.393	0.935
MANEMP	-0.021	1.416	0.395	-0.025	2.023	0.796
DMANEMP	-0.019	1.364	0.710	-0.025	2.061	0.849
NDMANEMP	-0.021	1.778	0.276	-0.016	1.981	0.469
SRVPRD	-0.004	0.188	0.977	-0.010	0.426	0.963

Summary statistics of macroeconomic variables' risk premium

	A: PCA			B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Macro varaibles	γ_g	$R_g^2(\%)$	*wald(p-value)*	γ_g	$R_g^2(\%)$	*wald(p-value)*
USTPU	-0.016	0.626	0.744	-0.024	1.508	0.723
USWTRADE	-0.010	0.444	0.638	-0.021	1.443	0.829
USTRADE	-0.014	0.393	0.850	-0.014	0.492	0.961
USFIRE	-0.016	0.744	0.645	-0.003	0.710	0.729
USGOVT	-0.006	1.021	0.746	-0.010	1.122	0.760
CES0600000007	-0.019	0.735	0.785	-0.030	1.280	0.866
AWOTMAN	-0.029	2.671	**0.036**	-0.002	1.999	**0.030**
AWHMAN	-0.024	0.918	0.603	-0.032	1.775	0.788
HOUST	-0.011	0.241	0.980	-0.008	0.281	0.969
HOUSTNE	-0.017	1.314	0.649	-0.021	1.260	0.713
HOUSTMW	0.010	0.175	0.913	0.009	0.349	0.978
HOUSTS	-0.020	0.342	0.760	-0.011	0.319	0.802
HOUSTW	-0.008	0.242	0.972	-0.009	0.268	0.980
PERMIT	-0.014	0.171	0.955	-0.007	0.156	0.963
PERMITNE	-0.007	0.805	0.804	-0.013	0.725	0.819
PERMITMW	0.008	0.349	0.861	0.012	0.537	0.839
PERMITS	-0.020	0.371	0.872	-0.010	0.271	0.931
PERMITW	-0.013	0.207	0.906	-0.008	0.190	0.927
ACOGNO	-0.018	2.695	0.734	0.001	2.714	0.720
AMDMNOx	0.003	0.102	0.991	-0.015	0.269	0.893
ANDENOx	0.008	1.143	0.660	-0.016	0.831	0.747
AMDMUOx	0.001	0.510	0.897	-0.022	0.912	0.807
BUSINVx	-0.039	2.774	**0.005**	-0.040	3.270	0.082
ISRATIOx	-0.013	3.626	**0.017**	-0.021	3.927	0.056

Summary statistics of macroeconomic variables' risk premium

	A: PCA			B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Macro varaibles	γ_g	$R_g^2(\%)$	*wald(p-value)*	γ_g	$R_g^2(\%)$	*wald(p-value)*
M1SL	-0.020	0.423	0.829	-0.012	0.323	0.818
M2SL	0.019	1.056	0.626	0.005	0.941	0.611
M2REAL	0.016	1.976	0.883	-0.004	1.711	0.874
BOGMBASE	-0.067	2.703	**0.019**	-0.040	1.546	0.128
TOTRESNS	-0.067	2.520	0.248	-0.028	1.036	0.846
NONBORRES	-0.040	1.832	0.161	-0.027	1.482	0.524
BUSLOANS	-0.030	1.281	0.638	-0.012	0.848	0.612
REALLN	-0.011	1.440	0.251	-0.020	1.523	0.566
NONREVSL	-0.015	0.638	0.709	0.003	0.496	0.723
CONSPI	-0.048	2.476	0.205	-0.014	1.464	0.513
S&P 500	0.125	**48.485**	**0.000**	0.154	**48.184**	**0.000**
S&P: indust	0.127	**48.025**	**0.000**	0.151	**47.505**	**0.000**
S&P div yield	-0.088	**42.730**	**0.000**	-0.108	**42.266**	**0.000**
S&P PE ratio	0.110	**22.399**	**0.000**	0.134	**21.362**	**0.000**
FEDFUNDS	-0.028	3.311	0.050	-0.022	3.753	0.440
CP3Mx	-0.021	3.641	**0.003**	-0.037	**5.545**	**0.034**
TB3MS	-0.016	1.638	0.335	-0.009	2.166	0.489
TB6MS	-0.018	1.155	0.603	-0.018	2.298	0.599
GS1	-0.009	0.708	0.618	-0.011	1.629	0.744
GS5	0.001	2.372	0.152	0.009	2.842	0.080
GS10	-0.013	3.136	0.067	0.008	3.285	0.053
AAA	-0.047	2.710	0.563	-0.030	2.334	0.615
BAA	-0.049	**5.646**	0.169	-0.048	**5.369**	0.201
COMPAPFFx	0.046	**11.357**	**0.000**	0.007	**8.885**	**0.003**

Summary statistics of macroeconomic variables' risk premium

	A: PCA			B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Macro varaibles	γ_g	$R_g^2(\%)$	*wald(p-value)*	γ_g	$R_g^2(\%)$	*wald(p-value)*
TB3SMFFM	0.005	1.507	0.372	0.014	1.499	0.310
TB6SMFFM	0.008	0.454	0.747	0.010	0.454	0.799
T1YFFM	0.013	0.417	0.775	0.010	0.400	0.941
T5YFFM	0.011	0.488	0.754	0.020	0.681	0.797
T10YFFM	0.008	0.703	0.749	0.021	0.780	0.703
AAAFFM	0.002	0.684	0.782	0.019	0.766	0.730
BAAFFM	0.007	0.550	0.683	0.021	0.865	0.756
TWEXAFEGSMTHx	-0.034	4.168	0.169	-0.035	4.289	0.204
EXSZUSx	-0.021	2.739	0.130	-0.001	2.437	0.253
EXJPUSx	-0.018	1.771	0.435	0.002	1.946	0.386
EXUSUKx	0.003	0.907	0.412	-0.001	1.200	0.769
EXCAUSx	-0.028	**16.135**	**0.000**	-0.078	**16.379**	**0.000**
WPSFD49207	0.022	1.233	0.503	0.025	1.173	0.605
WPSFD49502	0.025	1.472	0.274	0.026	1.408	0.428
WPSID61	0.019	1.284	0.344	0.026	1.191	0.580
WPSID62	0.008	1.144	0.179	0.030	1.270	0.108
OILPRICEx	0.032	1.174	0.176	0.038	1.179	0.239
PPICMM	0.054	2.397	**0.005**	0.044	1.705	0.206
CPIAUCSL	0.036	3.179	0.060	0.038	3.063	**0.041**
CPIAPPSL	0.004	0.407	0.583	0.020	0.540	0.173
CPITRNSL	0.062	3.702	**0.035**	0.042	3.130	0.067
CPIMEDSL	-0.002	1.887	0.061	0.006	1.315	0.169
CUSR0000SAC	0.056	3.229	0.112	0.040	2.837	0.128
CUSR0000SAD	0.003	0.319	0.947	0.001	0.387	0.904

Summary statistics of macroeconomic variables' risk premium

	A: PCA			B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Macro varaibles	γ_g	$R_g^2(\%)$	*wald(p-value)*	γ_g	$R_g^2(\%)$	*wald(p-value)*
CUSR0000SAS	-0.011	**5.466**	**0.002**	0.016	**5.309**	**0.003**
CPIULFSL	0.041	3.766	**0.040**	0.041	3.545	**0.028**
CUSR0000SA0L2	0.049	3.619	**0.046**	0.042	3.317	**0.037**
CUSR0000SA0L5	0.039	3.194	**0.043**	0.038	3.055	**0.033**
PCEPI	0.054	3.450	**0.027**	0.049	3.150	**0.024**
DDURRG3M086SBEA	0.007	1.565	0.147	0.027	1.860	0.123
DNDGRG3M086SBEA	0.059	3.524	0.073	0.038	3.126	0.081
DSERRG3M086SBEA	0.020	2.359	**0.037**	0.032	2.316	**0.027**
CES0600000008	0.003	0.402	0.807	0.003	0.589	0.553
CES2000000008	-0.005	0.522	0.078	-0.004	0.382	0.530
CES3000000008	0.009	0.846	0.276	0.011	1.582	**0.038**
UMCSENTx	0.060	**6.909**	**0.000**	0.075	**6.646**	**0.000**
DTCOLNVHFNM	-0.019	1.242	0.351	-0.029	1.591	0.206
DTCTHFNM	0.002	0.455	0.657	-0.012	1.308	0.155
INVEST	0.037	1.504	**0.026**	0.020	0.952	0.157
VXOCLSx	-0.021	**7.471**	**0.025**	-0.038	**8.884**	**0.030**
$R_g^2 > 5$		12			12	
$p < 0.05$ counts			28			22
$p < 0.01$ counts			13			10

Note: This table shows the results of Giglio and Xiu three-pass method for estimating risk premium of macroeconomic variables. Columns (2)-(4) display the estimated risk premium γ_g, regression fits R_g^2 and *Wald* test *p-value* for using PCA in the first step of three-pass method. Columns (5)-(7) represent the same results when RP-PCA with $\gamma = 5$ is used to derive intermediary latent factors. The regressions cover the time range from June 1987 to June 2017 (362 periods in total) with five latent factors. R_g^2 in columns (3) and (6) show the value larger than 5% in bold. In columns (4) and (7), **bold** numbers indicates joint significant *Wald's p-values* at the 5% level. The final three rows counts significant values, $R_g^2 > 5$, $p < 0.05$ and $p < 0.01$

Table 3.11 presents the key estimation results. Group A (columns (2)-(4)) reports the risk premium estimated from the conventional PCA latent factors, whereas Group B (columns (5)-(7)) shows summary statistics based on RP-PCA latent factors. Both groups contain the risk premium, R^2 from Equation **??**, and *Wald* test applied to examine the joint significance of PC factors on individual macroeconomic variables. To compare the results, I set the significant estimation thresholds with $R_g^2 > 5\%$ or *Wald* test $p - value < 0.05$.

From the result, there is no significant difference in the estimation of PCA and RP-PCA derived factors if only focusing on R_g^2. 12 macro variables can meet the requirement by filtering the R_g^2 greater than 5%. The only distinction is between *RPI* and *CP3Mx*. *RPI* is well explained by conventional PC factors, but it cannot be measured significantly by latent factors from RP-PCA. Instead, R_g^2 of the *CP3Mx* regression on RP-PCA latent factors is about 5.5%, but the conventional PC factors can only explain around 3.6% of *CP3Mx*. Four of the other variables are related to S&P, the market, and their R_g^2 can be more than 20%, especially the *S&P 500* (R_g^2 equals to 48% highest in 127 variables).

Next, using the criteria of $p < 0.05$, I select 28 and 22 significant macroeconomic variables from Group A: PCA and Group B: RP-PCA, respectively. Moreover, only 13 (PCA) and 10 (RP-PCA) macro variables can reject the null hypothesis of the *Wald* test at 1% significance level that the PC factors have the joint explanatory power to explain the particular macro variables. Likewise, I categorise the macroeconomic variables ($p < 0.05$) into different categories according to McCracken and Ng (2016) description: 1 for *Output and income*, 4 for *Labour market*, 5 for *Consumption, orders and inventiries*, 2 for *Money and credit*, 3 for *Interest and exchange rates*, 8 for *Prices* and 5 for *Stock market*. Notably, in the estimation of the 3PM, *Housing* related macroeconomic variables completely lose the explanatory power on characteristic-adjusted managed portfolios, whilst these variables can have a linear relation with managed portfolios via mimicking portfolio analysis. But Rapach and Zhou (2021) derive a significant *housing* factor from the 3PM, which contradicts this chapter.

Stock market related variables, *S&P 500, S&P industry, S&P div yield*, are insignificant as the macro factor-mimicking portfolios when using characteristic-adjusted managed portfolios as the base assets. However, characteristic-adjusted managed portfolio latent factors from the 3PM PCA or RP-PCA can significantly explain these variables.

In a nutshell, the risk premium results are consistent with Rapach and Zhou (2021) that the majority of individual macro variables do not have significant exposure to the risk of stock returns, even when the excess returns are adjusted for a large panel of firm characteristics.Furthermore, using the first-moment penalty to derive the intermediary latent factors (RP-PCA) does not outperform the conventional PCA. At 5% joint significant level, the former methodology produces 6 fewer significant risk premiums of macroeconomic variables than the latter. Moreover, conventional PCA can provide over one-fifth of the macro variables that are jointly significant at 5% level at least.

3.5.2.3 Interpretation of three-pass method

I can derive 28 out of 127 individual macroeconomic variables from the 3PM results that can be significantly explained by latent factors from characteristic-adjusted managed portfolios. Nevertheless, the correlation between them is still weak, which means that the majority of macroeconomic variables cannnot be the systematic risk factor to build the pricing model.

I expand the estimation to include recognisable systematic risk factors for comparison. According to Giglio and Xiu (2021), 3PM can be interpreted as the mimicking portfolio estimator, which means that this method can correspond to Fama and French (1993) three factors or Hou et al. (2015) q-theory five factors.

Table 3.12: Summary statistics of factors' risk premium

	Group A: PCA			Group B: RPPCA with $\gamma = 5$		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Factor	γ_g	$R_g^2(\%)$	*wald*	γ_g	$R_g^2(\%)$	*wald*
(1) MKT	0.613	68.443	0.000	0.776	67.842	0.000
(2) SMB	0.292	33.798	0.000	0.403	33.151	0.000
(3) HML	-0.171	4.461	0.131	-0.105	2.882	0.685
(4) MKT	0.625	70.134	0.000	0.790	69.452	0.000
(5) ME	0.263	31.922	0.000	0.371	31.579	0.000
(6) IA	-0.110	11.455	0.005	-0.131	10.904	0.082
(7) ROE	-0.203	42.644	0.000	-0.379	41.788	0.000
(8) EG	-0.178	41.400	0.000	-0.257	41.728	0.000

Note: This table shows the results of the estimated risk premium of factors from Fama and French (1993) and Hou et al. (2015) by Giglio and Xiu (2021) three-pass method (3PM). Columns (2)-(4) show the estimated risk premium γ_g, regression fits R_g^2 and *Wald* test *p-value* for applying PCA in the first step of 3PM. Columns (5)-(7) display the same components when RP-PCA with $\gamma = 5$ is used to derive intermediary latent factors. The regressions spans from June 1987 to June 2017 (362 periods in total) with five latent factors. Rowss (1)-(3) represent Fama and French (1993) three factors and rows (4)-(8) represent Hou et al. (2015) q-theory five factors, respectively.

Table 3.12 gives the details about the Fama and French (1993) three factor (FF3) and Hou et al. (2015) five-factor (HXZ5) models. According to Table 3.12 row (3), *HML* from FF3 has no explanatory power for the managed portfolio excess returns by PCA or RP-PCA, and the estimation of the *HML* regression on latent factors failed to reject the null hypothesis of joint significance. The result is consistent with Fama and French (2015) that *HML* is the redundant factor in Fama-French five-factor model (FF5). Aside from that, the other factors all have significant risk premiums that pass the joint *Wald* test and have a relative high R_g^2, and the market factor can have a significant performance in any conditions. Two market factors from FF3 (row (1)) and HXZ5 (row (4)), whose R_g^2 can be at the peak around 70% among all significant macroeconomic variables. Factors related to size, profitability and growth have a relatively acceptable R_g^2, but they still have joint significance at 1% level.

To summarise, the characteristics adjustment method reduces the explanatory power of equity factors, such as market factor, *SMB*. Giglio and Xiu (2021) test these factors with excess returns, which can have R_g^2 over 90%. Based on the results of R_g^2, the majority of individual macroeconomic variables cannot explain excess stock returns when compared to equity risk factors, owing to the weak correlation between macroeconomic and stock returns.

3.5.3 Extension test

From the findings of 3PM, only about one-fifth of macroeconomic variables with a resultant risk premium can be used to construct the systematic risk factors. But they do not have a strong correlation with the base assets (characteristic-adjusted managed portfolios). So I develop the macro factor to extract the common information from macroeconomic variables, particularly significant variables. I establish the analysis from two perspectives. One is based on the macro factors derived from all individual macroeconomic variables, while another focuses only on significant macroeconomic variables from the above 3PM.

Rapach and Zhou (2021) create the macro factors via sparse PCA. The macro factors may centralise the common information from various macroeconomic variables. I examine whether macroeconomic factors can capture risk exposure on characteristic-adjusted managed portfolios using two methods: conventional PCA and weak factor principal component analysis (WF-PCA) by Uematsu and Yamagata (2020).

3.5.3.1 Construct characteristic factors from all macroeconomic variables

Table 3.13 Panel A shows the summary information of WF-PCA. I compute six unobserved factors by using Onatski (2010) criteria, which gives 6 as the optimal number of factors for WF-PCA. In Table 3.13 Panel B, I also apply 3PM to 6 macro factors derived from conventional PCA.

In comparison, here are some findings: First, PCA outperform the WF-PCA. The former method can provide four significant factors, *K2, K3, K4, K6*, at 5% joint significant level of *Wald* test, while WF-PCA only has one macro factor(*K6*) significantly correlated with stock latent factors.[19] Secondly, the RP-PCA method still cannot improve the significance level for the macro related factors. Thirdly, macro factors *K1* to *k5* all haave have low R_g^2, in which the third factor *K3* having the maximum R_g^2 5.5%. However, the last macro factor *K6* has a strong correlation (about 80%) with stock market-related variables (*S&P div yield*).

3.5.3.2 Only significant macroeconomic variables

Incorporating all 127 macroeconomic variables in the 3PM can only yield one significant macro factor (*K6*). The insignificant macro variables have a neg-

[19] I denote *K1* to *K6* as the first to sixth latent factors derived from principal component analysis.

Table 3.13: Risk premium of macro factors derived from all individual macroeconomic variables

Panel A: Macro factors from WF-PCA

	PCA			RP-PCA with $\gamma = 5$		
Factor sequence	γ_g	$R_g^2(\%)$	*wald*	γ_g	$R_g^2(\%)$	*wald*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
K1	-0.022	0.485	0.840	-0.022	0.850	0.986
K2	0.048	3.388	0.098	0.047	3.184	0.053
K3	0.020	2.660	0.060	0.027	2.577	0.176
K4	-0.033	1.627	0.399	-0.028	1.630	0.491
K5	-0.004	0.195	0.987	-0.014	0.996	0.769
K6	-0.105	39.552	0.000	-0.136	39.683	0.000

Panel B: Macro factors from PCA

	PCA			RP-PCA with $\gamma = 5$		
Factor sequence	γ_g	$R_g^2(\%)$	*wald*	γ_g	$R_g^2(\%)$	*wald*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
K1	-0.023	0.527	0.830	-0.022	0.912	0.985
K2	0.058	4.636	0.039	0.059	4.359	0.014
K3	0.033	5.592	0.000	0.040	5.410	0.000
K4	-0.044	3.759	0.022	-0.045	3.466	0.096
K5	0.002	0.378	0.963	-0.008	1.259	0.790
K6	-0.095	33.605	0.000	-0.122	33.752	0.000

Note: This table shows the results of the estimated risk premium of macro factors from by Giglio and Xiu (2021) three-pass method (3PM). The macro factors are derived from 127 macroeconomic variables via Uematsu and Yamagata (2020) WF-PCA (Panel A) and PCA (Panel B). Columns (2)-(4) show the estimated risk premium γ_g, regression fits R_g^2 and *Wald* test *p-value* for applying PCA in the first step of 3PM. Columns (5)-(7) display the same components when RP-PCA with $\gamma = 5$ is used to derive intermediary latent factors. The regressions spans from June 1987 to June 2017 (362 periods in total) with five latent factors. Each row represent one factor (K1 to K6) from specific method.

ative impact on information centralisation. As a result, I only use the previous section's 28 significant macroeconomic variables as the base assets to obtain the macro factors using the above procedures. The 3PM is then used to calculate the risk premium of these derived macro factors. Table 3.14 provides information on the risk premiums of macro factors. To maintain consistency with the scenario of all individual macroeconomic variables, I continue to derive six factors.

According to Table 3.14 Panel A, WF-PCA can provide three jointly significant macro factors (*K1, K2, K6*) with a positive risk premium to measure managed portfolio excess returns with five conventional PC factors. If the latent factors of managed portfolios are derived by RP-PCA, the sixth factor, *K6*, will lose the significance at 5% level (*p-value = 0.053*). Macro factors *K1* and *K2* can be explained by the five PC factors with R_g^2 about 20% and 30%, respectively, while factor *K6* can only be captured by less than 5%. The strong correlation between *K1, K2* and excess stock returns, and weak correlation between *K6* and stock returns, support the criteria proposed by Onatski (2010), who suggests that the optimal number of macroeconomic variables from only significant macroeconomic variables is two.

Table 3.14 Panel B applies the 3PM to six macro factors by using conventional PCA. When the joint significance is considered, PCA still outperforms WF-PCA. The *Wald p-value* indicates that five latent factors from managed portfolios jointly explain four factors *K1, K2, K5* and *K6*. Even when the RP-PCA is used in the first step of the 3PM, the factors remain meaningful. Except the fifth factor *K5*, all of the significant factors have positive risk premiums, which is consistent with Table 3.14 Panel A. Remarkably, R_g^2 for *K5* is only 1.7%, suggesting that *K5* does not have a correlation with the latent factors despite passing the *Wald* test. Apart from that, the R_g^2 for *K2* increases from about 29.5% to 34.7%

Table 3.14: Risk premium of drived from 28 significant macroeconomic variables

Panel A: Macro factors from WF-PCA

	PCA			RPPCA with $\gamma = 5$		
Factor sequence	γ_g	$R_g^2(\%)$	*wald*	γ_g	$R_g^2(\%)$	*wald*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
K1	0.092	19.773	0.000	0.106	19.027	0.000
K2	0.080	29.474	0.000	0.110	29.868	0.000
K3	0.040	1.674	0.339	0.043	2.596	0.442
K4	0.004	1.945	0.148	-0.019	1.859	0.150
K5	0.033	1.480	0.157	0.000	0.646	0.610
K6	0.075	4.624	0.001	0.032	2.615	0.053

Panel B: Macro factors from PCA

	PCA			RPPCA with $\gamma = 5$		
Factor sequence	γ_g	$R_g^2(\%)$	*wald*	γ_g	$R_g^2(\%)$	*wald*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
K1	0.079	13.024	0.000	0.087	12.360	0.000
K2	0.090	34.662	0.000	0.122	34.955	0.000
K3	-0.033	1.311	0.248	-0.041	2.262	0.299
K4	0.047	2.257	0.062	0.022	1.662	0.148
K5	-0.018	1.769	0.000	-0.014	1.691	0.001
K6	0.073	4.585	0.001	0.049	2.861	0.011

Note: This table shows the results of the estimated risk premium of macro factors from by Giglio and Xiu (2021) three-pass method (3PM). The macro factors are derived from 28 significant macroeconomic variables via Uematsu and Yamagata (2020) WF-PCA (Panel A) and PCA (Panel B). Columns (2)-(4) show the estimated risk premium γ_g, regression fits R_g^2 and *Wald* test *p-value* for applying PCA in the first step of 3PM. Columns (5)-(7) display the same components when RP-PCA with $\gamma = 5$ is used to derive intermediary latent factors. The regressions spans from June 1987 to June 2017 (362 periods in total) with five latent factors. Each row represent one factor (K1 to K6) from specific method.

when PCA is used.

Results from macro factor analysis show that including only the most important macroeconomic variables can yield more useful macro factors than including all variables. This implies that some macroeconomic variables are unnecessary in pricing the managed portfolios. I also examine the correlation between significant macro factors(both significant macro variables and all individual macro variables) and all individual macroeconomic variables. the sixth-factor *K6* is highly correlated with the market-related variables, such as *S&P 500, S&P div yield*. Other factors have significant relations with *RPI* or *CPI*, but cannot be adequately explained by the latent factors.

However, the first and second factors, *K1* and *K2* reveal over 40% correlation with market-related information when macro factors are computed using only significant variables, like Table 3.14. *K1* has a correlation around 45% with three market indexes, but the correlations between *K1* and five *Prices* related variables are over 85%.[20] On the contrary, *K2* is not significantly associated with any specific group. While it has a high correlation (62% to 82%) with the market variables, *K2* can be designated as a market factor.

Overall, a group of 28 significant macroeconomic variables can fetch more significant macro factors than including all individual macroeconomic variables. They can achieve marginally significant results in the 3PM with convention PCA applied in the first stage.

3.5.3.3 Applying the PLS method to predicting the expected returns

To explain the characteristic-adjusted managed portfolios, the 3PM provides 28 macroeconomic variables with significant risk premiums. I can derive three feasible macro factors from 28 variables that are significantly correlated

[20]Five *Prices* related macroeconomic variables are *CPITRNSL, CPIULFSL, CUSR0000SA0L2, CUSR0000SA0L5* and *PCEPI*.

with managed portfolios.

The partial least square (PLS) is introduced in Wold (1966) to solve the problem of large-dimensional explanatory variables. He states that the primary goal of PLS is to identify the small dimension of factors that correlate with both responses and predictors. It can also solve the multicollinearity problem, which may exist in the large-dimension of firm characteristics or macroeconomic variables. In this part, I apply the PLS method to test the relationship between managed portfolio excess returns and 28 significant macroeconomic variables, which can aid in the identification of the common component. Then I compare the 3PM and PLS results.

In the beginning, I assume that the macroeconomic variables are linearly related to the managed portfolio excess returns:

$$Y = X\beta'$$

where Y is $T \times L$ 126 characteristic-adjusted managed portfolio excess returns matrix as the response, X is $T \times M$ 28 macroeconomic variable matrix as the predictor. The predictor only includes the variables whose *Wald p-value* is less than 0.05 at a 5% significance level from 128 macroeconomic variables.

Following Kelly and Pruitt (2015), the PLS procedure establishes the following steps: First, demean both managed portfolios and significant macroeconomic variables. Second, apply the PLS regression to calculate $\hat{\beta}$. Finally, obtain the estimated $\hat{y}$ and the forecasting $R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$. I also apply the same methodology to compute the out-of-sample R^2 to investigate the performance of the 28 significant macroeconomic variables.

Table 3.15 reports the percentage explained in-sample R^2 and out-of-sample R^2_{oos} via PLS. Column (1) represents the number of PLS components applied in

Table 3.15: Summary results of significant macroeconomic variables via PLS

# (1)	$var_{macro}(\%)$ (2)	$var_{port}(\%)$ (3)	R^2 (4)	R^2_{oos} (5)
C1	0.196	0.383	0.397	0.230
C2	0.387	0.417	0.430	0.367
C3	0.434	0.445	0.458	0.429
C4	0.473	0.457	0.469	0.484
C5	0.522	0.461	0.473	0.538
C6	0.573	0.463	0.475	0.581
C7	0.617	0.464	0.477	0.632
C8	0.646	0.467	0.479	0.684
C9	0.684	0.468	0.480	0.722
C10	0.720	0.469	0.481	0.760

Note: This table shows the results of partial least square (PLS) regression. I denote the characteristic-adjusted managed portfolios as the response and 28 macro variables with significant *Wald p-value* as the predictor. Column (1) represents the number of common components in the regression. Columns (2) and (3) show the percentage of variance explained by the PLS regression for macro variables and managed portfolios . Columns (4) and (5) show the goodness of fit, in-sample R^2 and out-of-sample R^2_{oos} for the regression of managed portfolios on macro variables with PLS estimated coefficients. The data sample covers from June 1987 to June 2017. The out-of-sample estimation are based on all past information up to period t and start from period 121.

the estimation, ranging from 1 component (C1) to 10 components (C10). Columns (2) and (3) express the percentage of the variance of macroeconomic variables and managed portfolios explained by the PLS component. Increasing the number of the PLS components has a significant effect on capturing the variance of the macroeconomic variables, from 19.6% with one component to 72% with ten. The result means that 28 macroeconomic variables are not tightly correlated with each other, so they require few PLS components to measure the common information. However, one PLS component explains 38.3% of managed portfolios' variance. Adding more components can improve the explanatory power marginally, capturing around 8.6% incremental percentage of explained variance to 46.9%. The low explanatory power indicates that macroeconomic variables have a limited impact on measuring managed portfolios. Column (4) exhibits the R^2 of managed portfolios' regression on macroeconomic variables. The results show that the significant macroeconomic variables can explain no more than 48% of managed portfolio returns using PLS regression with ten components. Column (5) displays the results of out-of-sample R^2, which are significantly affected by the number of PLS components in the regression. One component can only help the macro variables predict 23% of the managed portfolio returns whilst ten PLS components can predict 76% of managed portfolio returns. Unexpectedly, after incorporating four PLS components, the out-of-sample performance ($R^2 = 48.4\%$) surpasses in-sample estimation ($R^2 = 47.3\%$).

Overall, the PLS estimation result is in agreement with the 3PM. The 28 key macroeconomic variables cannot fully explain the managed portfolios. Only the core common macroeconomic information can have a significant contribution on pricing the characteristic-adjusted managed portfolios excess returns. Similar to the study of the 3PM, the market-related data is substantially associated with the first two PLS components.

3.5.4 The multi-factor model

According to the results of the 3PM and PLS regression analyses, the explanatory power is given by market and price associated factors. Using the aforementioned data, I extract the significant macro factors to build a multi-factor model to price the assets. The multi-factor model contains four factors, the market factor ($r_m - r_f$) and three macro factors from PCA, *K1*, *K2* and *K5*. Figure 3.3 exhibits the time-series trend from three significant macro factors from June 1987 to June 2017. During the recession, Figure 3.3a and Figure 3.3b exhibit a rising and then decreasing trend. They could be a sign of rising *inflation* or *prices*. The trend in Figure 3.3c is different, with a decline during the recession and an increase afterwards. However, it is unclear whether the trend is a sign of *growth*.

In this case, I compare the macro multi-factor model with other popular multi-factor models to examine whether the macro-factor model can price the portfolio returns with low pricing error. I evaluate the macro-factor model's findings against those of Fama and French (1993) three-factor model (FF3), Fama and French (2015) five-factor model (FF5), Carhart (1997) four-factor model (FFC4) and Hou et al. (2015) q-theory five-factor model (HXZQ).

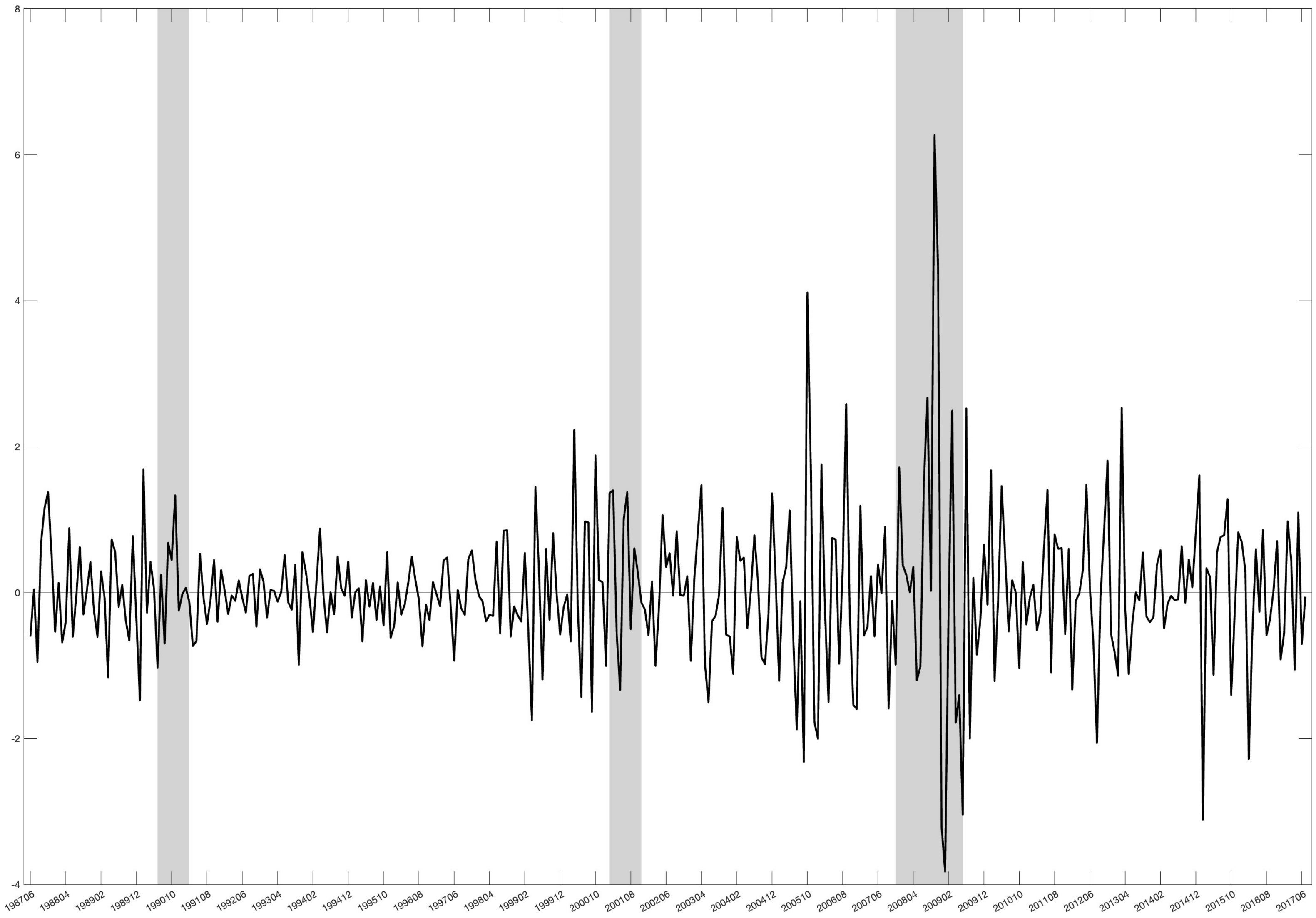
First of all, Table 3.16 Panel A shows how the FF5 and HXZQ factors correlate with the three most important macro factors. The correlation coefficients indicate that the third macro factor is uncorrelated with any of the five factors, suggesting that the third macro factor may be redundant in the multi-factor model. The second factor has a high correlation with the market factor in both model (about 58%), which means the explanatory power of the second macro factor is dominated by the stock market-related information. This result is consistent with the 3PM and PLS regression. In addition, it also have a reasonable correlation with profitability and investment factor in HXZQ.

I run the multi-factor regression with two different groups of test assets. I start with a sample of 183 long-short anomaly portfolio returns. Table 3.16 Panel B provides a breakdown of the estimation pricing errors for the 183 long-short anomaly portfolios returns.[21] The HXZQ has the best performance out of five different multi-factor models. With the highest R^2 and lowest percentage of significant pricing errors, it indicates that the chosen criteria are preponderant and sufficient for valuing the sample assets. The lowest fit among the four conventional multi-factor models is found in FF3, with R^2 of 26.5%. The macro-factor model is even weaker than FF3, with R^2 of 13%.

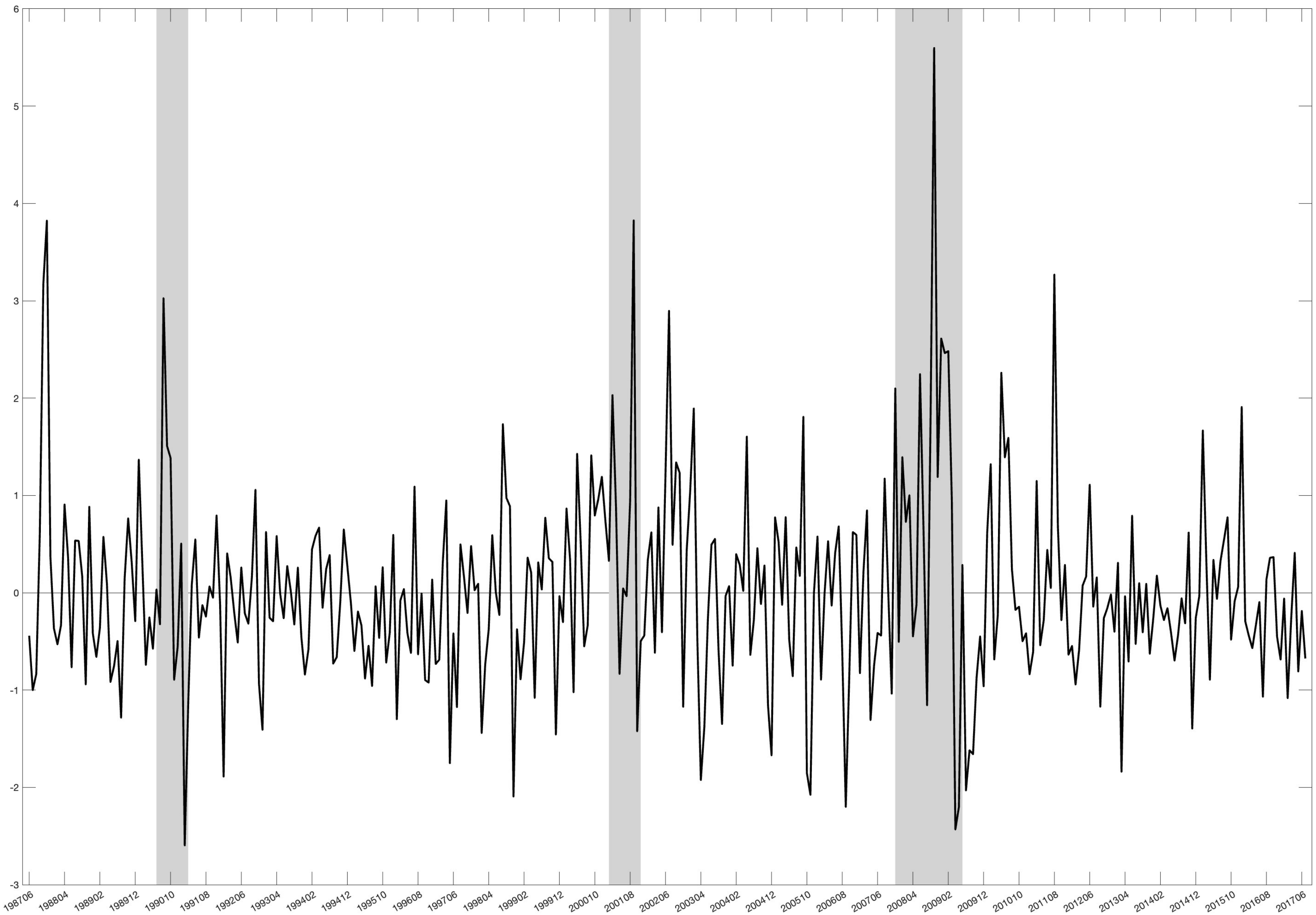
[21]The Open Source Asset Pricing database was developed by Chen and Zimmermann and contains 183 monthly long-short portfolio returns; the database can be accessed at `https://www.openassetpricing.com/data`. The original data sample consists of 205 portfolios. To maintain full data sample availability, I remove portfolios that have missing data.

Figure 3.3: Three macro factors derived from 28 significant macro variables via PCA

(a) Factor 1

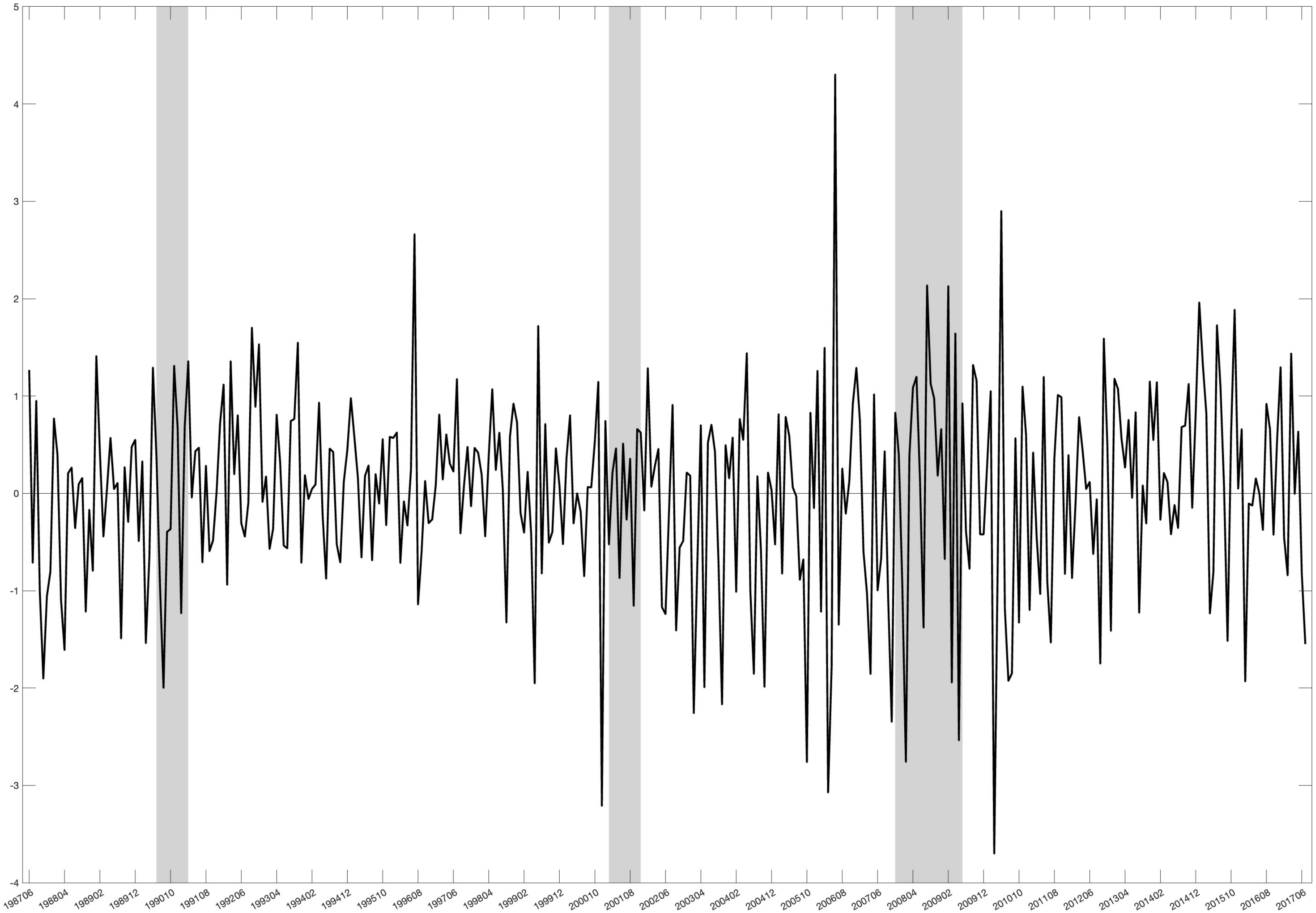


(b) Factor 2



6
5
4
3
2
1
0
-1
-2
-3
198706
198804
198902
198912
199010
199108
199206
199304
199402
199412
199510
199608
199706
199804
199902
199912
200010
200108
200206
200304
200402
200412
200510
200608
200706
200804
200902
200912
201010
201108
201206
201304
201402
201412
201510
201608
201706

(c) Factor 3



Note: This figure shows the summary statistics of managed portfolios for three different transformations. This figure shows the summary statistics for three data transformation from June 1987 to June 2017. It includes the mean value (Mean), 5% percentile value and 95% percentile value. The first one (a) is about the results after applying Green et al. (2017) winsorization. The second (b) is calculated with Kelly et al. (2019) rank transformation. The third (c) represents summary statistics of Kozak et al. (2020) normalized transformation. The vertical gray background shows the recession period by the National Bureau of Economic Research. The data is downloaded from `https://fred.stlouisfed.org/series/USREC`.

Besides, the macro factor can hardly provide an explanation for the long-short anomaly portfolio returns, with the proportion of the significant pricing errors being approximately 73.8% and 60% at 5% and 1% significance level, respectively, compared to 48.6% and 37.2% in HXZQ.

Additionally, this chapter's primary focus is on identifying the connection between between firm characteristics and macroeconomic variables. In Table 3.16 Panel C, I analyse the pricing mistakes made by the characteristic-adjusted managed portfolios. When the managed portfolio returns on the left are ordered and normalised, the summary statistics from all five multi-factor models are comparable. Moreover, all multi-factor model cannot measure the managed portfolio returns better than the long-short portfolio returns since the former scenario can only be evaluated with a 10% R^2. HXZQ still has the best explanatory power of all models with the lowest percentage of unexplained pricing errors and the highest R^2. The HXZQ is the only traditional model that outperforms the other three, meaning that they all fail to accurately price the managed portfolios once adjustments have been made. Due to the inferior 8.4% R^2 and 44.4% significant pricing errors, the macro-factor model cannot outperform the equity-related models. After considering the equity-based multi-factor model, the macro factors barely provide significant explanatory value to the managed portfolio prices. The new multi-factor model may achieve promising results with modest price errors. However, the macroeconomic information cannot perfectly handle the adjusted managed portfolio returns in the current stage.

3.6 Conclusion

The ideas presented in this chapter is inspired by Kelly et al. (2019) the Instrumented Principal Component analysis (IPCA). Characteristic-adjusted man-

Table 3.16: Multi-factor model tests

Panel A: Correlation between market factor and macro factors

Fama-French five-factor model					
Factor	mkt	smb	hml	rmw	cma
factor 1	0.238	0.137	-0.062	-0.186	-0.129
factor 2	0.579	0.204	0.010	-0.185	-0.124
factor 3	-0.081	-0.064	0.072	-0.013	0.080
HXZ q-factor model					
Factor	mkt	me	ia	roe	eg
factor 1	0.256	0.133	-0.101	-0.224	-0.218
factor 2	0.580	0.209	-0.113	-0.289	-0.313
factor 3	-0.066	-0.061	0.086	-0.019	-0.047

Panel B: In-sample Pricing error performance of 183 long-short anomalies returns

Model	Mean	std	R^2	$\%\lvert t\rvert > 1.96$	$\%\lvert t\rvert > 2.58$	$\%\lvert t\rvert > 3$
(1) Fama-French three-factor model	0.572	0.176	0.265	0.738	0.650	0.568
(2) Fama-French five-factor model	0.482	0.181	0.343	0.650	0.514	0.404
(3) Carhart four-factor model	0.499	0.174	0.340	0.656	0.568	0.448
(4) HXZ q-factor model	0.332	0.211	0.383	0.486	0.372	0.306
(5) macro-factor model	0.609	0.203	0.129	0.738	0.601	0.519

Panel C: In-sample Pricing error performance of characteristic-adjusted managed portfolios

Model	Mean	std	R^2	$\%\lvert t\rvert > 1.96$	$\%\lvert t\rvert > 2.58$	$\%\lvert t\rvert > 3$
(1) Fama-French three-factor model	-0.001	0.002	0.098	0.413	0.310	0.262
(2) Fama-French five-factor model	-0.001	0.003	0.109	0.437	0.294	0.262
(3) Carhart four-factor model	-0.001	0.002	0.116	0.444	0.310	0.254
(4) HXZ q-factor model	-0.001	0.003	0.127	0.389	0.302	0.254
(5) macro-factor model	-0.001	0.003	0.084	0.444	0.325	0.262

Note: This table exhibits the summary information of multi-factor model regression on characteristic-adjusted managed portfolios. Panel A shows the correlation between market factor ($r_m - r_f$) and derived macro factors via PCA. Panel B gives the details of estimated pricing error of 183 long-short portfolio returns from Fama-French three-factor model, Fama-French five-factor model, HXZ q-theory model and multi macro-factor model. The last three columns shows the percentage of t-statistics of pricing error at 5%, 1% and 0.1% significance levels. Panel C calculates the pricing error for 126 characteristic-adjusted managed portfolios. The display is the same as Panel B.

aged portfolios are those that I tramsform from $Z_t' R_{t+1}$ to $(Z_t' Z_t)^{-1} Z_t' R_{t+1}$. The primary goal of this chapter is to use the Giglio and Xiu (2021) three-pass method (3PM) to determine the risk premium of macroeconomic variables and to determine the relationship between characteristic-adjusted managed portfolios and these variables.

To maintain uniformity across the managed portfolios, I first use a variety of data correction techniques to tweak the characteristics matrix. They are winsorization at 1% and 99% (Green et al. (2017), GHX), rank transformation with 0.5 shifts (Kelly et al. (2019), KPS) and rank transformation with normalisation (Kozak et al. (2020), KNS). In this regard, the KPS approach may provide a higher degree of goodness-of-fit than the other two. KPS transformation underpins the entire empirical analysis.

Then, I obtain the risk premium of macroeconomic variables from 126 characteristic adjusted managed portfolios with PCA and Risk-premium PCA (RP-PCA, $\gamma = 5$) in the first step of the 3PM. I examine the R^2 and *Wald* test for the risk premiums of macroeconomic variables.

Some conclusions are supported by the empirical work: First, RP-PCA cannot improve the outcome on the basis of conventional PCA. Second, I only collect 28 of the 127 macroeconomic variables that potentially have considerable risk premiums, with the majority of them being stock market-related. Third, I classify 28 macroeconomic variables with significant risk premiums into seven categories, with *Prices* and *Stock market* being the most significant. In contrast to the mimicking portfolio analysis, the output from 3PM can have consistently significant *Prices* and *Stock market* associated variables.

Additionally, I use the 3PM to determine risk premia for macro components, which are obtained from macroeconomic variables by conventional PCA and Uematsu and Yamagata (2020) weak-factor (WF-PCA). Conventional PCA can provide four macro factors with significant risk premium while WF-PCA can only have three.

I also present PLS analysis as a means of identifying the common component between the macro variables and managed portfolios. The percentage of explained variance expresses that the macroeconomic variables have limited explanatory power on pricing the characteristic-adjusted managed portfolios. The stock market continues to play the biggest role.

By the end of the chapter, I have compared four widely used multi-factor models to the explanatory power of derived macro factors on managed portfolio and long-short portfolio returns. The findings suggest: First, the long-short portfolio returns can be better explained by any multi-factor model than the rank-transformed managed portfolio returns. Second, the macro-factor model is ineffective in pricing managed portfolios, as pricing mistakes from regressions are more pronounced than with other multi-factor models, and the R^2 is also lower than with other models. Third, only the macro factors related to the *Stock market* can have a significant impact on pricing the managed portfolios, while all other macro factors fail to provide sufficient explanatory power to capture the left-hand side managed portfolio returns in the tesed macro-factor model.

In sum, the 3PM is used in this chapter to extrapolate the risk premiums of macroeconomic variables from the characteristic-adjusted managed portfolios. The result is consistent with Rapach and Zhou (2021) that most independent macro variables or macro factors fail to price stock portfolios, whereas those tied to equities can. Further research related to characteristics adjustment and also study about the macro factors should be considered.

Appendix A

Supplementary Tables for Chapter 3

Table A.1: List of characteristics full details

abb.	Name	Frequency	Citation	Details
prctmaxprc	current price to high price over last year		George and Hwang (2004)	$\frac{prc_t}{\max prc_{t,t-1}}$
cash	Cash holdings scaled by total assets	annual	Palazzo (2012)	$\frac{che_t}{at_t}$
taxch	Change in tax expense	annual		$\frac{txt_t - txt_{t-1}}{at_t}$
cinvest	Change in investment over average of prior three years investment	quarterly	Francis et al. (2004)	$\frac{invest_t - invest_{t-1}}{(invest_{t-1}+invest_{t-2}+invest_{t-3})/3}, invest_t = \frac{ppentq_t - ppentq_{t-1}}{saleq_t}$
earnvol	earning volatility	annual	Francis et al. (2004)	$\sigma_{4Y}(ni_t)$
roavol	return on assets volatility	quarterly	Francis et al. (2004)	$\sigma_{16Q}(\frac{ibq_t}{atq_t})$
roaq	Return on assets	quarterly	Balakrishnan et al. (2010)	$\frac{ibq_t}{atq_{t-1}}$
roeq	Return on book equity	quarterly	Hou et al. (2015)	$\frac{ibq_t}{be_{t-1}}$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
rsup	Revenue surprise	quarterly	Jegadeesh and Livnat (2006)	$\frac{saleq_t - saleq_{t-4}}{cshoq_t \times prccq_t}$
stdacc	accrual volatility	quarterly		$\sigma(acc_t)$
cfvol	cash flow volatility	quarterly	Huang (2009)	$\sigma_4 Q((ibq_t + dpq_t - wcapq_t))$
abcinv	abnormal corporate investment		Titman et al. (2004)	$\frac{capx_t / sale_t}{[(capx_{t-1}/sale_{t-1}) + (capx_{t-2}/sale_{t-2}) + (capx_{t-3}/sale_{t-3})]/3} - 1$
oacc	Operating accruals	annual		$\frac{ib_t - oancf_t}{(at_t + at_{t-1})/2}$
absoacc	Absolute accruals	annual		$\lvert oacc_t \rvert$
pctacc	Percent accruals	annual	Hafzalla et al. (2011)	$\frac{oacc_t}{ib_t}$
zscore	Altman Z-score	annual	Dichev (1998)	$1.2 \times \frac{act_t - lct_t}{at_t} + 1.4 \times \frac{re_t}{at_t} + 3.3 \times \frac{ebitda_t}{at_t} + 0.6 \times \frac{prcc_t \times csho_t}{lt_t} + \frac{sale_t}{at_t}$
atpch	Asset growth 1 year	annual	Cooper et al. (2008)	$\frac{at_t}{at_{t-1}} - 1$
atan	asset tangibility	annual	Hahn and Lee (2009)	$\frac{che_t + (0.715 \times rect_t + 0.547 \times invt_t + 0.535 \times ppegt)}{at_t}$
ato	asset turnover	annual		$\frac{sale_t}{(at_t + at_{t-1})/2}$ or $\frac{sale_t}{noa_{t-1}}$
bleve	book leverage		Fama and French (1992)	$\frac{at_t}{be_t}$
cashtdebt	Cash flow to debt	annual	Ou and Penman (1989)	$\frac{ib_t + dp_t + txdi_t}{lt_t}$
cboplag	cash-based lagged operating profitability	annual	Ball et al. (2016)	$\frac{cop_t}{at_{t-1}}$
cbop	cash-based operating profitability			$\frac{cop_t}{at_t}$
cta	cash-to-assets	annual	Palazzo (2012)	$\frac{che_t}{at_t}$
ocfta	operating cash flow -to-assets	annual	Bouchaud et al. (2019)	$\frac{oancf_t}{at_t}$
dpch3	change in total debt 3 year		Lyandres et al. (2008)	$\frac{dltt_t + dlc_t}{dltt_{t-3} + dlc_{t-3}} - 1$
flch3	change in financial liabilities		Richardson et al. (2005)	$\frac{(dltt_t + dlc_t + pstk_t) - (dltt_{t-3} + dlc_{t-3} + pstk_{t-3})}{at_t}$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
coach	Change in current operating assets		Richardson et al. (2005)	$\frac{(act_t - che_t) - (act_{t-1} - che_{t-1})}{at_t}$
colch	Change in current operating liabilities		Richardson et al. (2005)	$\frac{(lct_t - dlc_t) - (lct_{t-1} - dlc_{t-1})}{at_t}$
cowcch	change in current operating working capital	annual	Richardson et al. (2005)	$\frac{coa_t - col_t}{at_t}$
invch	Change in inventory	annual	Belo and Lin (2012)	$\frac{invt_t - invt_{t-1}}{at_t}$ or $\frac{invch_t}{at_t}$
nfach	change in net financial assets		Richardson et al. (2005)	$\frac{(fa_t - fl_t) - (fa_{t-1} - fl_{t-1})}{at_t}$
nncoach	change in net noncurrent operating assets		Richardson et al. (2005)	$\frac{ncoa_t - ncol_t}{at_t}$
ncoach	change in noncurrent operating assets		Richardson et al. (2005)	$\frac{(at_t - act_t - ivao_t) - (at_{t-1} - act_{t-1} - ivao_{t-1})}{at_t}$
ncolch	change in noncurrent operating liabilities	annual	Richardson et al. (2005)	$\frac{(lt_t - lct_t - dltt_t) - (lt_{t-1} - lct_{t-1} - dltt_{t-1})}{at_t}$
dcv	convertible debt	annual		$\frac{dcvt_t}{dlc_t + dltt_t}$
dcvind	convertible debt indicator	annual	Valta (2016)	dcv is available, $dcvind = 1$
cr	Current ratio	annual	Ou and Penman (1989)	$\frac{act_t}{lct_t}$
crpch	% change in current ratio	annual	Ou and Penman (1989)	$\frac{cr_t}{cr_{t-1}} - 1$
depr	depreciation percentage	annual	Holthausen and Larcker (1992)	$\frac{dp_t}{ppent_t}$
deprpch	% change in depreciation	annual	Holthausen and Larcker (1992)	$\frac{dp_t}{dp_{t-1}} - 1$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
earnvari	Earnings variability	annual	Pontiff and Woodgate (2008)	$\frac{\sigma_{5Y}(ni_t/at_{t-1})}{\sigma_{5Y}(ocf_t/at_{t-1})}$
emppch	Employee growth rate	annual		$\frac{emp_t - emp_{t-1}}{0.5\times emp_t + 0.5\times emp_{t-1}}$
gp	Gross profitability	annual	Novy-Marx (2013)	$\frac{gp_t}{at_t}$
gplag	Gross profitability lagged assets	annual		$\frac{gp_t}{at_{t-1}}$
capxch1	Change in capital expenditures 1 year	annual	Xie (2001)	$\frac{capx_t - capx_{t-1}}{at_t}$
capxch2	Change in capital expenditures 2 years	annual	Anderson and Garcia-Feijóo (2006)	$\frac{capx_t - capx_{t-2}}{at_t}$
capxch3	Change in capital expenditures 3 years	annual	Anderson and Garcia-Feijóo (2006)	$\frac{capx_t - capx_{t-3}}{at_t}$
capxch2	Change in capital expenditures 2 years	annual	Anderson and Garcia-Feijóo (2006)	$\frac{capx_t}{capx_{t-2}} - 1$
ceqpch	Growth in common shareholder equity	annual	Richardson et al. (2005)	$\frac{ceq_t}{ceq_{t-1}} - 1$
ltdpch	growth in long-term debt	annual	Richardson et al. (2005)	$\frac{dltt_t}{dltt_{t-1}} - 1$
ltnoach	growth in long-term net operating assets scaled by average total assets	annual	Fairfield et al. (2003)	$\frac{lnoa_t - lnoa_{t-1}}{(at_t + at_{t-1})/2}, lnoa_t = ppent_t + intan_t + ao_t - lo_t + dp_t$
invpch	percent change in inventory 1 year	annual	J. K. Thomas and Zhang (2002)	$\frac{invt_t}{invt_{t-1}} - 1$
invest	property investment change scaled by assets	annual	Lyandres et al. (2008)	$\frac{ppeinv_t - ppeinv_{t-1}}{at_t}, ppeinv_t = ppegt_t + invt_t$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
lgr	Percent change in liability	annual	Lyandres et al. (2008)	$\frac{lt_t}{lt_{t-1}} - 1$
saleemppch	Labour force efficiency	annual	Abarbanell and Bushee (1998)	$\frac{sale_t/emp_t}{sale_{t-1}/emp_{t-1}} - 1$
liqat	liquidity of book assets	annual	Ortiz-Molina and Phillips (2014)	$\frac{che_t+0.75\times coa_t+0.5\times(at_t-act_t-intan_t)}{at_t}$
ndtp	net debt to price	annual	Penman et al. (2007)	$\frac{dt_t-che_t}{me_t}$
nceqiss	Net equity issuance	annual	Francis et al. (2004)	$\frac{sstk_t-prstkc_t}{at_t}$
noa	Net operating assets	annual	Hirshleifer et al. (2004)	$\frac{oa_t-ol_t}{at_t}$
age	number of year	month		age of the firm in month for Compustat database from the beginning until period t
oscore	Ohlson O-score	annual	Ohlson (1980)	$-1.32 - 0.407 \times at_{t-1} + 6.03 \times \frac{debt_t}{at_t} + 1.43 \times \frac{act_t-lct_t}{at_t} + 0.076 \times \frac{lct_t}{act_t} - 1.72 \times ind_{lt_t>at_t} - 2.37 \times \frac{ib_t}{at_t} - 1.83 \times \frac{pi_t+dp_t}{lt_t} + 0.285 \times ind_{ni_t<0,ni_{t-1}<0} - 0.52 \times \frac{ni_t-ni_{t-1}}{\lvert ni_t\rvert+\lvert ni_{t-1}\rvert}$
oleve	operating leverage	annual		$\frac{xopr_t}{at_t}$
op	Operating profitability	annual		$\frac{ebitda_t+xrd_t}{at_t}$
oplag	Operating profitability lagged book assets	annual	Ball et al. (2016)	$\frac{ebitda_t+xrd_t}{at_{t-1}}$
oaccni	Percent accruals	annual	Hafzalla et al. (2011)	$\frac{oacc_t}{\lvert ib_t\rvert}$
taccni	Percent total accruals	annual	Hafzalla et al. (2011)	$\frac{tacc_t}{\lvert ib_t\rvert}$, $tacc_t = oacc_t + (nfa_t - nfa_{t-1})$
fscore	Pitroski F-score	annual	Piotroski (2000)	$roa_{>0,t} + (\frac{oancf_t}{at_{t-1}})_{>0} + (roa_t - roa_{t-1})_{>0} + (\frac{oancf_t}{at_{t-1}} - roa_t)_{>0} + (\frac{dltt_t}{at_t} - \frac{dltt_{t-1}}{at_{t-1}})_{<0} + crch_{t,>0} + sstk_{t,=0} + gmch_{t,>0} + aturnch_{t,>0}$
qr	Quick ratio	annual	Ou and Penman (1989)	$\frac{act_t-invt_t}{lct_t}$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
rdctba	R&D capital to book assets	annual	Li (2011)	$\frac{\sum_{n=0}^{4}(1-.2\times n)(xrd_{t-n})}{at_t}$
rdat	R&D to assets	annual	Guo et al. (2006)	$\frac{xrd_t}{at_t}$
rdind	R&D increase	annual	Eberhart et al. (2004)	$(\frac{xrd_t}{at_t})_{>0.05})$
rdsale	R&D to sales	annual	Guo et al. (2006)	$\frac{xrd_t}{sale_t}$
roic	Return on invested capital	annual	Brown and Rowe (2007)	$\frac{ebit_t-nopi_t}{ceq_t+lt_q-che_t}$
ronoa	return on net operating assets	annual	Soliman (2008)	$\frac{ebit_t}{icapt_t+dlc_t-che_t}$
salepch3	Sales growth 3 years	annual	Lakonishok et al. (1994)	$\frac{sale_t}{sale_{t-3}}-1$
salecash	Sales to cash	annual	Ou and Penman (1989)	$\frac{sale_t}{che_t}$
Pchgmpchsale	% change in gross margin - % change in sales	annual	Abarbanell and Bushee (1998)	$(\frac{sale_t-cogs_t}{sale_{t-1}-cogs_{t-1}}-1)(\frac{sale_t}{sale_{t-1}}-1)$
Pchsalepchxsga	% change in sales - % change in SG&A	annual	Abarbanell and Bushee (1998)	$(\frac{sale_t}{sale_{t-1}}-1)(\frac{xsga_t}{xsga_{t-1}})$
saleinv	Sales to inventory	annual	Ou and Penman (1989)	$\frac{sales_t}{invt_t}$
pchsaleinv	% change in sales-to-inventory ratio	annual	Ou and Penman (1989)	$\frac{sale_invt_t}{sale_invt_{t-1}}-1$
salerec	Sales to receivables	annual	Ou and Penman (1989)	$\frac{sale_t}{rect_t}$
Secured	Secured debt scaled by total liabilities	annual		$\frac{dm_t}{lt_t}$
securedind	Secured debt indicator	annual	Valta (2016)	$(dm_t)_{>0}$
stinvch	change in short-term investment	annual	Richardson et al. (2005)	$\frac{ivst_t-ivst_{t-1}}{at_t}$
txtsup	tax expenses surprise	annual	Thomas and Zhang (2011)	$\frac{txt_t-txt_{t-1}}{at_t}$
tacc	total accruals	annual	Richardson et al. (2005)	$\frac{oacc_t+nfna_t}{at_t}$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
mom1m	1-month momentum (short-term reversal)	monthly	Jegadeesh and Titman (1993)	ret_{t-1}
mom12m	12-month momentum (prior one year return)	monthly	Fama and French (1996)	$mom12 = \prod_{m \in \{t-12:t-1\}} (ret_m + 1) - 1$
mom6	6-month momentum 6-1	monthly	Jegadeesh and Titman (1993)	$\prod_{m \in \{t-6:t-1\}} (ret_m + 1) - 1$
mom36	36-month momentum skip last year	monthly	Jegadeesh and Titman (1993)	$\prod_{m \in \{t-36:t-13\}} (ret_m + 1) - 1$
mret60	Prior five years return	monthly	Daniel and Titman (2006)	$\prod_{m \in \{t-36:t-13\}} (ret_m + 1) - 1$
chmom	Change in 6-month momentum	monthly		$mom_{t-1,t-6} - mom_{t-7,t-12}$
iss	Composite issuance	annual		$\frac{me_t}{me_{t-3}} - ret_{t,t-3}$
dolvol	dollar trading volumn	monthly	Chordia et al. (2001)	$\sum_{t,t-2} prc_t \times vol_t$
onep	1/share price	monthly	Miller and Scholes (1982)	$\frac{1}{prc_t}$
turn	Share turnover	monthly	Datar et al. (1998)	$\frac{(\sum_{t=0}^{2})/3}{share_t}$
maxret	maximum daily return	daily		$\max_m onth(ret_t)$
retvol	Return volatility	daily	Ang et al. (2006)	$\sigma_1 m(ret_t)$
baspread	high-low bid-ask spread	monthly	Corwin and Schultz (2012)	$S = \frac{2(e^\alpha - 1)}{1+e^\alpha}$ $\alpha = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma}{3-2\sqrt{2}}}$ $\beta = E\left\{\sum_{j=0}^{1}\left[\ln\left(\frac{H_{t+j}^0}{L_{t+j}^0}\right)\right]^2\right\}$ $\gamma = \left[\ln\left(\frac{H_{t,t+1}^0}{L_{t,t+1}^0}\right)\right]^2$
zerotrade	Zero trading days	monthly		
dolvol	dollar trading volumn	monthly	Chordia et al. (2001)	$\sum_{t,t-2} prc_t \times vol_t$
beta	market beta	monthly	Fama and MacBeth (1973)	$\frac{covar(ret_t, mret_t)}{var_(mret_t)}$
beta	Square of market beta	monthly	Fama and MacBeth (1973)	β^2

Table A.1 continued from previous page

abb.	name	frequency	citation	details
ill	Illiquidity	monthly (daily mean)		$mean(\frac{\|ret_t\|}{dvol_t})$
stddolvol	Volatility of liquidity (dollar trading volume)	daily	Chordia et al. (2001)	$\frac{\sigma_{1m}(dvol_t)}{dvol_t}$
stdturn	Volatility of liquidity (share turnover)	daily	Chordia et al. (2001)	$\frac{\sigma_{1m}(turn_t)}{turn_t}$
beta_capm	Downside beta	monthly	Ang et al. (2006)	$r_{it} = a_i + \beta_i r_{mt} + \epsilon_{it}$
beta_down	Downside beta	monthly	Ang et al. (2006)	$r_{it} = a_i^- + \beta_i^- r_{mt} + \epsilon_{it}^-$, whenever $r_{mt} \leq \bar{r}_m - \sigma_{r_m}$
ivolcapm	idiosyncratic volatility CAPM	monthly (daily mean)		$\sigma(\epsilon_t^{capm})$
ivolff3	idiosyncratic volatility ff3	monthly (daily mean)	Ang et al. (2006)	$\sigma(\epsilon_t^{ff3})$
ivolq	idiosyncratic volatility qmodel	monthly (daily mean)		$\sigma(\epsilon_t^{q})$
ivoldown	idiosyncratic volatility downside beta from CAPM	monthly (daily mean)		$\sigma(\epsilon_t^{down})$
iskewcapm	idiosyncratic skewness CAPM	monthly	Bali et al. (2016)	$skew(\epsilon_t^{capm})$
iskewff3	idiosyncratic skewness ff3	monthly	Bali et al. (2016)	$skew(\epsilon_t^{ff3})$
iskewq	idiosyncratic skewness qmodel	monthly	Bali et al. (2016)	$skew(\epsilon_t^{q})$
iskewq	idiosyncratic skewness downside	monthly	Bali et al. (2016)	$skew(\epsilon_t^{down})$
me	Size	monthly	Banz (1981)	me_t
atme	Assets-to-market capitalization	annual	Bhandari (1988)	$\frac{at_t}{me_t}$
btm	Book-to-market ratio	annual	Rosenberg et al. (1985)	$\frac{be_t}{me_t}$

Table A.1 continued from previous page

abb.	name	frequency	citation	details
cashpr	Cash productivity	annual	Chandrashekar and Rao (2006)	$\frac{me_t+dltt_t-at_t}{che_t}$
cfp	Cash-flow-to-price ratio	annual	Desai et al. (2004)	$\frac{oancf_t}{me_t}$
divy	Dividend Yield	annual	Litzenberger and Ramaswamy (1979)	$\frac{dv_t \times csho_t}{me_t}$
ep	Earnings to price	annual		$\frac{ni_t}{me_t}$
rdme	R&D to market capitalization	annual	Guo et al. (2006)	$\frac{xrd_t}{me_t}$
stp	Sales to price	annual	William C. Barbee et al. (1996)	$\frac{sale_t}{me_t}$

Note: This table delivers the details of all characteristics. The first column is the abbreviations that we use in this chapter, and the second column shows the name of each characteristic. The third column is the calculation frequency. The fourth column shows the citation paper. The fifth final one displays the detail of the calculate methodology and all items can be collected from CRSP/Compustat database.

References

Abarbanell, J. S. and B. J. Bushee (1998, January). Abnormal returns to a fundamental analysis strategy. *Accounting Review 73*(1), 19.

Alam, M. M. and G. Uddin (2009). Relationship between Interest Rate and Stock Price: Empirical Evidence from Developed and Developing Countries.

Allozi, N. M. and G. S. Obeidat (2016). The Relationship between the Stock Return and Financial Indicators (Profitability, Leverage): An Empirical Study on Manufacturing Companies Listed in Amman Stock Exchange. *Journal of Social Sciences (COES&RJ-JSS) 5*(3), 408–424.

Anderson, C. W. and L. Garcia-Feijóo (2006). Empirical Evidence on Capital Investment, Growth Options, and Security Returns. *The Journal of Finance 61*(1), 171–194.

Ang, A. and G. Bekaert (2004, March). How Regimes Affect Asset Allocation. *Financial Analysts Journal 60*(2), 86–99.

Ang, A. and G. Bekaert (2007, May). Stock Return Predictability: Is it There? *The Review of Financial Studies 20*(3), 651–707.

Ang, A., J. Chen, and Y. Xing (2006, December). Downside Risk. *The Review of Financial Studies 19*(4), 1191–1239.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The Cross-Section of Volatility and Expected Returns. *The Journal of Finance 61*(1), 259–299.

Asness, C., A. Frazzini, R. Israel, and T. Moskowitz (2015, October). Fact, Fiction, and Value Investing. *The Journal of Portfolio Management 42*(1), 34–52.

Asness, C. S. (1995, January). The Power of Past Stock Returns to Explain Future Stock Returns.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica 70*(1), 191–221.

Balakrishnan, K., E. Bartov, and L. Faurel (2010, May). Post loss/profit announcement drift. *Journal of Accounting and Economics 50*(1), 20–41.

Bali, T. G., R. F. Engle, and S. Murray (2016, February). *Empirical Asset Pricing: The Cross Section of Stock Returns*. John Wiley & Sons.

Ball, R. (1978, June). Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics 6*(2), 103–126.

Ball, R. and P. Brown (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research 6*(2), 159–178.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2016, July). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics 121*(1), 28–45.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2020, January). Earnings, retained earnings, and book-to-market in the cross section of expected returns. *Journal of Financial Economics 135*(1), 231–254.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. V. Nikolaev (2015, August). Deflating profitability. *Journal of Financial Economics 117*(2), 225–248.

Banz, R. W. (1981, March). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*(1), 3–18.

Basu, S. (1983, June). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics 12*(1), 129–156.

Beccalli, E., B. Casu, and C. Girardone (2006). Efficiency and Stock Performance in European Banking. *Journal of Business Finance & Accounting 33*(1-2), 245–262.

Belo, F. and X. Lin (2012, January). The Inventory Growth Spread. *The Review of Financial Studies 25*(1), 278–313.

Berk, J. B. (1995). A critique of size-related anomalies. *The Review of Financial Studies 8*(2), 275–286.

Bhandari, L. C. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *The Journal of Finance 43*(2), 507–528.

Blackburn, D. W. and N. Cakici (2019, June). Book-To-Market Decomposition, Net Share Issuance, and the Cross Section of Global Stock Returns. *Journal of Risk and Financial Management 12*(2), 90.

Bouchaud, J.-P., P. Krüger, A. Landier, and D. Thesmar (2019). Sticky Expectations and the Profitability Anomaly. *The Journal of Finance 74*(2), 639–674.

Brown, D. P. and B. Rowe (2007, June). The Productivity Premium in Equity Returns.

Campbell, J. Y. (2017). *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton University Press.

Campbell, J. Y. and R. J. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies 1*(3), 195–228.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance 52*(1), 57–82.

Chandrashekar, S. and R. Rao (2006). The Productivity of Cash and the Cross-Section of Expected Stock Returns.

Chang, K. J., D. C. Chichernea, and H. R. HassabElnaby (2014, January). On the DuPont analysis in the health care industry. *Journal of Accounting and Public Policy 33*(1), 83–103.

Chaudhry, N. (2021, June). Tax Aggressiveness and Idiosyncratic Volatility.

Chen, A. Y. and T. Zimmermann (2021, May). Open Source Cross-Sectional Asset Pricing. SSRN Scholarly Paper ID 3604626, Social Science Research Network, Rochester, NY.

Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic Forces and the Stock Market. *The Journal of Business 59*(3), 383–403.

Chordia, T., A. Subrahmanyam, and V. R. Anshuman (2001, January). Trading activity and expected stock returns. *Journal of Financial Economics 59*(1), 3–32.

Cochrane, J. H. (1996). A Cross-Sectional Test of an Investment-Based Asset Pricing Model. *Journal of Political Economy 104*(3), 572–621.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of Finance 66*(4), 1047–1108.

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset Growth and the Cross-Section of Stock Returns. *The Journal of Finance 63*(4), 1609–1651.

Corwin, S. A. and P. Schultz (2012). A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *The Journal of Finance 67*(2), 719–760.

Cutler, D. M., J. M. Poterba, and L. H. Summers (1988, March). What Moves Stock Prices?

Daniel, K. and S. Titman (2006). Market Reactions to Tangible and Intangible Information. *The Journal of Finance 61*(4), 1605–1643.

Daniel, K. and S. Titman (2016, May). Another Look at Market Responses to Tangible and Intangible Information. *Critical Finance Review 5*(1), 165–175.

Datar, V. T., N. Y. Naik, and R. Radcliffe (1998, August). Liquidity and stock returns: An alternative test. *Journal of Financial Markets 1*(2), 203–219.

De BONDT, W. F. M. and R. Thaler (1985). Does the Stock Market Overreact? *The Journal of Finance 40*(3), 793–805.

DeAngelo, H., L. DeAngelo, and R. M. Stulz (2006, August). Dividend policy and the earned/contributed capital mix: A test of the life-cycle theory. *Journal of Financial Economics 81*(2), 227–254.

Desai, H., S. Rajgopal, and M. Venkatachalam (2004). Value-Glamour and Accruals Mispricing: One Anomaly or Two? *The Accounting Review 79*(2), 355–385.

Dhaliwal, D., L. Krull, and O. Z. Li (2007, March). Did the 2003 Tax Act reduce the cost of equity capital? *Journal of Accounting and Economics 43*(1), 121–150.

Dhaliwal, D., L. Krull, O. Z. Li, and W. Moser (2005). Dividend taxes and implied cost of equity capital. *Journal of Accounting Research 43*(5), 675–708.

Dichev, I. D. (1998). Is the Risk of Bankruptcy a Systematic Risk? *The Journal of Finance 53*(3), 1131–1147.

Dimitrov, V. and P. C. Jain (2008, April). The Value-Relevance of Changes in Financial Leverage Beyond Growth in Assets and GAAP Earnings. *Journal of Accounting, Auditing & Finance 23*(2), 191–222.

Dittmar, A. K. (2000). Why Do Firms Repurchase Stock. *The Journal of Business 73*(3), 331–355.

Eberhart, A. C., W. F. Maxwell, and A. R. Siddique (2004). An Examination of Long-Term Abnormal Stock Returns and Operating Performance Following R&D Increases. *The Journal of Finance 59*(2), 623–650.

Eisfeldt, A. L. and D. Papanikolaou (2013). Organization Capital and the Cross-Section of Expected Returns. *The Journal of Finance 68*(4), 1365–1406.

Eriotis, N. P., Z. Frangouli, and Z. Ventoura-Neokosmides (2002, April). Profit Margin And Capital Structure: An Empirical Relationship. *Journal of Applied Business Research (JABR) 18*(2).

Fairfield, P. M., J. S. Whisenant, and T. L. Yohn (2003). Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing. *The Accounting Review 78*(1), 353–371.

Fairfield, P. M. and T. L. Yohn (2001, December). Using Asset Turnover and Profit Margin to Forecast Changes in Profitability. *Review of Accounting Studies 6*(4), 371–385.

Fama, E. F. (1981). Stock Returns, Real Activity, Inflation, and Money. *The American Economic Review 71*(4), 545–565.

Fama, E. F. and K. R. French (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance 47*(2), 427–465.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (1995). Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance 50*(1), 131–155.

Fama, E. F. and K. R. French (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance 51*(1), 55–84.

Fama, E. F. and K. R. French (2006, December). Profitability, investment and average returns. *Journal of Financial Economics 82*(3), 491–518.

Fama, E. F. and K. R. French (2008a). Average Returns, B/M, and Share Issues. *The Journal of Finance 63*(6), 2971–2995.

Fama, E. F. and K. R. French (2008b). Dissecting anomalies. *The Journal of Finance 63*(4), 1653–1678.

Fama, E. F. and K. R. French (2015, April). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and J. D. MacBeth (1973). Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy 81*(3), 607–636.

Fan, J., Y. Liao, and W. Wang (2016, February). PROJECTED PRINCIPAL COMPONENT ANALYSIS IN FACTOR MODELS. *Annals of statistics 44*(1), 219–254.

Francis, J., R. LaFond, P. M. Olsson, and K. Schipper (2004, October). Costs of Equity and Earnings Attributes. *The Accounting Review 79*(4), 967–1010.

Freyberger, J., A. Neuhierl, and M. Weber (2020, May). Dissecting Characteristics Nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

George, T. J. and C.-Y. Hwang (2004). The 52-Week High and Momentum Investing. *The Journal of Finance 59*(5), 2145–2176.

George, T. J. and C.-Y. Hwang (2010, April). A resolution of the distress risk and leverage puzzles in the cross section of stock returns. *Journal of Financial Economics 96*(1), 56–79.

Gerakos, J. and J. T. Linnainmaa (2014, July). Market Reactions to Tangible and Intangible Information Revisited.

Gerakos, J. and J. T. Linnainmaa (2016, May). Market Reactions to Tangible and Intangible Information Revisited. *Critical Finance Review 5*(1), 135–163.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004, June). The MIDAS Touch: Mixed Data Sampling Regression Models.

Giglio, S. and D. Xiu (2021, July). Asset Pricing with Omitted Factors. *Journal of Political Economy 129*(7), 1947–1990.

Graham, B., S. Cottle, C. Tatham, and D. L. Dodd (1962). Security analysis: Principles and technique.

Graham, B., D. Dodd, and D. L. F. Dodd (1934). *Security Analysis: The Classic 1934 Edition*. McGraw Hill Professional.

Green, J., J. R. M. Hand, and X. F. Zhang (2013, September). The supraview of return predictive signals. *Review of Accounting Studies 18*(3), 692–730.

Green, J., J. R. M. Hand, and X. F. Zhang (2017, December). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies 30*(12), 4389–4436.

Grullon, G. and R. Michaely (2004). The Information Content of Share Repurchase Programs. *The Journal of Finance 59*(2), 651–680.

Guo, R.-J., B. Lev, and C. Shi (2006). Explaining the Short- and Long-Term IPO Anomalies in the US by R&D. *Journal of Business Finance & Accounting 33*(3-4), 550–579.

Hafzalla, N., R. Lundholm, and E. M. Van Winkle (2011). Percent Accruals. *The Accounting Review 86*(1), 209–236.

Hahn, J. and H. Lee (2009). Financial Constraints, Debt Capacity, and the Cross-section of Stock Returns. *The Journal of Finance 64*(2), 891–921.

Hail, L., S. Sikes, and C. Wang (2017). Cross-country evidence on the relation between capital gains taxes, risk, and expected returns. *Journal of Public Economics 151*, 56–73.

Harkavy, O. (1953). The relation between retained earnings and common stock prices for large, listed corporations. *The Journal of Finance 8*(3), 283–297.

Harvey, C. R., Y. Liu, and H. Zhu (2016). … and the cross-section of expected returns. *The Review of Financial Studies 29*(1), 5–68.

Haugen, R. A. and N. L. Baker (1996, July). Commonality in the determinants of expected stock returns. *Journal of Financial Economics 41*(3), 401–439.

Heath, D. and C. Mace (2020, April). The Strategic Effects of Trademark Protection. *The Review of Financial Studies 33*(4), 1848–1877.

Hirshleifer, D., Kewei Hou, S. H. Teoh, and Yinglei Zhang (2004, December). Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics 38*, 297–331.

Holthausen, R. W. and D. F. Larcker (1992, June). The prediction of stock returns using financial statement information. *Journal of Accounting and Economics 15*(2), 373–411.

Hou, K., C. Xue, and L. Zhang (2015, March). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies 28*(3), 650–705.

Hovakimian, A., T. Opler, and S. Titman (2001). The Debt-Equity Choice. *The Journal of Financial and Quantitative Analysis 36*(1), 1–24.

Huang, A. G. (2009, June). The cross section of cashflow volatility and expected stock returns. *Journal of Empirical Finance 16*(3), 409–429.

Huang, R. D. and H. Jo (1995, September). Data frequency and the number of factors in stock returns. *Journal of Banking & Finance 19*(6), 987–1003.

Ikenberry, D., J. Lakonishok, and T. Vermaelen (2000). Stock Repurchases in Canada: Performance and Strategic Trading. *The Journal of Finance 55*(5), 2373–2397.

Ikenberry, D. L. and T. Vermaelen (1996). The Option to Repurchase Stock. *Financial Management 25*(4), 9–24.

Jegadeesh, N. and J. Livnat (2006, April). Revenue surprises and stock returns. *Journal of Accounting and Economics 41*(1), 147–171.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance 48*(1), 65–91.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (2021, January). Is There a Replication Crisis in Finance? SSRN Scholarly Paper ID 3774514, Social Science Research Network, Rochester, NY.

Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics 186*(2), 294–316.

Kelly, B. T., S. Pruitt, and Y. Su (2019, December). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kelly, B. T., S. Pruitt, and Y. Su (2020, December). Instrumented Principal Component Analysis.

Kim, J.-B., Y. Li, and L. Zhang (2011, June). Corporate tax avoidance and stock price crash risk: Firm-level analysis. *Journal of Financial Economics 100*(3), 639–662.

Koh, P.-S. and D. M. Reeb (2015, August). Missing R&D. *Journal of Accounting and Economics 60*(1), 73–94.

Kozak, S., S. Nagel, and S. Santosh (2020, February). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

Lakonishok, J., A. Shleifer, and R. W. Vishny (1994). Contrarian Investment, Extrapolation, and Risk. *The Journal of Finance 49*(5), 1541–1578.

Lamont, O. (1998). Earnings and Expected Returns. *The Journal of Finance 53*(5), 1563–1587.

Lang, M. H. and D. A. Shackelford (2000). Capitalization of capital gains taxes: Evidence from stock price reactions to the 1997 rate reduction. *Journal of Public Economics 76*(1), 69–85.

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance 56*(3), 815–849.

Lettau, M. and M. Pelger (2020). Factors That Fit the Time Series and Cross-Section of Stock Returns. *The Review of Financial Studies 33*(5), 2274–2325.

Lewellen, J. (1999, October). The time-series relations among expected return, risk, and book-to-market. *Journal of Financial Economics 54*(1), 5–43.

Li, F. (2011). Earnings Quality Based on Corporate Investment Decisions. *Journal of Accounting Research 49*(3), 721–752.

Liadaki, A. and C. Gaganis (2010, October). Efficiency and stock performance of EU banks: Is there a relationship? *Omega 38*(5), 254–259.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains From Diversification. *The Journal of Finance 20*(4), 587–615.

Litzenberger, R. H. and K. Ramaswamy (1979, June). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics 7*(2), 163–195.

Ludvigson, S. C. and S. Ng (2007). The empirical risk–return relation: A factor analysis approach. *Journal of Financial Economics 83*(1), 171–222.

Ludvigson, S. C. and S. Ng (2009). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

Lyandres, E., L. Sun, and L. Zhang (2008, November). The New Issues Puzzle: Testing the Investment-Based Explanation. *The Review of Financial Studies 21*(6), 2825–2855.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *The Journal of Finance 71*(1), 5–32.

Miller, M. H. and M. S. Scholes (1982). Dividends and Taxes: Some Empirical Evidence. *Journal of Political Economy 90*(6), 1118–1141.

Modigliani, F. and M. H. Miller (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review 48*(3), 261–297.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica 34*(4), 768–783.

Muradoğlu, Y. G. and S. Sivaprasad (2012, June). Capital structure and abnormal returns. *International Business Review 21*(3), 328–341.

Narayan, P. K. and S. S. Sharma (2015, May). Does data frequency matter for the impact of forward premium on spot exchange rate? *International Review of Financial Analysis 39*, 45–53.

Newey, W. K. and K. D. West (1987). Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review 28*(3), 777–787.

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

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research 18*(1), 109–131.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics 92*(4), 1004–1016.

Ortiz-Molina, H. and G. M. Phillips (2014, February). Real Asset Illiquidity and the Cost of Capital. *Journal of Financial and Quantitative Analysis 49*(1), 1–32.

Ou, J. A. and S. H. Penman (1989, November). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics 11*(4), 295–329.

Palazzo, B. (2012, April). Cash holdings, risk, and expected returns. *Journal of Financial Economics 104*(1), 162–185.

Patin, J.-C., M. Rahman, and M. Mustafa (2020, May). Impact of Total Asset Turnover Ratios on Equity Returns: Dynamic Panel Data Analyses. *Journal of Accounting, Business and Management (JABM) 27*(1), 19–29.

Paulo, A. (2018, September). Abnormal retained earnings around the world. *Journal of Multinational Financial Management 46*, 63–74.

Penman, S. H., S. A. Richardson, and İ. Tuna (2007). The Book-to-Price Effect in Stock Returns: Accounting for Leverage. *Journal of Accounting Research 45*(2), 427–467.

Pontiff, J. and L. D. Schall (1998, August). Book-to-market ratios as predictors of market returns. *Journal of Financial Economics 49*(2), 141–160.

Pontiff, J. and A. Woodgate (2008). Share Issuance and Cross-sectional Returns. *The Journal of Finance 63*(2), 921–945.

Pukthuanthong, K., R. Roll, J. L. Wang, and T. Zhang (2019, February). Testing Asset Pricing Model with Non-Traded Factors: A New Method to Resolve (Measurement/Econometric) Issues in Factor-Mimicking Portfolio. SSRN Scholarly Paper ID 3341604, Social Science Research Network, Rochester, NY.

Rapach, D. and G. Zhou (2021, January). Sparse Macro Factors.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010, February). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies 23*(2), 821–862.

Reinganum, M. R. (1981, March). Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics 9*(1), 19–46.

Richardson, S. A., R. G. Sloan, M. T. Soliman, and İ. Tuna (2005, September). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics 39*(3), 437–485.

Rosenberg, B., K. Reid, and R. Lanstein (1985, April). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management 11*(3), 9–16.

Ross, S. A. (1976, December). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk*. *The Journal of Finance 19*(3), 425–442.

Shumway, T. (1997). The delisting bias in CRSP data. *The Journal of Finance 52*(1), 327–340.

Sialm, C. (2009, September). Tax Changes and Asset Pricing. *American Economic Review 99*(4), 1356–1383.

Sikes, S. A. and R. E. Verrecchia (2012). Capital gains taxes and expected rates of return. *The Accounting Review 87*(3), 1067–1086.

Sivaprasad, S. and Y. G. Muradoglu (2009, November). Leverage,Stock Returns,Taxes and Industry Concentration.

Sloan, R. G. (1996). Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings? *The Accounting Review 71*(3), 289–315.

Soliman, M. T. (2008). The Use of DuPont Analysis by Market Participants. *The Accounting Review 83*(3), 823–853.

Spyrou, S. I. (2001, July). Stock returns and inflation: Evidence from an emerging market. *Applied Economics Letters 8*(7), 447–450.

Stattman, D. (1980). Book values and stock returns. *The Chicago MBA: A Journal of Selected Papers 4*, 25–45.

Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics 14*(1), 11–30.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association 97*(460), 1167–1179.

Swaminathan, S. and J. Weintrop (1991). The information content of earnings, revenues, and expenses. *Journal of Accounting Research 29*(2), 418–427.

Thomas, J. and F. X. Zhang (2011). Tax Expense Momentum. *Journal of Accounting Research 49*(3), 791–821.

Titman, S., K. C. J. Wei, and F. Xie (2004). Capital Investments and Stock Returns. *The Journal of Financial and Quantitative Analysis 39*(4), 677–700.

Uematsu, Y. and T. Yamagata (2020). Estimation of weak factor models. *ISER DP* (1053).

Valta, P. (2016, February). Strategic Default, Debt Structure, and Stock Returns. *Journal of Financial and Quantitative Analysis 51*(1), 197–229.

Vassalou, M. (2003, April). News related to future GDP growth as a risk factor in equity returns. *Journal of Financial Economics 68*(1), 47–73.

Vuolteenaho, T. (2002). What Drives Firm-Level Stock Returns? *The Journal of Finance 57*(1), 233–264.

Welc, J. (2017, August). Impact of Non-Controlling Interests on Reliability of Consolidated Income Statement and Consolidated Balance Sheet.

William C. Barbee, J., S. Mukherji, and G. A. Raines (1996). Do Sales-Price and Debt-Equity Explain Stock Returns Better than Book-Market and Firm Size? *Financial Analysts Journal 52*(2), 56–60.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate analysis*, 391–420.

Yook, K. C. (2010, August). Long-run stock performance following stock repurchases. *The Quarterly Review of Economics and Finance 50*(3), 323–331.