Three Essays on Market Anomalies and Financial Econometrics

by

Junyong Kim

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Management Science

at

The University of Wisconsin–Milwaukee

August 2020

Abstract

Three Essays on Market Anomalies and Financial Econometrics

by

Junyong Kim

The University of Wisconsin–Milwaukee, 2020

Under the Supervision of Professor Valeriy Sibilkov

This dissertation consists of two chapters about the momentum and idiosyncratic volatility anomalies, respectively, and one chapter about estimating clustered standard errors.

Chapter 1, Flights to Quality and Momentum Crashes, relates crashes of momentum strategies in stock markets around the world to investor behavior called flight to quality phenomena. The momentum crashes, defined as extremely negative returns of momentum portfolios, occur in most developed stock markets and are centered in economic recovery periods after recessions. I find that their negative returns and negative market betas are associated with investor behavior known as flights to quality (FTQ). Low quality—*i.e.*, high default risk—stocks experience larger investor withdrawals and consequential stock price plunges at financial market collapses, featuring higher market betas particularly during recessions. So the momentum strategies, which tend to sell these plunging stocks, exhibit negative market betas before their crashes and underperform once those stocks bounce back to an economic recovery phase. Worldwide momentum returns and two FTQ proxies, US institutional ownership changes and stock market-bond market disagreements, show consistent results.

Chapter 2, Which Volatility Drives the Anomaly? Cash Flow Versus Discount Rate, ex-

amines whether the cross-sectional idiosyncratic volatility anomaly is because of the volatility's cash flow news part or its discount rate news counterpart. In detail, I reexamine the idiosyncratic volatility anomaly of Ang et al. (2006) and investigate the relative importance of cash flow news and discount rate counterpart in driving this anomaly using the news decomposition of Vuolteenaho (2002). The results from idiosyncratic volatility-sorted portfolios show that the arbitrage portfolio with two extreme portfolios earns about 1.3 (1.2) percent quarterly alpha return after the market factor (the Fama–French factors). I also create two decile portfolios sorted on discount rate news volatilities and cash flow news counterparts. While the average return of the arbitrage portfolio from discount rate news volatilities is insignificant, the counterpart from cash flow news volatilities exhibits about 1.5 (1.2) percent quarterly alpha return after the market factor (the Fama–French factors). These findings indicate that cash flow news rather than discount rate counterpart governs most of the anomaly. The results suggest that investors prefer cash flow news volatilities to discount rate news counterparts, and hence not all idiosyncratic volatilities are equally priced in the cross-section.

Chapter 3, Multiway Clustered Standard Errors in Finite Samples, proposes new clustered standard errors less biased than existing clustered standard error estimators in finite samples. Specifically, I demonstrate the downward bias of existing one-way and two-way clustered standard error estimators (Petersen, 2009; Thompson, 2011) in finite samples using Monte Carlo simulations. When there exist both firm and time effects in a panel regression with $N \gg T$, a firm clustered standard error is always the worst. A clustered standard error estimator by time is the third best, but worsens as T increases. A clustered standard error estimator by both firm and time is the second best, but is biased downward in finite samples. I suggest two first best standard error estimators that always outperform the other competitors.

To my family

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express my sincere gratitude to Valeriy Sibilkov for his considerate guidance and cordial encouragement. His philosophy and attitude greatly influenced not only this dissertation but also my academic life.

I am deeply grateful to John R. Huck for his enlightening discussions and insightful comments. I sharpened and polished this dissertation tremendously based on his challenging questions and meticulous instructions.

I am very much indebted to Donghyun Kim for his continual support and intellectual inspiration. I cannot imagine completing this dissertation without the amount of time and effort he spent. I wish to extend my genuine appreciation to Chang-Mo Kang for his strong insights and helpful suggestions.

I truly appreciate Richard D. Marcus, Steven M. Trick, and Xiaoting Hao for their timely counseling and invaluable motivation. I was really fortunate to have such exceptional educators as my close mentors.

I would like to also acknowledge Ioannis V. Floros, Yong-Cheol Kim, Yao Li, Pengcheng Zhu, and G. Kevin Spellman for their detailed advice, and Dawn M. Koerton and James J. Hardy for their careful assistance. This dissertation benefitted a lot from their help.

I further wish to thank Jangsu Yoon and Ehsan S. Soofi for their untiring teaching and experienced consultation. Their immeasurable expertise and unmatched professionalism broadened my eyes to econometrics significantly.

My special thanks go to Xiaoqiong Wang, Chengcheng Li, Yi Liu, Siyuan Fan, Yangcheng Chen, and Donglai Bao for the friendship and enthusiasm they always gave. I owe Eun Jung Lee, Charles Kang, Yoonsun Jeong, and Jiwon Kim enough for keeping me company and cheering me up.

I must convey my greatest acknowledgement to my father Kyouk Kim, my mother Sungsin Cho, my brother Heeyong Kim, and my wife Youngeui Kim for their eternal love and infinite sacrifice. I would never have been able to accomplish anything without their generous tolerance and silent patience.

Chapter 1

Flights to Quality and Momentum Crashes

1.1 Introduction

Since the seminal work of Jegadeesh and Titman (1993a), momentum strategies have been considered to be one of the most well-known investment strategies and have been followed by money managers. Despite the momentum strategies' strong long-run returns and high Sharpe ratios, recent studies have reported episodes of momentum crashes, which refers to the strategies' extreme negative returns (Cooper et al. 2004, Stivers and Sun 2010, Barroso and Santa-Clara 2015, Ali et al. 2017). These momentum crashes, which tend to take place when economies recover from recessions, rarely occur but significantly damage the momentum portfolios. Finance literature introduces some methods to forecast or avoid the momentum crashes[1] but pays less attention to the crashes' economic origins. Daniel and Moskowitz (2016), for example, show that the momentum portfolios exhibit negative market

[1] Ehsani and Linnainmaa (2019) show that the momentum crashes take place when other factors become less autocorrelated. Novy-Marx (2015) shows that momentum strategies based on intermediate horizon past performance, unlike the strategies based on recent past performance, reduce the momentum crashes.

betas before these crashes but don't show what drives their negative betas.[2]

I examine whether these momentum crashes and that pre-crash beta behavior originate from investors' flights to quality (FTQ). FTQ, which refers to investors' shifting their portfolios to safer assets, is a well-documented phenomenon around financial market collapses.[3] When these extreme events happen, investors tend to hoard safe assets in fear of the worst case scenario (Caballero and Krishnamurthy 2008). Institutional investors also seek safe and liquid assets because of concerns of retail investors' withdrawals (Vayanos 2004), margin constraints (Krishnamurthy 2010), or lower capital and risk-bearing capacity (He and Krishnamurthy 2012).[4] FTQ affects not only capital markets but also real estate markets and economic growth (Boudry et al. 2019).

Why would investors' FTQ relate to momentum crashes? This prediction is motivated by the findings of Baele et al. (2019). They find that FTQ events, over short horizons, decrease equity prices relative to safer assets (*e.g.*, sovereign or corporate bonds) and increase expected returns of high default risk (*i.e.*, low quality) stocks relative to low default risk (high quality) stocks. These findings suggest that momentum portfolios, which buy past winner stocks and sell past loser stocks, effectively sell those low quality stocks around recessions. This tendency brings about the momentum crashes as the low quality stocks realize higher returns once the economy enters to a recovery phase. Furthermore, the low quality stocks, which the momentum portfolios sell, exhibit high market betas due to their procyclical returns, so the momentum portfolios feature negative market betas during recessions.

Figure 1.1 shows intuition for how the market beta of the momentum factor behaves differently from those of the other factors. I estimate monthly market betas of the momentum, size, value, profitability, and investment factors using daily data and plot their one-year mov-

[2] Momentum betas reflect past factor realizations in part by nature as these factors determine the momentum winners and losers (Kothari and Shanken 1992). However, the momentum betas that control these factors also show similar time-varying behavior (Blitz et al. 2011).

[3] Some examples include the market crash in 1987, the Russian default and sovereign debt crisis in 1998, and the global financial crisis in 2007.

[4] See also Barsky (1989) and Bekaert et al. (2009) who explain FTQ in consumption-based asset pricing models.

ing averages. The momentum factor's market beta, which signals the momentum crashes in turn, fluctuates more often than the other factors' market beta and becomes significantly negative especially around shaded recessions.[5] This pattern implies that, during those times, buying winners and selling losers coincides buying low betas and selling high betas. I examine if this coincidence is driven by the FTQ events and explore how these FTQ events affect momentum returns using worldwide stock returns.

This cross-country analysis is crucial for my research objective because both the momentum crashes and the FTQ events are respectively rare in one country, but at least the momentum profitability and the FTQ behavior are commonly found in many countries (Asness et al. 2013, Baele et al. 2019). Therefore, I first investigate whether the momentum crashes, as well as the momentum effects *per se*, are universal around the world, and then relate the FTQ events to these crashes. Moreover, the international data enable to further explore a number of novel research questions: Do these crashes show similar patterns in terms of timing and magnitude? Are the crashes proportional to the momentum profits across countries? Can investors also hedge the foreign crashes using the existing methods? Do the FTQ events lead to the momentum crashes in all developed markets?

I first test whether momentum crashes occur internationally. I find that momentum returns are negatively skewed in most countries. Among 23 MSCI developed countries, 20 countries exhibit negative skewness.[6] In particular, momentum exhibits its worst performance in the recovery period after recessions. During the recovery, winners underperform losers by a significant -1.28% per month. Momentum crashes during the recovery period are most severe in the United States, Canada, and Sweden (-5.13%, -4.23%, and -3.71% per month, respectively), whereas least severe in Israel (1.77% per month).[7] Unlike the momentum returns per se that differ by country (Griffin et al. 2003, Chui et al. 2010, Goyal and

[5]Though unreported, the market beta of the residual momentum factor (Blitz et al. 2011) shows similar behavior.

[6]Australia, Switzerland, and Portugal show positive skewness in the momentum return distributions.

[7]Israel, Denmark, and Austria do not experience momentum crashes in the recovery period.

Wahal 2015)[8], the momentum crashes are universal, which suggest the separate origins of the momentum returns and their crashes (Novy-Marx 2015, Goetzmann and Huang 2018).

Next, I estimate the time-varying market beta of momentum portfolios for each country conditioning on the recovery period. The estimation shows that momentum portfolios have a significantly negative market beta during the recovery, consistent with the US evidence of Daniel and Moskowitz (2016). The market beta of momentum returns during the recovery period is significantly lower by -0.630. Sweden has the largest discrepency (-1.439), whereas Norway has the smallest difference (-0.306). I also implement constant and dynamic volatility strategies widely accepted in recent literature (Moskowitz et al. 2012, Barroso and Santa-Clara 2015, Daniel and Moskowitz 2016, Hurst et al. 2017, Harvey et al. 2018). Both strategies change the weights of the portfolio over time by reducing the leverage of the strategies if the market is predicted to be volatile. These strategies work better than simple momentum strategies in most countries, increasing the average return by 0.41% per month. The strategies also generate much higher Sharpe ratios—0.44 and 0.48 per annum for the constant and dynamic strategies, respectively—than their pure counterparts—0.32 per annum.

The high beta of the losers in the recovery comes from two effects, changes in the composition of the portfolio, that is, more procyclical stocks in the portfolio, and changes in the beta of the constituent stocks. To disentangle the effects, I compare the market betas of momentum portfolios estimated in different windows: of (t-5, t+6), of (t-23, t-12), and of (t+13, t+24) months. During a recession, the market beta of the loser portfolio measured in (t-5, t+6) is higher than the beta of the same portfolio measured in (t-23, t-12) or (t+13, t+24), whereas the beta of the winner portfolio does not change by the estimation windows. The results indicate that the beta of the stocks in the loser portfolio increase temporarily. It is consistent with the hypothesis that the prices of the loser stocks temporarily drop due to FTQ and recovers when the economy improves.

[8] Asness et al. (2013) show that the momentum returns in Japan are positive and significant after accounting other pricing factors.

Furthermore, I investigate whether FTQ is related to a sudden fall in the stock prices of the losers and subsequent increase. To capture FTQ, I use fund flows of U.S. institutional investors.[9] From the panel regression, I find that institutional flows have a positive relationship with next quarter stock returns over the normal period, but they are negatively associated with next quarter stock returns over the recovery period. I disentangle the effects of outflows and inflows, and show that the outflows actually lead to a larger increase in stock returns over the next recovery period. In addition, I study the interaction between institutional flows and past stock returns and find that the increase in returns during the recovery is more pronounced for the loser stocks with institutional outflows. Next, I investigate whether the loser stocks during the recovery are those with more institutional outflows and higher default risks. I confirm that the losers indeed experience larger institutional outflows during the recession and have greater leverage during the recovery and therefore have a higher risk of default. My results indicate that the outperformance of the losers relative to the winners during the recovery is keenly related to the trading pattern of the investors.

Finally, following the recent FTQ literature (Baele et al. 2019, Boudry et al. 2019), I also identify both FTQ and flight to risk (FTR) events using both stock and bond market information rather than institutional investors behavior. Since the FTQ and FTR events identified by this method are short and infrequent, I estimate their effects to momentum returns using daily US momentum deciles as well as monthly momentum portfolios in other countries. Consistent with other findings, the results demonstrate that the momentum returns positively (negatively) react to FTQ (FTR) events and their reaction is greater to FTR rather than FTQ events, which are similar to the previous market rebounds.

The remainder of the paper is organized as follows. Section 1.2 describes the data, and Section 1.3 documents the international evidences of momentum crashes. Section 1.4 explores the relation between momentum crashes and FTQ. Section 1.5 concludes.

[9] U.S. institutional investors are less restricted in transferring funds from one country to another, so their tradings are similar to trading patterns of FTQ.

1.2 Data

1.2.1 Stock Returns

The returns and market capitalizations of international securities are obtained from Datastream. The intersection of the data contains 45,815 securities across 124 countries from December 1964 to July 2015. Bermuda, Cayman Islands, Côte d'Ivoire, Malawi, Monaco, Virgin Islands, Zambia, Canada, and United States are excluded. I include the observations if at least 10 securities are available for each country and month. From the data, I only keep the stocks identified by the Datastream type (EQ). I use the SEDOL codes to identify the securities and the Datastream return indices to compute the security returns. I obtain the securities' regional (headquarter) information through FactSet using the SEDOL codes.

Following the literature, I include large securities that comprise the first 90% of each country's market capitalization in each month. I adopt the 95% alternatively for Israel and Spain because the 90% threshold drops too many samples in late 1980s and early 1990s.[10] According to the classification of Morgan Stanley Capital International (MSCI), 79 countries are classified by developed, emerging, or frontiers. In detail, 21 countries are classified as developed markets, 25 countries are classified as emerging markets, and 33 countries are classified as frontier markets. In the main analysis, only the results of developed countries are reported, and the results of emerging and frontier countries are included in the appendix.

The return index and market value of each security are denominated in its home currency. The local return in month t is the difference between the return indices in t and t-1 divided by the return index in t-1. The value-weighted portfolio return in month t is based on each participating security's local market capitalization. I rank the securities each month by the (-12,-2) returns, the cumulative returns from t-12 to t-2, and form either the quintile portfolios in the countries with 300 or more securities available or the tercile portfolios otherwise.

In each country, the bear market indicator in month t is 1 if the excess market return

[10] The size distribution of Israel and Spain is quite positively skewed. The largest company in each country has a market share of nearly 40%.

from t-25 to t-1 is negative, and the down market indicator in month t is 1 if the one-month or two-month excess market return is negative. The backward, central, and forward betas of each security in month t are the simple regression estimates of the excess security returns on the excess market counterparts from t-24 to t-12, t-6 to t+6, and t+12 to t+24, respectively.

The leverage variables are from Datastream. I calculate the total liabilities to shareholders' equity ratio and the current liabilities to shareholders' equity counterpart as the default risk measures. I require shareholders' equity (WC03995) observations to have positive values but allow current liabilities (WC03101) observations with negative values. Each total liabilities observation is the difference between the total liabilities and shareholders' equity (WC03999) and the shareholders' equity (WC03995). I alter negative total liabilities observations by zero.

1.2.2 Currencies

The matching U.S. dollar return is based on the exchange rate return in month t. I collect the United States dollar exchange rates against other currencies from FactSet. I identify each country and each currency by the ISO 3166 and 4217 codes, respectively. I collect the one-month Treasury bill rates as the risk-free rates from the Kenneth R. French's website. Each excess return is the difference between the U.S. dollar return and the risk-free rate. Portfolios' total and excess returns are converted from local currencies to US dollars pursuant their respective definitions.

1.2.3 Institutional Investments

The institutional ownership data are from Global Ownership of Thomson Reuters. I consider observations with the country code 21 as the U.S. institutions and measure each security's ownership change in each quarter based on the SEDOL code. The institutional ownership change in quarter t is the aggregated value change from t-1 to t divided by the market capitalization in t. Unlike the returns above, I compute each institutional ownership change using the U.S. dollar market capitalization as well as the U.S. dollar value change.

1.2.4 Stock Market-Bond Market Disagreements

In addition to US institutional investment data, the FTQ and FTR events are identified following Baele et al. (2019) and Boudry et al. (2019). Since this literature employs stock and bond market return data around the world, I use daily total market indices (TOTMK) and benchmark 10 year government bond indices (BM10) from Datastream for stock and bond market returns, respectively. Due to their unavailability issue, I exclude Hong Kong and Israel from this analysis.

1.3 International Momentum Crashes

Summary statistics from the 21 countries categorized as developed by Morgan Stanley Capital International (MSCI) are presented in Table 1.1 . I sort the countries in the first column based on the numbers of the stocks available in my data. I append a dagger (†) to a country with less than 300 securities. Each country, I report the beginning and ending dates in the second and third columns, respectively. For example, the data from the United Kingdom and Ireland start early in January 1966, but those from Spain and Portugal start late in April 1988 and September 1989, respectively. Each in the fourth and fifth columns exhibits the time-series average number of available stocks and the number of available months by country.[11] Following Fama and French (2017), I only consider large stocks that account for up to 90% of the country's total market capital each month.[12] For Israel and Spain, where the largest company accounts for nearly 50% of the country's total market capitalization, I apply the threshold of 95% instead of 90%, as the 90% rule drops too many small-caps in

[11]For example, in Japan, there are 498 months, with an average of more than 554 securities per month. In Portugal, there exist 294 months, with less than 10 securities per month on average.

[12]For example, in Japan, there exist 752 stocks available at the beginning of February 1974. According to the rules, I only include the largest 376 securities, with a total capital of 29.12 trillion yen, because they account for 90% of the total capital of 32.37 trillion yen.

the late 1980s and early 1990s. The sixth to ninth columns present the sample means, standard deviations, skewness estimates, and annualized Sharpe ratios of the countries' monthly excess market returns. Each month, I calculate the value-weighted average market return of each country, convert the average return to the dollar return based on the percentage change of the exchange rate, and calculate the excess market return based on a one month U.S. Treasury bill rate. The average excess market return is between 0.213% and 1.041%. Hong Kong has the highest market return (1.04%), whereas Portugal has the lowest market return (0.213%) among developed countries. The average standard deviation of excess market return for developed countries is 6.952%. The excess returns show a positive skewness in the United Kingdom (0.95) and Singapore (0.81), but a negative skewness in Norway (-0.56) and Australia (-0.72). The Sharpe ratios are relatively high in Switzerland (0.51), Sweden (0.45), and Hong Kong (0.40), but relatively low in Austria (0.16), Italy (0.15), and Portugal (0.10).

The last seven columns in Table 1.1 labeled momentum(-12,-2) display the sample means, standard deviations, and skewness estimates of the countries' winner-minus-loser (WML) portfolios, and then the average returns and annualized Sharpe ratios of the countries' winner and loser portfolios. For each country, I form the value-weighted quintile or tercile (†) portfolios based on the returns from t-12 to t-2 of the sample stocks, rebalance them each month, and adopt the highest and lowest portfolios with past returns as the winner and loser portfolios following Daniel and Moskowitz (2016). Consistent with the international momentum literature, the average monthly returns of WML portfolios are positive in all countries except Spain. Instead, the average momentum return in Spain is positive for the equal-weighted momentum strategy without sample selection.[13] The average returns of the momentum portfolios in Israel (1.69%) and Denmark (1.32%) are relatively high, but those in Japan (0.19%) and Belgium (0.0009%) are relatively low.[14] The WML portfolios in the

[13]Without sample screening, the mean, standard deviation, and skewness of WML portfolio of Spain are 0.64%, 5.83%, and -1.08, respectively.

[14]In Belgium, the average return of the momentum portfolio becomes much higher (0.77%) if I use the equal-weighted tercile portfolios and do not exclude any stocks based on the market capital.

United Kingdom (5.91%) and Australia (6.15%) are riskier than those in Sweden (11.63%) and Israel (14.26%). In my sample periods, the value-weighted momentum strategies show a significance greater than 10% for 14 out of the 21 countries. As documented in the previous literature, momentum strategies are profitable in most countries.

It is worth noting that 18 of the 21 skewness estimates of momentum portfolios are negative, indicating that the momentum strategy is embedded in the possibility of a crash. The skewness estimates are positive in Australia (0.08), Portugal (0.03), and Switzerland (0.03), but are very negative in Singapore (-1.16), Sweden (-1.47), and Spain (-2.16). The correlation between the means and skewness of the momentum is 0.51. It shows that the occasional momentum collapse seriously affects the average of momentum gains.

Following Daniel and Moskowitz (2016), if the excess market returns in the previous two years are negative, I define time as a bear market, otherwise it will be a normal market. In the bear market, I divide time into up and down markets. If the excess market return for the current month is negative (positive), then it is a down (up) market. Table 1.2 displays the average returns of the market, winner, loser, and WML portfolios in the developed countries during the normal, bear-down, and bear-up markets. For each country in the first column, I report the average returns of the market, winner, loser, and WML portfolios during the normal period in Panel A. For each WML portfolio, I also report the t-statistic and the number of observations in the sixth and seventh columns respectively. Panel B and Panel C present the same statistics during the bear-down and the bear-up markets respectively. The bottom row exhibits the averages of the average returns and their t-statistics.

Consistent with the momentum literature, the average returns of momentum portfolios are positive in non-bear markets (0.38%) and bear markets (0.42%). In detail, the average returns during the bear-down and bear-up markets are asymmetric. In the bear market, the sample mean is largely positive (3.36%), whereas in the bear market it is negative (-1.17%). However, the bear-up months (101 months on average) are about twice more often the bear-down months (55 months on average). The average returns during the bear-down markets are positive in all countries except Israel, and those during the bear-up markets are negative in the most countries, 18 out of 21 countries. In 16 out of 21 t-statistics during the bear-down markets are greater than 1.64 and 7 out of 21 t-statistics during the bear-up markets are less than -1.64. The results show that the momentum strategy is profitable in normal period and at the beginning of recession periods, but in the recovery period, the momentum returns are globally negative. Based on the findings that momentum crashes occur during a given period, the momentum crash seems to be a systematic problem. This suggests that the stocks included mechanically in the momentum portfolio during the recovery are systematically different from those of other periods.

The average returns during the bear-down markets are relatively high in Sweden (8.53%), Norway (6.29%), and Hong Kong (4.91%), but relatively low in Australia (0.72%), Austria (0.32%), and Israel (-0.32%). Likewise, the average returns during the bear-up markets are relatively low in Sweden (-3.71%), Singapore (-3.52%), and Switzerland (-2.65%), but relatively high in Denmark (0.16%), Austria (0.48%), and Israel (1.77%). The countries with the wide gaps are the countries with the high excess market returns. The average market returns in Hong Kong, Sweden, Singapore, Switzerland, and Norway are 1.04%, 0.99%, 0.76%, 0.75%, and 0.73%, respectively. The correlation coefficient between the WML gaps during the bear-down and bear-up markets and the excess market returns are 0.58.

Table 1.3 displays the market timing regression estimates and the corresponding t-statistics from the developed countries. Following Daniel and Moskowitz (2016), I consider (1) a regression with one alpha and three betas during non-bear, bear-up, and bear-down markets; (2) a simple market regression with one alpha and one beta; (3) a regression with two alphas and two betas during non-bear and bear markets respectively; and (4) a regression with two alphas during non-bear and bear markets, and three betas during non-bear, bear, bear-up markets. I report the estimates of the first regression obtained from the developed countries in Table 1.3, and those of the other regression specifications and those obtained from the emerging and unclassified countries in Appendix. For each country in the first column, I report the estimates (t-statistics) of the alpha, normal beta, bear-down beta, and bear-up beta in the even (odd) columns, respectively. The bottom row exhibits the averages of the countries' regression estimates and their t-statistics.

Consistent with the literature, the alpha estimates are positive in all countries but Belgium. In Belgium, alternatively, the alpha estimate of the equal-weighted counterpart is also positive (0.90%) and statistically significant at 1% level. The estimates are relatively high in Israel (2.98%), Portugal (2.57%), and Denmark (1.91%), but relatively low in Japan (0.31%), Spain (0.27%), and Belgium (-0.02%). Fifteen of 21 t-statistics are greater than 1.64.[15]

Non-bear beta estimates are positive in 18 of the 21 countries and significant in 11 countries. The estimates are relatively high in Japan (0.40), Portugal (0.31), and Israel (0.27), but relatively low in Germany (0.06), New Zealand (0.04), and the United Kingdom (0.02). The estimates are negative in France (-0.004), Spain (-0.007), and Italy (-0.08), but insignificant. The average of the estimates is 0.14.[16] I confirm that the non-bear market beta of WML is negative in the U.S., but positive outside the U.S., consistent with the literature (Novy-Marx 2012, Daniel and Moskowitz 2016).

The bear-down beta estimates are negative in all countries but Israel and significant in 15 out of 21 countries. The bear-up beta estimates are negative and significant in all countries. Note that the bear-up beta captures the time-varying amount of beta from the bechmark beta, that is, non-bear beta. In all countries, the sum of non-bear beta and bear-up beta is also siginficantly negative, suggesting that momentum portfolios have negative market betas in the recovery. The averages of the bear-down beta estimates and the bear-up counterparts are -0.45 and -0.72, respectively. The results are consistent with (a) Grundy and Martin (2001) as the betas are time-varying and negative during bear markets, (b) Cooper et al. (2004) as the WML portfolios underperform during bear markets, and (c) Daniel and

[15] In the appendix, the average alphas from the regression models (2), (3), and (4) are 0.89%, 0.92%, and 0.92%, respectively.

[16] The averages from the regressions (2), (3), and (4) above are -0.14, 0.14, and 0.14, respectively.

Moskowitz (2016) as there exists the optionality in the upside and downside beta estimates during the bear markets. The bear-up beta estimates are greater in magnitude than the bear-down counterparts in 17 out of 21 countries by 0.28 on average, and the F-statistics about their respective differences are significant in 7 out of 21 countries.[17]

There is an interesting relationship between the alphas and the difference between the bear-down beta and bear-up beta. The alphas are relatively high in the countries where the differences between betas are greater in magnitude. In other words, the market beta of momentum portfolio is more time-varying in the countries with the momentum profits. It suggests that in countries with momentum profits, momentum crash is more likely to occur during the recovery periods. The differences between the bear-down and bear-up betas in Israel, Portugal, and Denmark are -1.38, -0.85, and -0.59, respectively. The correlation coefficient between the alphas and the differences is -0.85. The same coefficient between the bear (non-bear) alphas and the bear-beta differences is -0.86 (-0.39).

The time-varying beta of momentum portfolio can be dissected by the beta changes of winners and losers. It can be further divided into a time-varying composition of the portfolio and a time-varying betas of constituent stocks. We examine the channels of time-varying beta of momentum portfolio in Table 1.4 by looking at the behavior of the central, backward, and forward 12-month betas from the winner, loser, and WML portfolios during the non-bear, bear-up, and bear-down markets. I obtain the equal-weighted betas of the countries' winner, loser, and WML portfolios based on the betas of the constituent stocks each month and report their time-series averages during the non-bear, bear-down, and bear-up markets. I regress the securities' returns on the corresponding market returns to acquire the stocks' market betas. In each month t, I estimate (i) the central betas using the monthly returns from t-5 to t+6, (ii) the backward betas using the monthly returns from t-23 to t-12, and (iii) the forward betas using the monthly returns from t+13 to t+24.

[17]The bear beta estimates from the regressions (3) and (4) are significant in all countries and 16 out of 21 countries respectively. The averages of the bear beta estimates from the regressions (3) and (4) are -0.61 and -0.40 respectively.

In Panel A, I present the central betas for each country. The backward betas and the forward betas are reported in Panel B and Panel C, respectively. Each panel contains the averages of the winner, loser, and WML betas during the non-bear, bear-down, and bear-up markets, respectively. The asterisks present the significance of the averages from the respective WML portfolios. The results with the central window in Panel A are consistent with the results in Table 1.3. The averages of the WML portfolios' equal-weighted betas during the non-bear markets are positive in all countries but New Zealand and significant in 16 out of 21 countries. The averages during the bear-down and bear-up markets are negative and significant in all countries. The non-bear, bear-down, and bear-up averages of the respective 21 WML averages are 0.14, -0.39, and -0.36. Both beta changes in winners and losers seem to contribute beta changes in WML portfolio. During the recovery period, the loser's beta change is more pronounced, at 0.3, whereas the winner's beta decreases by 0.2.

The results with the backward beta in Panel B are also consistent as the non-bear betas are positive in 16 out of 21 countries and the bear betas in all countries except the bear-up beta in Switzerland are negative. Unlike the central betas, however, the backward counterparts exhibit a relatively weaker pattern. Fifty-five out of 63 backward WML betas are less in magnitude by 0.15 on average than the central counterparts. The non-bear, bear-down, and bear-up averages of the respective 21 WML averages are 0.03, -0.23, and -0.16. Likewise, the pattern of the forward betas in Panel C is relatively weaker than that in the central betas. Forty-seven out of 63 forward WML betas are less in magnitude by 0.10 on average than the central counterparts. The non-bear, bear-down, and bear-up averages of the respective 21 WML averages are 0.08, -0.29, and -0.20.

It is worth noting that during a recession, central beta of the losers (winners) are higher (less) than backward and forward betas of the same portfolio. The loser's market beta has increased over the past year before being included in the loser portfolio, and after the formation period, the market beta has declined. It can be inferred that the increase in the beta of losers during this period is mainly due to changes in the beta of individual stocks rather than changes in composition. If the loser's stock price temporarily falls more than other stocks and then rebounds more, then the beta change pattern can be similar. The time-varying beta of the losers is due to the capture of this fluctuation.

Table 1.5 displays the returns and Sharpe ratios of (i) the plain WML strategy, (ii) the constant volatility strategy, and (iii) the dynamic volatility strategy by country. Following Daniel and Moskowitz (2016), I estimate the full-sample GJR-GARCH(1,1)-M with each country's WML portfolio and scale the portfolio based on the mean and variance estimates to implement the constant and dynamic volatility strategies.[18] I adjust the time-invariant scalars in the scaling factors of the constant and dynamic strategies to equate the strategies' full-sample volatilities and the plain strategy's counterpart.

For each country in the first column, I report the average returns of the plain, constant, and dynamic strategies in the second, third, and fourth columns, respectively. The brackets below the averages display the annualized Sharpe ratios. I also report the GARCH estimates of each country in the fifth to tenth columns. The parentheses below the estimates exhibit the t-statistics. The bottom row presents the averages of the countries' estimates and their t-statistics.

The results in Table 1.5 are consistent with not only Barroso and Santa-Clara (2015) as the constant volatility strategies outperform the plain WML strategies in all countries, but also Daniel and Moskowitz (2016) as the dynamic volatility strategies outperform the constant volatility strategies in all countries but Finland. The return improvements from the plain strategies to the constant counterparts are relatively high in Spain (0.75%), Israel (0.71%), and Sweden (0.61%), but relatively low in New Zealand (0.07%), Norway (0.06%), and Finland (0.03%). The return improvements from the constant strategies to the dynamic counterparts are relatively high in Spain (1.22%), Singapore (0.64%), and Belgium

[18] As aforementioned in Daniel and Moskowitz (2016), these GARCH-based strategies are not implementable in real time. Some strategies using trailing volatilities allow real time implementation but require daily returns as well as monthly returns (Moreira and Muir 2017, Harvey et al. 2018).

(0.58%), but relatively low in Norway (0.03%), New Zealand (0.02%), and Sweden (0.01%), and negative in Finland (-0.03%). The average returns (Sharpe ratios) of the countries' plain, constant, and dynamic strategies are 0.77% (0.30), 1.06% (0.41), and 1.31% (0.51), respectively. The improvements by the constant strategies are not in some countries, but those by the dynamic counterparts are in most countries. The correlation coefficients of the plain strategies' averages with the improvements by the constant and dynamic strategies are -0.03 and -0.57, respectively.

The results in the 8th column are also consistent with the momentum literature. The asymmetry parameters are negative in all countries but Germany and significant in 18 out of 21 countries. The pattern suggests the momentum crashes as the volatility react strongly to the positive shocks and weakly to the negative counterparts. Unlike the literature, however, the feedback parameters in the 10th column are significant in 3 out of 21 countries, though their average is -0.59.

The unreported alphas of the plain strategies are positive in all countries but Spain (-0.16%) and significant in 17 out of 21 countries. The alphas of the constant strategies are positive in all countries including Spain. They are significant in 18 out of 21 countries. The alphas of the dynamic strategies are positive and significant in all countries. The averages from the unreported alphas of the plain, constant, and dynamic strategies are 0.85%, 1.13%, and 1.34%, respectively. Likewise, I examine whether the plain strategies span the constant and dynamic strategies based on the regression alphas of the constant and dynamic strategies on the plain counterparts. The unreported alphas of the constant and dynamic strategies after the plain counterparts are significant in 17 out 21 countries respectively. The averages of the unreported alphas from the constant and dynamic strategies are 0.34% and 0.74% respectively.

In addition, the averages of the plain, constant, and dynamic strategies' unreported skewness estimates are -0.54, -0.24, and 0.36, respectively. Consistent with Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016), the results indicate that the constant and dynamic volatility strategies mitigate the momentum crashes of the plain WML strategies and improve them significantly in the international markets. The results also prove that the performance improvements by the dynamic volatility strategies are more effective and stable than those by the constant counterparts.

1.4 Flights to Quality and Momentum Crashes

1.4.1 FTQs by Institutional Investments

In Table 1.6 panel A, I investigate the stock return consequences of flight-to-quality. FTQ is proxied by ownership changes of U.S. financial institutions. The rationale for this proxy is that U.S. institutions are less constrained to withdraw their money from one country, their investment behavior would resemble the investment pattern of FTQ investors. Due to the filing frequency of ownership data, I use quarterly returns and quarterly holding changes of U.S. institutions. I run the following panel regression:

$$R_{cit+1} = \alpha_t + \left(\beta + \left(\gamma + \delta I_{Uct+1}\right) I_{Bct}\right) \Delta IO_{it} + \varepsilon_{cit+1}, \tag{1.1}$$

where R is the dollar return, I_B is the bear market indicator, I_U is the up market indicator, and ΔIO is the institutional ownership change at the stock level. For the up market indicator (I_{Uct+1}), I use the beginning month of the quarter (I_{U1}) or the same quarter $\left(I_{U(1,3)}\right)$ relative to the quarter of stock return. The subscripts c, i, and t indicate each country, stock, and quarter, respectively. In the regression, I include time fixed effects and cluster standard errors by country. Column 1 shows that changes in U.S. institutional ownership is positively and significantly associated with the next quarter stock returns. It is consistent with the literature of smart money effect in foreign equity sector by U.S. institutions. I decompose the ownership changes by positive and negative changes. This decomposititon allows me to examine whether return predictability comes from inflows, outflows, or both. The result in column 2 shows that outflows predominantly predict returns, but not inflows. When U.S. institutions collectively sell, then the returns of the stocks are lower in the next quarter. In columns 3–8, I divide the periods to distinguish the effect of institutional holding changes over different periods. In columns 5 and 7, I find that the stock return becomes opposite to the direction of institutional flows in the next quarter if the next quarter is up market during the recession. The results in columns 6 and 8 indicate that the reversal during bear-up market period is concentrated in stocks with institutional outflows.

In panel B of Table 1.6, I examine the interaction effect of institutional flows and momentum. In the previous section, I show that momentum crash is due to the extremely high return of losers in the recovery. The results in columns 6 and 8 indicate that there is an interaction between momentum and FTQ. The returns of the losers, which also experienced U.S. institutional outflows, are particularly higher during the bear-up market period. It is consistent that the reversal of the losers are associated with the flight-to-quaility and the subsequent flight-from-quality, that is, investors over-sell risky stocks at high risk of default, but the price declines of losers rebound rapidly during the market recovery phase, when market-wide fears about default risks are resolved.

Table 1.7 displays changes in institutional ownership prior to the formation of the winner and loser portfolios during the non-bear, bear-down, and bear-up markets. I obtain the institutional ownership changes of the securities from Global Ownership and compute the value-weighted averages of the winner and loser portfolios. Changes in ownership of the winner and loser portfolios are calculated monthly, even if institutional ownership changes are quarterly, as the portfolio is rebalanced each month. For example, value-weighted averages for October, November and December are derived from changes in institutional ownership from the previous quarter, from July through the end of September. I winsorize the ownership changes at the 0.5th and 99.5th percentiles by country and months.

For each country in the first column, I report the means of the monthly value-weighted average institutional ownership changes based on the previous quarter data (ΔIO from t-2 to t-1) in Panel A, based on the data from two quarters before (ΔIO from t-3 to t-2) in Panel B, and based on the previous two quarter data (ΔIO from t-3 to t-1) in Panel C, respectively. Each Panel of 9 columns exhibits the winner and loser portfolios' means and their differences during the non-bear, bear-down, and bear-up markets respectively. The bottom row presents the averages of the countries' estimates and their t-statistics.

The results indicate that the institutional investors increase their ownership of the winner securities more and the ownership of the loser counterparts less on average. In panel A, the differences between the winner and loser average ownership changes are positive in 16 out of 21 countries during the non-bear, bear-down, and bear-up markets, respectively. In panel B, the differences are positive in 16, 13, and 18 countries during the non-bear, bear-down, and bear-up markets, respectively. In panel C, the differences are positive in 17, 16, and 18 countries during the non-bear, bear-down, and bear-up markets, respectively. The averages of the differences during the non-bear, bear-up, and bear-down markets are 0.10, 0.19, and 0.18 in panel A, 0.13, -0.01, and 0.19 in panel B, and 0.22, 0.00, and 0.30 in panel C, respectively. This evidence suggests that the institutional investors prefer the momentum strategy to the reversal counterpart.

The results also indicate that the institutional investors increase their ownership more during the non-bear markets and less during the bear-down and bear-up markets on average. In panel C, the averages from the countries' winner portfolios during the non-bear, bear-down, and bear-up markets are 0.55, 0.02, and 0.17, respectively. Likewise, the same averages from the countries' loser counterparts are 0.33, 0.02, and -0.14, respectively. This evidence suggests that the institutions relatively inject capital into the international markets prior to the non-bear markets and relatively eject from the markets prior to the bear-down and bear-up markets.

An interesting pattern exists between the countries' average WML returns and the institutional ownership changes. The correlation coefficients of the average returns with the differences between the winner and loser portfolios' ownership changes during the non-bear,

bear-down, and bear-up markets are 0.12, 0.32, and 0.60, respectively. The same coefficients with the central (backward) betas in Table 1.4 are 0.08 (0.28), 0.11 (-0.54), and 0.48 (-0.25), respectively. The results suggest that the momentum strategies by the institutional investors are effective as the returns of the WML portfolios are proportional to the differences between the winner and loser portfolios' ownership changes.

Table 1.8 exhibits the winner and loser portfolios' equal-weighted averages of the book leverages with the total and current liabilities respectively during the non-bear, bear-down, and bear-up markets. I acquire the current liabilities, total liabilities, and shareholders' equity of each participating security from Datastream and compute the current liabilities-to-shareholders' equity ratio and the total liabilities-to-shareholders' equity counterpart as the default risk measures. I estimate the monthly equal-weighted averages of the winner and loser portfolios with the annual balance sheet items as I rebalance the portfolios each month. For example, I calculate the monthly equal-weighted averages in year t with the leverage ratios based on the accounting data in year t-1. I exclude observations with negative shareholders' equity values and alter negative total liabilities values by 0s. I winsorize the ratios at the 99th percentile by country and year.

For each country in the first column, I report the means from the monthly equal-weighted averages with the total liabilities-to-shareholders' equity ratios of the winner and loser portfolios in the second to tenth columns, and the same means with the current liabilities-to-shareholders' equity ratios in the eleventh to nineteenth columns, respectively. Each partition of nine columns display the winner and loser portfolios' means and their differences during the non-bear, bear-down, and bear-up markets. The bottom row exhibits the averages of the countries' estimates and their t-statistics.

The results present that the differences between the leverage ratios of the winner and loser portfolios increase (decrease) prior to the non-bear (bear-down and bear-up) markets on average. In panel A, 13 (8) out of 21 differences are positive (significant) during the non-bear markets and the average is 0.19. In contrast, 15 (6) differences are negative (significant) during the bear-down markets and 14 (9) differences are negative (significant) during the bear-up markets. The averages during the bear-down and bear-up markets are -1.02 and -1.03 respectively. In panel B, likewise, 13 (8) out of 21 differences are positive (significant) during the non-bear markets, but 14 (9) differences are negative (significant) during the bear-down markets and 14 (10) differences are negative (significant) during the bear-up markets. The averages during the non-bear, bear-down, and bear-up markets are 0.02, -0.24, and -0.22, respectively. The correlation coefficient between the WML returns during all the periods and the leverage ratios with the total liabilities (current liabilities) is 0.29 (0.16). The unreported results with the value-weighted leverage ratios are also consistent with the results in Table 1.8. This evidence demonstrates that the default risk measures explain the time-varying returns of the WML portfolios in part as the leverage ratios allegedly proxy the default probabilities.

One characteristic of flight-to-quality events is the enhanced negative correlation between the stock and bond markets (Baele et al. 2019). If the flight-to-quality events affect these two markets more during the bear markets than the bull markets, then the correlation between the stock and bond markets will decrease. Following Baele et al. (2019), I investigate whether the daily return correlations between the stock and bond markets in the countries decrease during the bear markets. Table 1.9 exhibits the daily return performance of the stock and bond markets in the MSCI developed markets except the United States, Canada, Hong Kong, and Israel.

Among 19 countries, the correlation between the stock and bond markets is negative in 14 countries and significant in 10 countries during the bull markets, while the correlation is positive in 5 countries and significant in Italy, Ireland, Spain and Portugal. During the bear-down markets, in contrast, the correlation is negative and significant in 11 countries, but positive and significant in Italy and Portugal. The positive and significant correlations in Spain and Ireland become insignificant and negative, respectively. In detail, during the bear-down markets, the correlation drops in 16 countries and the decreases are significant in 8 countries. Similarly, during the bear-up markets, the correlation drops in 15 countries but significant in only 4 countries. This is consistent with the literature as the flight-to-quality events are more likely during the bear-down markets according to the definition of the flight-to-quality. Consistent with Baele et al. (2019), these findings imply that the flight-to-quality episodes affect both the stock and bond markets in the countries during the bear markets, especially during the bear-down markets, and that the bear market indicator variables capture the flight-to-quality effects consistently.

1.4.2 FTQs by Stock Market-Bond Market Disagreements

FTQ behavior by investor is informative but difficult to directly observe. Researchers, therefore, also use market information to indirectly identify FTQ events and measure their effects (Baele et al., 2019; Boudry et al., 2019). They identify the FTQ events as the periods with large negative stock market returns and large positive bond market returns together. Since these events are short and infrequent, the researchers examine daily data from stock and bond markets.

Following this literature, I examine how FTQ affects momentum crashes using the indirect method as well. Following Baele et al. (2019), I compute daily stock and bond market returns from US Datastream market total return index (TOTMKUS) and US benchmark 10 year Datastream government bond total return index (BMUS10Y), respectively. Daily FTQ events are identified as follows.

$$FTQ_t = I\left\{r_t^b > \kappa\sigma_t^b\right\} \times I\left\{r_t^e < -\kappa\sigma_t^e\right\},$$

where r_t^b and r_t^e are the daily bond and stock market returns, respectively. σ_t^b and σ_t^e are the second moments of these returns estimated using a one-sided normal kernel with a bandwidth of 250 days skipping the nearest 5 days to avoid any influence of FTQ and FTR events. κ is 1.5 following Boudry et al. (2019) while the results are robust regardless of the choice.

While both FTQ and flight to risk (FTR) affect momentum returns, momentum crashes are affected more by FTR events because these crashes often occur during market recoveries. Therefore, I identify FTR events in addition to the FTQ counterparts using the opposite inequalities. That is, the FTR events are identified as the periods with large negative bond market returns and large positive stock market returns together.

Figure 1.3 shows the annual distribution of FTQ and FTR days identified using the stock and bond market returns from the 21 countries. The results confirm that the FTQ and FTR days tend to cluster over time. For example, there were 33 FTQ days in the United States and 387 FTQ days outside the United States in 2007 and 2008. These FTQ days account for 29.2% and 26.7% of all the FTQ days identified in and outside the United States, respectively. Likewise, 19.7% of FTR days in the United States and 28.2% of FTR days outside the United States are concentrated in 2007 and 2008.

Though untabulated, the unconditional likelihood of the identified FTQ events is 1.13% and their estimated daily price impact is 4.01%. These results are consistent with the 1.74% likelihood of Baele et al. (2019) and the 1.54% likelihood of Boudry et al. (2019) but lower than these two. Similarly, the unconditional likelihood of the identified FTR events is 0.76% and their estimated daily price impact is 3.69%.

Since these FTQ and FTR events are identified at a monthly frequency, I first examine daily momentum returns. Due to the lack of daily momentum return data by country, I investigate US daily momentum deciles. After than, I generalize the definition of the FTQ and FTR events to incorporate monthly momentum returns around the world. Using these events, I estimate the following regression.

$$R_t^e = \alpha + \beta^\top f_t + \gamma FTQ_t + \beta_Q FTQ_t f_t^{MKTRF} + \delta FTR_t + \beta_R FTR_t f_t^{MKTRF} + \varepsilon_t,$$

where R_t^e is a daily excess return of each momentum decile or a daily winner-minus-loser (WML) return from the deciles, and f_t is a vector of pricing factors (Fama and French, 2015).

I estimate γ and δ with and without these factors, and then test whether γ, δ, β_Q, and β_R are significant.

Table 1.10 exhibits the regression results. The first partition displays the results without the pricing factors. The intercept of the daily WML returns is 4.8 basis points (12.1% per annum) and its t-statistics is 2.86. Consistent with previous results, the γ and δ estimates are 0.9% and -1.3% and their t-statistics are 2.60 and -3.32, respectively. The results show that the WML portfolio experiences positive returns during FTQ events and negative returns during FTR events, and that the magnitude of the negative returns is greater than that of the positive returns. The unreported Wald statistic that tests their difference is 13.82.

The second and third partitions display the results with the pricing factors. The intercept after controlling the factors is 6.2 basis points (12.1% per annum) and its t-statistics is 4.12. Following the literature, the daily WML returns exhibit negative loadings to the market, size, and value factors but positive loadings to the profitability and investment factors, respectively. Furthermore, consistent with previous results, the β_R is negative and significant. That is, the WML returns react more negatively to the market factor during FTR events.

Similar to Boudry et al. (2019), I also examine how FTQ and FTR events affect momentum returns using monthly FTQ and FTR variables as well as monthly momentum returns around the world. After excluding the United States, Canada, Hong Kong, and Israel, I compute the daily FTQ and FTR variables of the 19 MSCI developed countries using their stock (TOTMK) and bond (BM10Y) market returns (Baele et al., 2019) and examine whether there are FTQ and FTR events occurred in each month to incorporate monthly information. Monthly FTQ (FTR) variables are one in month t if one or more FTQ (FTR) events identified in month t. Similar to the previous regression, both γ and δ are estimated with monthly WML returns by country but without other factors.

Table 1.11 exhibits the regression results. The intercept is positive in all countries but Belgium and significant in 5 out of 19 countries. When all countries are used, it is positive ($\alpha = 0.008$) and significant ($t = 6.52$). The γ for the FTQ events is positive (significant) in 14

(6) out of 19 countries. When all countries are used, it is positive ($\gamma = 0.012$) and significant ($t = 3.67$). The δ for the FTR events is negative in 15 out of 19 countries and significant in France, Norway, and Austria. When all countries are used, it is negative ($\delta = -0.015$) and significant ($t = 4.12$). The results are also consistent when NFTQ and NFTR replace FTQ and FTR, respectively. Consistent with previous findings, these results show that the WML returns react positively (negatively) to FTQ (FTR) events, and that the magnitude of the return reaction to FTR events is greater than its counterpart to FTQ events.

1.5 Conclusion

Many papers have explored the momentum effect and the momentum crashes. These papers are also suggesting some successful trading methods that improve the static momentum portfolios, which are vulnerable to those momentum crashes. In contrast, the economic reason of these momentum crashes has drawn relatively less attention from those papers. I examine the momentum crashes in international stock markets and introduce flights to quality in financial markets to help explain the way the momentum crashes take place. While the flights to quality are being studied actively, I relate them to the momentum crashes to better understand the cross-sectional phenomena.

Like the momentum effect per se, the momentum crashes are universal in the international security markets, as well as the United States market. Alongside the positive and economically significant Sharpe ratios of the momentum portfolios around the world, their skewness estimates are negative and economically significant as well on average. In particular, the crashes are concentrated during the bear-up rather than non-bear and bear-down markets, and the tendencies are captured by the market timing regressions with the respective indicators. The time-varying betas explain the crashes in part as the betas are weakly positive during the non-bear markets and strongly negative during the bear markets, and the bear-up betas are greater in magnitude than the bear-down counterparts. These patterns are clear with the central betas, but unclear with the backward and forward betas.

The momentum crashes are also noticeable in the countries where the momentum performance is insignificant. The constant and dynamic volatility strategies proposed by Barroso and Santa-Clara (2015) and Daniel and Moskowitz (2016) respectively are effective to avoid the crashes and improve the WML portfolios. The constant and dynamic volatility portfolios with the GJR-GARCH(1,1)-M estimates outperform not only the corresponding market portfolios but also the plain momentum counterparts. The improvements of the dynamic volatility portfolios are proportional, but those of the constant volatility counterparts are disproportionate. The decrease in the negative skewness estimates after these portfolios is economically significant. The momentum returns across the countries are weakly correlated with each other.

Flight-to-quality by U.S. institutional investors precedes the momentum crashes. The institutional ownership changes proxy the flight-to-quality. The returns react positively to the flight-to-quality during the non-bear markets. This relation becomes negative during the bear-down markets. This reversal confirms both the smart money effect by the institutional investors and the price rebounds of the securities followed by the flight-to-quality. The smart money effect comes from the negative institutional ownership changes, i.e. the outflows, rather than the positive counterparts, i.e. the inflows. Likewise, the reversal arises after the outflows but does not after the inflows. The rebounds are more pronounced among the loser securities than the winner counterparts, resulting in the collapses of the WML portfolios.

Momentum returns exhibit consistent patterns when both FTQ and FTR events are identified by stock and bond market data rather than US institutional investor data. The FTQ and FTR events identified by Baele et al. (2019) and Boudry et al. (2019) demonstrate the positive reaction of the WML returns during the FTQ events and the negative reaction during the FTR events. The results also show that the negative FTR reaction is greater than the positive FTQ reaction. These empirical findings answer why the momentum crashes occur in part.

The flight-to-quality coincides with how default risk behaves around the bear markets. Two leverage ratios, the total liabilities and current liabilities-to-shareholders' equity ratios respectively, proxy the default risk. During the bear markets, the institutional flows to the loser securities fall more than those to the winner counterparts do. At the same time, the loser securities become riskier than the winner counterparts. The evidence advocates the time-varying risk literature as the default risk measures as well as the central betas exhibit a negative relation with the flight-to-quality proxies. The results associate the increased loser default risk during the bear markets with the institutional flight-to-quality and exhibit the post flight loser rebounds that crashes momentum.

Table 1.1. Market and Momentum Portfolios by Country

This table presents summary statistics of market and momentum returns of each country. The returns are US dollar return, i.e. local returns are coverted into US dollar retruns using exchange rates. Monthly US and Canadan return data are collected from Compustat. Monthly local return data are collected from Datastream and classified by the FactSet ISO country codes—classes other than EQ are excluded. If Datastream, I only include the top 90% of each country by size each month to exclude outliers following the previous literature. For each country, the value-weighted market portfolio and the winner-minus-loser momentum portfolio are formed respectively. I construct the WML portfolio using quintile or tercile († is displayed next to the country name) depending on the number of stocks available in the country following Daniel and Moskowitz (2016). The second and third columns display the beginning and ending periods of my samples. N is the time-series average of the number of stocks in the country and T is the number of monthly observations. The market and momentum columns report the mean, standard deviation, skewness, and annualized Sharpe ratio of market portfolio and momentum portfolio. The last row displays the pooled estimates from all country returns and the corresponding t-statistics. The table shows the results of MSCI developed countries and the results of MSCI emerging and other countries are available in the appendix.

					Market (Value-Weighted)				Momentum (-12,-2)						
Country	Start	Finish	N	T	Mean (%)	St Dev (%)	Skew	Sharpe	Mean (%)	St Dev (%)	Skew	Winner Mean (%)	Loser Mean (%)	Winner Sharpe	Loser Sharpe
United States	19500731	20181231	3,572.1	822	0.634	4.247	-0.555	0.517	1.046	6.074	-1.146	1.049	0.003	0.679	0.002
Canada	19510131	20181231	106.5	816	0.511	5.532	-0.524	0.320	1.138	10.494	-0.557	0.905	-0.233	0.400	-0.076
Japan	19740228	20150731	554.4	498	0.389	6.042	0.270	0.223	0.195	6.391	-0.499	0.573	0.378	0.293	0.178
United Kingdom	19660131	20150731	136.3	595	0.637	6.173	0.945	0.357	0.571	5.909	-0.585	0.878	0.308	0.461	0.134
Australia	19740228	20150731	72.4	498	0.656	7.039	-0.720	0.323	0.680	6.150	0.075	0.982	0.302	0.436	0.127
France	19740228	20150630	66.3	497	0.615	6.188	-0.281	0.344	0.480	6.935	-0.563	0.938	0.458	0.512	0.183
Germany	19740228	20150630	39.9	497	0.387	5.880	0.047	0.228	0.810	9.859	-0.193	0.993	0.183	0.462	0.072
Hong Kong	19740228	20150731	68.7	498	1.041	8.993	-0.310	0.401	0.830	8.087	-0.881	1.410	0.580	0.519	0.185
Singapore	19740228	20150731	49.6	498	0.761	8.136	0.806	0.324	0.417	8.019	-1.160	1.132	0.714	0.419	0.224
Sweden	19830228	20150630	26.0	389	0.988	7.536	-0.140	0.454	0.966	11.625	-1.472	1.839	0.873	0.753	0.253
Israel	19870228	20150630	95.6	341	0.648	7.946	-0.554	0.283	1.691	14.264	-0.629	-0.296	1.395	0.410	-0.080
Italy	19740228	20150630	24.9	482	0.297	6.955	0.332	0.148	0.778	10.635	-0.448	0.533	-0.196	0.205	-0.061
Switzerland	19740228	20150630	27.5	497	0.747	5.112	-0.360	0.506	0.575	7.007	0.031	0.950	0.374	0.530	0.169
Norway	19810228	20150630	25.5	413	0.725	7.441	-0.558	0.338	1.132	9.322	-0.067	1.334	0.202	0.508	0.065
Netherlands	19740228	20150731	23.9	498	0.618	5.696	-0.288	0.376	0.770	9.418	-0.168	1.101	0.331	0.535	0.123
Denmark	19740228	20150630	23.8	497	0.638	5.924	-0.337	0.373	1.316	9.195	-0.383	1.386	0.071	0.639	0.026
Belgium†	19740228	20150630	15.8	497	0.519	5.677	-0.356	0.316	0.001	7.749	-1.002	0.593	0.615	0.331	0.266
Spain†	19880430	20121130	28.8	296	0.348	7.273	-0.307	0.166	-0.277	9.152	-2.162	0.737	0.460	0.230	0.251
New Zealand†	19870228	20150630	20.0	341	0.492	6.620	-0.448	0.257	0.837	7.110	-0.156	0.928	0.091	0.468	0.037
Finland†	19880229	20150630	18.7	317	0.632	9.199	0.239	0.238	1.097	9.253	-0.155	1.386	0.289	0.544	0.094
Austria†	19740228	20150630	15.4	497	0.307	6.515	-0.217	0.163	0.837	6.557	-0.693	0.715	-0.122	0.361	-0.055
Ireland†	19660131	20150731	10.2	578	0.614	6.379	-0.078	0.333	1.305	9.405	-0.225	1.115	-0.168	0.502	-0.064
Portugal†	19890930	20150630	9.7	294	0.213	7.433	0.462	0.099	1.142	11.242	0.033	0.814	-0.349	0.307	-0.106
Total					0.594***	6.592***	-0.076	0.312***	0.805***	8.720***	-0.576***	1.015***	0.209***	0.455***	0.076***
					(9.46)	(85.17)	(-0.62)	(9.38)	(9.86)	(65.67)	(-2.99)	(11.39)	(2.35)	(13.62)	(2.37)

Table 1.2. Momentum Profits for Different Time Periods

This table reports average monthly returns of all, winner, loser, and winner-minus-loser (WML) portfolios for three different time periods: (i) normal periods, (ii) bear down market periods, and (iii) bear up market periods. I_B is the bear market indicator, I_{BD} is the bear down market indicator, and I_{BU} is the bear market and up market indicator. I define the bear market if the 2-year excess market return is negative, or 0 otherwise. I define the down (up) market if the contemporaneous excess market return is negative (positive), or 0 otherwise. The loser and winner portfolios are the bottom and top quintile—or tercile if less than 300 stocks are available—portfolios. The fourth and fifth columns of each period exhibit t-statistics of the winner-minus-loser portfolios and number of monthly observations. The returns of all, loser, winner portfolios are the excess returns after the one-month T-bill rates and the return of the winner-minus-loser portfolios is the long-short return. The last row displays the pooled estimates from all countries and their bootstrap t-statistics. This table includes 23 MSCI developed countries - other emerging and unclassified countries are also available in the appendix.

	Panel A						Panel B						Panel C					
	$I_B = 0$						$I_{BD} = 1$						$I_{BU} = 1$					
Country	All	Winner	Loser	WML	t Stat	T	All	Winner	Loser	WML	t Stat	T	All	Winner	Loser	WML	t Stat	T
United States	0.969	1.430	0.160	1.270	(6.35)	718	-4.944	-4.241	-9.006	4.765	(4.23)	40	5.329	4.664	9.798	-5.134	(-3.77)	50
Canada	0.656	1.174	-0.547	1.722	(4.67)	636	-3.651	-3.409	-7.042	3.633	(2.90)	72	5.391	5.113	9.347	-4.234	(-2.76)	99
Japan	0.224	0.353	0.544	-0.190	(-0.43)	200	-5.435	-4.860	-6.421	1.561	(2.60)	78	3.841	3.687	4.997	-1.310	(-2.20)	122
United Kingdom	0.646	0.709	0.904	-0.195	(-0.30)	152	-6.499	-5.771	-8.311	2.540	(2.66)	56	4.813	4.488	6.280	-1.791	(-2.14)	96
Australia	1.631	1.489	1.710	-0.221	(-0.47)	154	-5.135	-4.515	-5.232	0.717	(0.91)	51	4.981	4.462	5.147	-0.685	(-1.18)	103
France	0.820	1.073	1.129	-0.057	(-0.09)	177	-6.123	-4.339	-8.333	3.994	(4.05)	56	4.033	3.577	5.508	-1.931	(-2.58)	121
Germany	0.643	1.244	0.591	0.653	(0.75)	178	-4.906	-2.951	-6.274	3.322	(2.91)	64	3.758	3.599	4.445	-0.846	(-0.71)	114
Hong Kong	0.805	1.043	0.770	0.273	(0.32)	130	-8.849	-6.642	-11.553	4.911	(4.08)	44	5.745	4.975	7.074	-2.100	(-2.00)	86
Singapore	1.095	1.305	2.044	-0.739	(-0.87)	144	-5.540	-4.480	-8.518	4.038	(5.32)	53	4.960	4.675	8.196	-3.521	(-3.01)	91
Sweden	1.000	2.482	1.471	1.011	(0.55)	96	-7.430	-2.715	-11.248	8.533	(4.19)	37	6.287	5.741	9.447	-3.706	(-1.48)	59
Israel	-0.083	0.753	-0.835	1.588	(2.23)	193	-7.182	-8.487	-8.164	-0.323	(-0.10)	28	4.882	5.770	4.000	1.770	(0.96)	96
Italy	0.226	0.947	0.120	0.898	(1.24)	217	-5.525	-4.590	-7.448	2.857	(3.44)	88	4.149	4.753	5.282	-0.449	(-0.42)	129
Switzerland	0.823	0.701	0.945	-0.244	(-0.30)	126	-4.654	-3.477	-8.045	4.568	(2.97)	42	3.562	2.790	5.440	-2.650	(-3.12)	84
Norway	1.436	1.959	0.619	1.340	(1.29)	130	-6.703	-4.853	-11.139	6.285	(3.19)	37	4.674	4.670	5.297	-0.627	(-0.54)	93
Netherlands	0.378	0.321	0.446	-0.125	(-0.14)	155	-6.211	-5.251	-9.426	4.175	(2.65)	48	3.334	2.820	4.874	-2.054	(-1.92)	107
Denmark	1.588	2.217	1.235	0.982	(1.13)	165	-5.876	-5.154	-8.556	3.402	(2.15)	42	4.137	4.733	4.578	0.156	(0.15)	123
Belgium	0.385	0.489	0.235	0.254	(0.36)	186	-4.649	-3.390	-6.184	2.793	(2.47)	70	3.422	2.830	4.108	-1.278	(-1.43)	116
Spain	0.714	0.759	1.199	-0.440	(-0.59)	136	-7.722	-5.345	-9.067	3.722	(3.17)	46	3.988	3.591	4.633	-1.042	(-1.13)	90
New Zealand	0.260	1.104	-0.042	1.147	(1.53)	120	-7.509	-4.975	-8.936	3.961	(3.35)	42	4.444	4.378	4.747	-0.369	(-0.40)	78
Finland	0.661	1.474	0.796	0.679	(0.66)	120	-9.122	-5.806	-10.564	4.758	(2.73)	42	5.929	5.394	6.912	-1.518	(-1.25)	78
Austria	0.088	0.466	0.057	0.409	(0.90)	254	-3.913	-3.519	-3.839	0.320	(0.62)	112	3.243	3.609	3.130	0.479	(0.68)	142
Ireland	0.624	0.526	0.176	0.350	(0.47)	193	-5.823	-5.285	-7.210	1.925	(1.63)	66	3.974	3.546	4.014	-0.469	(-0.49)	127
Portugal	0.608	1.162	0.558	0.640	(0.53)	118	-6.996	-5.756	-8.256	2.500	(1.51)	45	5.295	5.486	5.992	-0.523	(-0.31)	73
Total	0.608***	1.086***	-0.019**	1.101**			-5.838***	-4.620***	-7.760***	3.140***			4.374***	4.204***	5.490***	-1.281***		
	(8.51)	(12.08)	(-0.21)	(12.51)			(-41.09)	(-26.01)	(-32.23)	(12.32)			(36.79)	(30.91)	(23.84)	(-5.52)		

Table 1.3. Time-Varying Betas of Momentum Portfolios in Bear and Up Markets

Following Daniel and Moskowitz (2016), I estimate the market timing regression for each country to capture the changing behavior of WML portfolios during bear down (BD) and bear up (BU) markets.

$$R_{\mathrm{WML}it} = \alpha_i + (\beta_i + (\beta_{BDi} I_{BDit} + \beta_{BUi} I_{BUit}))\, R_{Mit} + \varepsilon_{it}$$

Where the subscripts i and t represent each country and month, respectively. R_{WML} is the return of the WML portfolio, R_M is the market return of each country, I_{BD} is the bear down market indicator, and I_{BU} is the bear up market indicator, respectively. I define the bear market if the 2-year excess market return is negative, or 0 otherwise. I define the down (up) market if the contemporaneous excess market return is negative (positive), or 0 otherwise. This table only covers the results of the specification above - the regression results of other specifications are available in the appendix. The last row displays the pooled estimates from all countires and their t-statistics. This table includes 23 MSCI developed countries - other emerging and unclassified countries are also available in the appendix. The asterisks indicate the significance at 10% (*), 5% (**), and 1% (***) levels.

Country	α (%)	$t\,(\alpha)$	β	$t\,(\beta)$	β_{BD}	$t\,(\beta_{BD})$	β_{BU}	$t\,(\beta_{BU})$
United States	1.275***	(6.02)	-0.054	(-1.00)	-0.632***	(-4.04)	-1.143***	(-7.87)
Canada	1.780***	(4.66)	0.054	(0.73)	-0.537**	(-2.18)	-1.322***	(-7.41)
Japan	0.315	(1.03)	0.405***	(6.17)	-0.652***	(-5.08)	-0.888***	(-8.49)
United Kingdom	0.787***	(3.05)	0.024	(0.45)	-0.225**	(-2.00)	-0.475***	(-5.44)
Australia	1.023***	(3.40)	0.181***	(3.69)	-0.251*	(-1.77)	-0.489***	(-4.94)
France	1.094***	(3.42)	-0.004	(-0.06)	-0.438***	(-3.26)	-0.858***	(-7.63)
Germany	1.393***	(2.99)	0.064	(0.63)	-0.314	(-1.50)	-0.680***	(-3.87)
Hong Kong	0.974***	(2.75)	0.157***	(3.25)	-0.612***	(-5.43)	-0.601***	(-6.30)
Singapore	0.747**	(2.02)	0.113*	(1.91)	-0.567***	(-3.59)	-0.788***	(-7.06)
Sweden	1.343**	(2.24)	0.260***	(2.81)	-1.153***	(-5.38)	-1.439***	(-8.15)
Israel	2.976***	(3.66)	0.274**	(2.12)	0.383	(1.30)	-0.998***	(-4.31)
Italy	0.924	(1.64)	-0.079	(-0.77)	-0.227	(-1.14)	-0.313*	(-1.84)
Switzerland	0.475	(1.38)	0.166**	(2.20)	-0.968***	(-4.59)	-0.670***	(-4.08)
Norway	0.734	(1.45)	0.083	(1.03)	-0.536***	(-2.75)	-0.306**	(-1.99)
Netherlands	1.007**	(2.22)	0.198**	(2.02)	-0.690***	(-3.46)	-1.067***	(-5.67)
Denmark	1.914***	(4.20)	0.079	(0.82)	-0.078	(-0.39)	-0.664***	(-3.98)
Belgium	-0.017	(-0.04)	0.254***	(3.23)	-0.972***	(-5.89)	-0.870***	(-6.02)
Spain	0.265	(0.48)	-0.067	(-0.56)	-0.406**	(-2.29)	-0.535***	(-3.16)
New Zealand	1.146***	(3.12)	0.036	(0.47)	-0.265**	(-1.97)	-0.551***	(-4.38)
Finland	1.516***	(2.66)	0.168**	(2.18)	-0.622***	(-4.03)	-0.797***	(-6.30)
Austria	1.129***	(3.40)	0.083	(1.37)	-0.194	(-1.60)	-0.518***	(-4.53)
Ireland	1.522***	(3.44)	0.132	(1.56)	-0.427**	(-2.31)	-0.701***	(-4.49)
Portugal	2.566***	(3.14)	0.307	(1.62)	-0.129	(-0.46)	-0.976***	(-4.01)
Total	1.041***		0.098***		-0.458***		-0.630***	
	(11.66)		(2.93)		(-6.77)		(-8.89)	

Table 1.4. Backward and Forward Market Betas of Momentum Portfolios

This table reports equal-weighted average of market beta (β) of winner, loser, and winner-minus-loser (WML) portfolios for three different time periods: (i) normal periods, (ii) bear down market periods, and (iii) bear up market periods. Market beta (β) is measured from rolling regressions for three different horizons for each stock: from -5 to 6 months $\left(\beta_{(-5,6)}\right)$, from -23 to -12 months $\left(\beta_{(-23,-12)}\right)$, and from 13 to 24 months $\left(\beta_{(13,24)}\right)$ relative to the portfolio formation period. I_B is the bear market indicator, I_{BD} is the bear down market indicator, and I_{BU} is the bear market and up market indicator. I define the bear market if the 2-year excess market return is negative, or 0 otherwise. I define the down (up) market if the contemporaneous excess market return is negative (positive), or 0 otherwise. The loser and winner portfolios are the bottom and top quintile—or tercile if less than 300 stocks are available—portfolios. This table only includes 21 MSCI developed countries - other emerging and unclassified countries are also available in the appendix. WML columns of each period include asterisks, which indicate the significance at 10% (*), 5% (**), and 1% (***) levels.

	Panel A								
	$\beta_{(-5,6)}$								
	$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$		
Country	W	L	WML	W	L	WML	W	L	WML
Japan	1.08	0.83	0.25***	0.91	1.21	-0.3***	0.86	1.17	-0.31***
United Kingdom	1.09	0.99	0.1***	0.91	1.42	-0.51***	0.92	1.29	-0.37***
Australia	1.06	0.93	0.13***	0.93	1.06	-0.13**	0.9	1.07	-0.17***
France	1.04	0.94	0.1***	0.84	1.39	-0.55***	0.79	1.32	-0.54***
Germany	0.95	0.77	0.18***	0.59	0.95	-0.37***	0.65	1.07	-0.42***
Hong Kong	0.99	0.93	0.06**	0.82	1.2	-0.38***	0.89	1.21	-0.32***
Singapore	1.11	1.02	0.09***	0.93	1.4	-0.47***	0.87	1.21	-0.34***
Sweden	0.93	0.9	0.04	0.49	1.35	-0.86***	0.64	1.37	-0.73***
Israel	1.17	0.96	0.21***	0.83	1.2	-0.37***	0.84	1.21	-0.37***
Italy	1.1	0.88	0.22***	0.87	1.33	-0.46***	0.8	1.26	-0.46***
Switzerland	1.02	0.96	0.06***	0.86	1.69	-0.83***	0.87	1.31	-0.44***
Norway	0.99	0.97	0.02	0.95	1.27	-0.32***	1.04	1.22	-0.18**
Netherlands	1.05	0.94	0.11***	0.89	1.11	-0.23**	0.85	1.21	-0.36***
Denmark	1.03	0.91	0.12***	0.88	1.07	-0.19**	0.89	1.14	-0.26***
Belgium	0.92	0.68	0.25***	0.77	1.08	-0.31***	0.72	1.04	-0.32***
Spain	0.92	0.87	0.05	0.66	1.27	-0.59***	0.57	1.2	-0.63***
New Zealand	0.9	0.91	-0.01	0.87	1	-0.13**	0.78	1.01	-0.23***
Finland	0.92	0.67	0.25***	0.61	0.99	-0.38***	0.61	0.97	-0.36***
Austria	1.07	0.86	0.21***	0.95	1.16	-0.21**	0.89	1.03	-0.14**
Ireland	0.95	0.89	0.06	0.76	1.05	-0.29***	0.75	1	-0.25***
Portugal	1.15	0.75	0.39***	0.8	1.12	-0.34***	0.79	1.17	-0.41***
Total	1.02***	0.88***	0.14***	0.82***	1.21***	-0.39***	0.81***	1.17***	-0.36***
	(58.13)	(42.63)	(6.28)	(29.27)	(30.59)	(-9.09)	(31.83)	(46.21)	(-11.35)

	Panel B									Panel C								
	$\beta_{(-23,-12)}$									$\beta_{(13,24)}$								
	$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$			$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$		
Country	W	L	WML	W	L	WML	W	L	WML	W	L	WML	W	L	WML	W	L	WML
Japan	0.99	0.97	0.02	0.9	1.07	-0.17***	0.89	1.07	-0.18***	1.01	0.9	0.11***	0.91	1.15	-0.23***	0.9	1.08	-0.18***
United Kingdom	1.05	1.02	0.02	0.95	1.11	-0.17**	1.02	1.12	-0.1*	1.1	1.01	0.09***	1.03	1.19	-0.16**	1.04	1.14	-0.1**
Australia	0.97	1	-0.03	0.9	1.02	-0.13**	0.89	1.01	-0.12***	1.01	0.94	0.08***	0.94	1.09	-0.15***	0.89	1.06	-0.17***
France	1.05	1	0.05*	0.89	1.14	-0.26***	0.85	1.09	-0.24***	1.05	0.95	0.11***	0.94	1.14	-0.19***	0.88	1.09	-0.21***
Germany	0.86	0.86	0	0.53	0.89	-0.36***	0.69	0.94	-0.25***	1	0.78	0.22***	0.68	0.89	-0.2***	0.81	0.95	-0.14***
Hong Kong	1.03	0.99	0.04	0.77	1.09	-0.32***	0.98	1.06	-0.08	1.03	0.89	0.14***	0.85	1.07	-0.22***	0.91	1.03	-0.12**
Singapore	1.08	1.17	-0.09***	0.86	0.98	-0.12**	0.91	1.02	-0.11**	1.04	1.05	-0.01	0.92	1.32	-0.4***	1.01	1.3	-0.3***
Sweden	0.98	0.89	0.09***	0.59	1.01	-0.38***	0.65	0.97	-0.31***	0.93	0.8	0.13***	0.72	1.31	-0.59***	0.87	1.21	-0.33***
Israel	1.01	1.06	-0.06*	0.8	1.04	-0.24***	0.75	1.04	-0.29***	1.13	0.99	0.13***	0.79	0.98	-0.19**	0.86	0.96	-0.09*
Italy	1.03	0.89	0.14***	0.87	1.07	-0.19***	0.81	1.13	-0.32***	0.98	1.01	-0.02	0.83	1.25	-0.42***	0.8	1.26	-0.45***
Switzerland	0.98	0.95	0.03	0.97	1.08	-0.11	1.11	1.06	0.05	1.05	0.91	0.14***	1	1.15	-0.15	1.08	1.08	0
Norway	1.02	1.03	-0.02	0.86	1.28	-0.42***	0.87	1.1	-0.23***	1.02	0.93	0.09***	0.97	1.29	-0.32***	0.86	1.16	-0.3***
Netherlands	0.99	0.99	0	0.83	1.05	-0.22**	0.93	0.97	-0.04	1.01	0.83	0.17***	0.96	1.09	-0.13	0.96	1.11	-0.15**
Denmark	1.06	0.98	0.08***	0.91	1.07	-0.17*	0.96	1.1	-0.15**	0.89	0.95	-0.06**	1	1.18	-0.18*	0.97	1.05	-0.08*
Belgium	0.88	0.77	0.11***	0.74	0.88	-0.14**	0.83	0.9	-0.06	0.86	0.83	0.03	0.85	0.94	-0.1	0.92	0.9	0.02
Spain	0.78	0.79	-0.01	0.79	1.07	-0.3***	0.85	1.14	-0.3***	0.9	0.89	0	0.79	1.72	-0.93**	0.83	1.41	-0.54***
New Zealand	0.9	0.85	0.05***	0.8	1.03	-0.23***	0.76	1	-0.24***	0.86	0.9	-0.03*	0.81	0.96	-0.15**	0.78	1.01	-0.22***
Finland	0.83	0.71	0.12***	0.71	0.85	-0.14**	0.83	0.89	-0.06	0.77	0.62	0.16***	0.65	0.95	-0.3***	0.73	0.98	-0.25***
Austria	1.01	0.87	0.15***	0.96	1.07	-0.11	0.96	0.97	-0.02	1.01	0.92	0.09**	0.9	1.36	-0.46***	0.95	1.03	-0.08
Ireland	0.88	0.87	0	0.77	0.95	-0.19**	0.75	0.95	-0.2***	0.92	0.81	0.12***	0.88	1.01	-0.13	0.87	0.93	-0.06
Portugal	1.03	1	0.03	0.87	1.26	-0.39***	0.92	1.11	-0.2**	1.05	0.91	0	0.76	1.15	-0.39***	0.84	1.2	-0.36***
Total	0.97***	0.94***	0.03**	0.82***	1.05***	-0.23***	0.87***	1.03***	-0.16***	0.98***	0.90***	0.08***	0.87***	1.15***	-0.29***	0.89***	1.09***	-0.20***
	(53.47)	(39.60)	(2.55)	(33.38)	(45.41)	(-10.45)	(35.91)	(60.97)	(-7.06)	(51.15)	(42.74)	(4.84)	(37.21)	(28.07)	(-6.63)	(47.51)	(38.34)	(-6.21)

Table 1.5. Time-Varying Maximum Sharpe Ratio Momentum Strategies with GARCH

Following Daniel and Moskowitz (2016), I implement the constant and dynamic volatility strategies in addition to plain WML strategy. I estimate full-sample GJR-GARCH(1,1)-M using the plain momentum portfolio, and apply the scaling factor, $1/\left(2\lambda\sigma_{t-1}\right)$ for the constant volatility strategy and $\mu_{t-1}/\left(2\lambda\sigma_{t-1}^2\right)$ for the dynamic volatility strategy, respectively. I scale the constant and dynamic strategies using the time-invariant scalar λ, so that the full sample volatilities of the plain, constant, and dynamic strategies are equal. The specification of the GARCH model estimated by the maximum likelihood method is as follow:

$$\begin{aligned} R_t &= \mu + \delta h_t + \varepsilon_t \\ h_t &= \omega + \left(\alpha + \gamma I_{t-1}\right)\varepsilon_{t-1}^2 + \beta h_{t-1}. \end{aligned}$$

R_t is the return of the plain momentum portfolio, h_t is the time-varying volatility, and I_t is the indicator variable for negative ε_t. ω, α, and β govern the GARCH process. The parameters γ and δ reflect the asymmetric effect and the volatility-to-return effect, respectively. WML, CVOL, and DVOL are the average monthly returns of the plain, constant volatility, and dynamic volatility portfolios, respectively. The numbers in the sqaure brackets are annualized Sharpe ratios from the three momentum portfolios and the numbers in the parentheses are the t-statistics of GARCH parameter estimates. The last row, Total, displays the pooled statistics and estimates from all countries and their t-statistics. This table includes 23 MSCI developed countries - other emerging and unclassified countries are also available in the appendix.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Country	WML	CVOL	DVOL	μ	ω	α	γ	β	δ
United States	1.0455	1.5070	1.6416	0.0135***	0.0001***	0.3748***	-0.3390***	0.7730***	-1.1249*
	[0.596]	[0.859]	[0.936]	(7.542)	(3.983)	(6.815)	(-5.669)	(34.204)	(-1.683)
Canada	1.1380	1.2688	1.2945	0.0118*	0.0003***	0.1521***	-0.1404***	0.8951***	-0.1129
	[0.376]	[0.419]	[0.427]	(1.875)	(2.816)	(4.922)	(-4.242)	(-43.663)	(-0.170)
Japan	0.1948	0.3676	0.5691	0.0078**	0.0003***	0.3020***	-0.2740***	0.7831***	-1.4265
	[0.106]	[0.199]	[0.308]	(2.263)	(2.680)	(4.039)	(-3.939)	(14.266)	(-1.346)
United Kingdom	0.5706	0.7387	0.7941	0.0116***	0.0002***	0.1765***	-0.1234***	0.8280***	-2.0606*
	[0.334]	[0.433]	[0.466]	(3.633)	(3.470)	(4.310)	(-3.209)	(23.773)	(-1.815)
Australia	0.6803	0.7925	0.8404	0.0079	0.0004***	0.2745***	-0.1785**	0.7126***	-0.2213
	[0.383]	[0.446]	[0.473]	(1.495)	(2.603)	(3.858)	(-2.530)	(10.062)	(-0.136)
France	0.4797	0.6489	0.7993	0.0089*	0.0004***	0.3675***	-0.3317***	0.7327***	-0.7871
	[0.240]	[0.324]	[0.399]	(1.859)	(3.198)	(3.841)	(-3.383)	(15.318)	(-0.648)
Germany	0.8101	1.1905	1.5066	0.0159***	0.0010***	0.4978***	0.0086	0.5748***	-0.5263
	[0.285]	[0.418]	[0.529]	(3.049)	(4.771)	(6.516)	(0.092)	(14.821)	(-1.028)
Hong Kong	0.8297	1.3068	1.7546	0.0176***	0.0002***	0.2639***	-0.2181***	0.8253***	-1.3529
	[0.355]	[0.560]	[0.752]	(3.715)	(3.144)	(5.310)	(-4.382)	(43.591)	(-1.417)
Singapore	0.4174	0.8255	1.4634	0.0164***	0.0004***	0.3509***	-0.2894***	0.7807***	-1.5718*
	[0.180]	[0.357]	[0.632]	(3.079)	(3.390)	(5.100)	(-4.388)	(20.427)	(-1.646)
Sweden	0.9658	1.5714	1.5810	0.0111**	0.0003***	0.3052***	-0.3052***	0.8310***	-0.2677
	[0.288]	[0.468]	[0.471]	(2.266)	(2.625)	(5.309)	(-5.309)	(34.302)	(-0.514)
Israel	1.6906	2.4004	2.7618	0.0150***	0.0002	0.2918***	-0.1520*	0.8053***	0.0296
	[0.411]	[0.583]	[0.671]	(2.596)	(1.514)	(3.566)	(-1.897)	(20.031)	(0.072)
Italy	0.7777	1.1357	1.3932	0.0066	0.0001	0.2459***	-0.1587***	0.8602***	0.0664
	[0.253]	[0.370]	[0.454]	(1.408)	(1.355)	(7.250)	(-4.137)	(52.772)	(0.123)
Switzerland	0.5753	0.7459	0.8496	0.0105**	0.0002**	0.1995***	-0.1499**	0.8238***	-0.9567
	[0.284]	[0.369]	[0.420]	(2.183)	(2.276)	(2.900)	(-2.346)	(15.669)	(-0.822)
Norway	1.1321	1.1967	1.2231	0.0094	0.0007*	0.1982***	-0.1469**	0.8055***	0.2759
	[0.421]	[0.445]	[0.455]	(0.877)	(1.847)	(2.754)	(-2.041)	(11.663)	(0.205)
Netherlands	0.7699	1.0371	1.1343	0.0117**	0.0005***	0.2793***	-0.2706***	0.8089***	-0.4887
	[0.283]	[0.381]	[0.417]	(2.046)	(2.687)	(4.409)	(-4.649)	(17.415)	(-0.613)
Denmark	1.3156	1.6527	1.7616	0.0138**	0.0006***	0.2982***	-0.1666*	0.7272***	-0.2219
	[0.496]	[0.623]	[0.664]	(2.074)	(3.442)	(3.535)	(-1.757)	(13.600)	(-0.238)
Belgium	0.0009	0.2559	0.8383	0.0166***	0.0008***	0.3522***	-0.2312***	0.6407***	-2.7491**
	[0.000]	[0.114]	[0.375]	(2.999)	(3.975)	(4.595)	(-3.122)	(12.439)	(-2.376)
Spain	-0.2765	0.4764	1.6956	0.0132***	0.0001	0.4482***	-0.1515	0.7246***	-0.9493
	[-0.105]	[0.180]	[0.642]	(3.017)	(1.169)	(4.557)	(-1.498)	(15.846)	(-1.175)
New Zealand	0.8367	0.9088	0.9295	0.0034	0.0001***	0.1382***	-0.0343	0.8479***	1.0355
	[0.408]	[0.443]	[0.453]	(0.786)	(2.656)	(3.696)	(-0.860)	(30.437)	(0.914)
Finland	1.0967	1.1256	1.0985	0.0020	0.0003**	0.1784***	-0.0883*	0.8261***	1.0353
	[0.411]	[0.421]	[0.411]	(0.229)	(2.066)	(4.140)	(-1.837)	(15.819)	(0.912)
Austria	0.8367	0.9520	1.0185	0.0115**	0.0004***	0.2223***	-0.1496***	0.7551***	-0.8106
	[0.442]	[0.503]	[0.538]	(2.046)	(2.772)	(4.812)	(-3.028)	(14.020)	(-0.562)
Ireland	1.3047	1.3840	1.4169	0.0112	0.0004***	0.1236***	-0.0582*	0.8672***	0.1916
	[0.481]	[0.510]	[0.522]	(1.235)	(2.656)	(3.705)	(-1.692)	(30.427)	(0.177)
Portugal	1.1417	1.6473	2.0376	0.0157**	0.0002**	0.1753***	-0.1687***	0.8927***	-0.6355
	[0.352]	[0.508]	[0.628]	(2.533)	(2.297)	(3.532)	(-3.134)	(35.316)	(-0.878)
Total	0.8067***	1.0880***	1.2933***	0.0098***	0.0002***	0.2772***	-0.1848***	0.8053***	-0.2381
	[0.321]	[0.432]	[0.514]	(11.876)	(21.304)	(31.921)	(-21.695)	(196.847)	(-1.751)

Table 1.6. Institutional Ownership Changes and Stock Returns

In this table, I examine the effect of flight-to-quality, which is proxied by US institutional ownership changes, on stock returns. I use the following panel regression specification.

$$R_{cit+1} = \alpha_t + (\beta + (\gamma + \delta I_{Uct+1})\, I_{Bct})\, \Delta IO_{it} + \varepsilon_{cit+1},$$

where R is the dollar return, I_B is the bear market indicator, I_U is the up market indicator, and ΔIO is the U.S. institutional ownership change for each stock. For the up market indicator (I_{Uct+1}), I use the beginning month of the quarter (I_{U1}) or the same quarter $\left(I_{U(1,3)}\right)$ relative to the quarter of stock return. The subscripts c, i, and t imply each country, stock, and quarter, respectively. Monthly dollar return data are collected from Datastream. Quarterly investment flow data of institutional investors are collected from Global Ownership. I employ time fixed effects and clustered standard errors by country. For each country each quarter, I include the stocks that consist the top 90% market capitalization of the country. For each country also, I winsorize the top and bottom 1% of the ΔIO observations to prevent any outlier effect. ΔIO+ is equal to ΔIO if positive or zero otherwise. ΔIO- is equal to ΔIO if negative or zero otherwise. Panel A shows the results of return on ΔIO. Panel B contains the results of interaction between ΔIO and momentum loser portfolio. The asterisks indicate the significance at 10% (*), 5% (**), and 1% (***) levels.

Panel A: Regressions of return on ΔIO, changes in institutional ownership								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔIO	2.879***		2.943***		2.943***		2.943***	
	(13.10)		(9.04)		(9.04)		(9.04)	
ΔIO+		-0.274		-0.521		-0.520		-0.520
		(-0.64)		(-0.57)		(-0.57)		(-0.57)
ΔIO-		4.923***		5.021***		5.021***		5.021***
		(14.00)		(12.49)		(12.49)		(12.49)
$I_B \times$ ΔIO			-1.178		0.811		0.739	
			(-1.54)		(0.82)		(0.79)	
$I_B \times$ ΔIO+				-4.780***		-7.116***		-8.763***
				(-2.86)		(-3.11)		(-3.79)
$I_B \times$ ΔIO-				0.650		9.433***		8.062***
				(0.61)		(6.41)		(5.39)
$I_B \times I_{U1} \times$ ΔIO					-3.900***			
					(-4.48)			
$I_B \times I_{U1} \times$ ΔIO+						7.007**		
						(2.02)		
$I_B \times I_{U1} \times$ ΔIO-						-14.45***		
						(-6.95)		
$I_B \times I_{U(1,3)} \times$ ΔIO							-5.197***	
							(-4.83)	
$I_B \times I_{U(1,3)} \times$ ΔIO+								15.60***
								(3.52)
$I_B \times I_{U(1,3)} \times$ ΔIO-								-17.16***
								(-6.54)
N Observations	130648	130648	127912	127912	127904	127904	127883	127883
Adjusted R-Sq	0.242	0.242	0.241	0.241	0.241	0.241	0.241	0.242
Fixed Effects	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Clustered SE	Country	Country	Country	Country	Country	Country	Country	Country

Panel B: Regressions of return on interaction between ΔIO and momentum loser								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ΔIO	2.108***		1.214***		1.214***		1.214***	
	(10.74)		(4.12)		(4.12)		(4.12)	
ΔIO × Loser	11.09***		23.08***		23.08***		23.08***	
	(16.86)		(28.11)		(28.11)		(28.11)	
ΔIO+		-1.263***		-1.636*		-1.635*		-1.636*
		(-2.96)		(-1.91)		(-1.91)		(-1.91)
ΔIO-		4.585***		2.998***		2.998***		2.999***
		(19.23)		(7.81)		(7.81)		(7.81)
ΔIO+ × Loser		11.64		10.66*		10.66*		10.66*
		(1.52)		(1.80)		(1.80)		(1.80)
ΔIO- × Loser		8.871***		23.36***		23.36***		23.36***
		(8.12)		(29.73)		(29.73)		(29.73)
I_B × ΔIO			2.288***		1.231**		1.345***	
			(5.08)		(2.43)		(3.15)	
I_B × ΔIO × Loser			-30.64***		-9.708***		-6.342**	
			(-3.42)		(-3.84)		(-2.54)	
I_B × ΔIO+				-4.732***		-6.000***		-8.117***
				(-2.75)		(-2.79)		(-3.65)
I_B × ΔIO-				7.938***		11.50***		9.593***
				(7.72)		(8.52)		(7.19)
I_B × ΔIO+ × Loser				14.12		-40.17**		-159.7
				(0.37)		(-2.55)		(-1.05)
I_B × ΔIO- × Loser				-38.95***		-18.02***		-15.37***
				(-4.04)		(-4.17)		(-3.70)
$I_B \times I_{U1}$ × ΔIO					2.438***			
					(3.82)			
$I_B \times I_{U1}$ × ΔIO × Loser					-30.60***			
					(-16.15)			
$I_B \times I_{U1}$ × ΔIO+						4.540		
						(1.56)		
$I_B \times I_{U1}$ × ΔIO-						-6.523***		
						(-3.70)		
$I_B \times I_{U1}$ × ΔIO+ × Loser						272.3***		
						(2.69)		
$I_B \times I_{U1}$ × ΔIO- × Loser						-27.54***		
						(-7.83)		
$I_B \times I_{U(1,3)}$ × ΔIO							4.544***	
							(4.29)	
$I_B \times I_{U(1,3)}$ × ΔIO×Loser							-37.35***	
							(-23.29)	
$I_B \times I_{U(1,3)}$ × ΔIO+								19.67***
								(3.73)
$I_B \times I_{U(1,3)}$ × ΔIO-								-7.042**
								(-2.31)
$I_B \times I_{U(1,3)}$ × ΔIO+×Loser								172.9
								(1.22)
$I_B \times I_{U(1,3)}$ × ΔIO-×Loser								-28.69***
								(-7.14)
N Observations	120834	120834	119895	119895	119895	119895	119874	119874
Adjusted R-Sq	0.247	0.247	0.247	0.247	0.247	0.247	0.247	0.247
Fixed Effects	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter	Quarter
Clustered SE	Country	Country	Country	Country	Country	Country	Country	Country

Table 1.7. Institutional Ownership Changes of Momentum Portfolios

This table reports value-weighted average of institutional ownership changes of winner and loser portfolios for three different time periods: (i) normal periods, (ii) bear down market periods, and (iii) bear up market periods. Institutional ownership changes are changes of institutional ownership for each stock during the quarter before the portfolio formation period. I_B is the bear market indicator, I_{BD} is the bear down market indicator, and I_{BU} is the bear market and up market indicator. I define the bear market if the 2-year excess market return is negative, or 0 otherwise. I define the down (up) market if the contemporaneous excess market return is negative (positive), or 0 otherwise. The loser and winner portfolios are the bottom and top quintile—or tercile if less than 300 stocks are available—portfolios. This table only includes 21 MSCI developed countries - other emerging and unclassified countries are also available in the appendix.

	Panel A								
	$\Delta IO_{(t-3,t-1)}$								
	$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$		
Country	W	L	Diff	W	L	Diff	W	L	Diff
Japan	0.52	0.30	0.23***	0.08	-0.15	0.23***	0.12	-0.21	0.33***
United Kingdom	0.50	0.34	0.16***	0.09	0.04	0.05	0.27	0.12	0.15
Australia	0.41	0.23	0.18***	0.04	-0.06	0.09	0.07	0.10	-0.03
France	0.40	0.34	0.06	0.03	-0.14	0.17	0.06	-0.11	0.17**
Germany	0.26	0.26	0.00	0.00	-0.05	0.05	0.01	-0.11	0.12
Hong Kong	0.43	0.30	0.13**	0.02	-0.09	0.11**	0.13	-0.09	0.22***
Singapore	1.47	0.45	1.02***	0.11	-0.32	0.42	0.56	0.04	0.52**
Sweden	0.28	0.44	-0.16**	0.05	0.09	-0.05	0.04	0.18	-0.14
Israel	1.65	0.15	1.40***	0.59	-1.93	2.52**	0.90	-1.60	2.50***
Italy	0.23	0.31	-0.09	0.22	0.00	0.22**	0.09	0.01	0.08*
Switzerland	0.62	0.77	0.01	0.38	2.13	-1.75	0.49	0.03	0.46
Norway	0.63	0.05	0.57***	-0.33	-0.22	-0.13	0.10	-0.39	0.44**
Netherlands	0.23	0.39	-0.16	0.05	0.22	-0.17	-0.06	0.18	-0.24
Denmark	0.60	0.21	0.33**	0.47	-0.59	1.07**	0.39	-0.20	0.58**
Belgium	0.40	0.30	0.11	0.12	-0.12	0.24**	0.51	-0.11	0.62***
Spain	0.20	0.12	0.08	0.08	0.00	0.08	0.18	0.00	0.18**
New Zealand	0.20	0.40	-0.21***	-0.31	0.04	-0.36**	0.00	0.06	-0.06
Finland	0.32	0.41	-0.09	-0.18	-0.06	-0.12	0.24	-0.17	0.40***
Austria	0.27	0.17	0.10	0.27	-0.22	0.49***	0.36	-0.21	0.58***
Ireland	1.35	0.53	0.77***	0.18	0.31	-0.16	-0.36	-0.49	0.16
Portugal	0.41	0.38	0.02	0.17	-0.15	0.32	0.14	0.11	0.03
Total	0.54***	0.33***	0.21**	0.10*	-0.06	0.16	0.20***	-0.14	0.34**
	(5.90)	(9.55)	(2.38)	(2.07)	(-0.40)	(0.98)	(3.46)	(-1.66)	(2.79)

	Panel B									Panel C								
	$\Delta IO_{(t-3,t-2)}$									$\Delta IO_{(t-2,t-1)}$								
	$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$			$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$		
Country	W	L	Diff	W	L	Diff	W	L	Diff	W	L	Diff	W	L	Diff	W	L	Diff
Japan	0.26	0.14	0.13***	0.06	-0.05	0.11***	0.09	-0.11	0.21***	0.26	0.16	0.10***	0.02	-0.10	0.12***	0.06	-0.09	0.15***
United Kingdom	0.27	0.18	0.09***	0.12	0.08	0.04	0.10	0.02	0.08	0.24	0.16	0.08**	-0.03	-0.04	0.01	0.18	0.10	0.08
Australia	0.21	0.13	0.08**	0.03	-0.04	0.07	0.02	-0.01	0.03	0.21	0.11	0.10***	0.00	-0.02	0.02	0.04	0.10	-0.06
France	0.18	0.15	0.03	0.13	-0.05	0.18**	0.04	-0.10	0.14***	0.22	0.19	0.04	-0.10	-0.09	-0.01	0.01	-0.02	0.03
Germany	0.17	0.08	0.10	-0.07	0.29	-0.36	0.03	0.03	0.00	0.09	0.19	-0.09**	0.07	-0.34	0.41***	-0.02	-0.14	0.12
Hong Kong	0.23	0.14	0.08***	0.04	-0.02	0.06	0.10	-0.05	0.15***	0.21	0.17	0.04	-0.03	-0.08	0.05*	0.04	-0.04	0.07**
Singapore	0.87	0.24	0.63***	0.24	-0.08	0.32	0.11	-0.03	0.14	0.61	0.21	0.40***	-0.09	-0.23	0.14	0.50	0.06	0.44
Sweden	0.16	0.21	-0.05	0.10	0.20	-0.10	0.09	-0.06	0.15*	0.13	0.23	-0.10**	-0.05	-0.11	0.06	-0.05	0.22	-0.27***
Israel	0.83	0.11	0.72***	0.05	-0.61	0.66	0.19	-1.24	1.43***	0.81	0.04	0.75***	0.54	-1.31	1.85**	0.70	-0.37	1.07*
Italy	0.15	0.17	-0.02	0.10	0.03	0.07	0.06	0.00	0.06	0.09	0.13	-0.06	0.12	-0.03	0.15**	0.03	0.01	0.02
Switzerland	0.25	0.42	-0.08	0.55	2.26	-1.71	-0.15	-0.09	-0.06	0.37	0.34	0.08	-0.13	-0.12	-0.01	0.64	0.12	0.52
Norway	0.34	0.07	0.28***	-0.15	-0.03	-0.12	0.15	-0.26	0.38**	0.34	0.00	0.33***	-0.19	-0.20	0.00	-0.03	-0.14	0.08
Netherlands	0.17	0.20	-0.02	0.01	0.17	-0.16	-0.09	0.06	-0.14	0.07	0.23	-0.16	0.03	0.05	-0.02	0.02	0.12	-0.10
Denmark	0.28	0.08	0.24**	0.24	-0.33	0.57**	0.17	-0.14	0.30*	0.32	0.12	0.21*	0.23	-0.27	0.50**	0.22	-0.06	0.27
Belgium	0.20	0.18	0.02	0.06	-0.02	0.08*	0.30	-0.09	0.39***	0.20	0.11	0.09*	0.07	-0.10	0.17**	0.21	-0.03	0.23
Spain	0.11	0.06	0.05	0.08	0.02	0.06	0.09	-0.04	0.13**	0.11	0.06	0.05	-0.01	-0.02	0.02	0.09	0.03	0.05
New Zealand	0.12	0.20	-0.09**	-0.08	0.03	-0.11	-0.02	0.03	-0.05	0.08	0.20	-0.12**	-0.21	0.01	-0.22*	0.02	0.03	-0.01
Finland	0.18	0.17	0.01	-0.05	0.00	-0.04	0.10	-0.07	0.16	0.16	0.24	-0.08	-0.13	-0.05	-0.08	0.14	-0.10	0.24*
Austria	0.16	0.11	0.05	0.10	-0.06	0.16***	0.16	-0.15	0.31***	0.13	0.07	0.07	0.18	-0.16	0.34***	0.20	-0.07	0.27***
Ireland	0.84	0.27	0.56***	0.12	0.31	-0.19	-0.24	-0.35	0.13	0.60	0.26	0.34**	0.35	0.01	0.35**	0.03	-0.62	0.68
Portugal	0.18	0.18	0.00	0.14	-0.10	0.24**	0.02	-0.03	0.06	0.23	0.21	0.02	0.03	-0.05	0.07	0.11	0.13	-0.02
Total	0.29***	0.17***	0.13**	0.09**	0.10	-0.01	0.06**	-0.13**	0.19**	0.26***	0.16***	0.10**	0.03	-0.16**	0.19*	0.15***	-0.03	0.18***
	(5.65)	(9.37)	(2.64)	(2.80)	(0.82)	(-0.08)	(2.47)	(-2.12)	(2.75)	(6.10)	(9.21)	(2.14)	(0.82)	(-2.52)	(2.05)	(3.26)	(-0.85)	(2.85)

Table 1.8. Default Risk of Momentum Portfolios for Different Time Periods

This table reports equal-weighted average of book leverage of winner and loser portfolios for three different time periods: (i) normal periods, (ii) bear down market periods, and (iii) bear up market periods. Market leverage is a proxy for default risk and is defined by two different ways: Total Liabilities/Shareholder's Equity and Current Liabilities/Shareholder's Equity, which is based on the previous accounting year-end total liabilities and the previous calendar year december month-end market capitalizations. Total liabilities (max(WC03999-WC03995,0)), current liabilities (WC03101), and Shareholder's Equity (WC03995) are collected from Datastream. I_B is the bear market indicator, I_{BD} is the bear down market indicator, and I_{BU} is the bear market and up market indicator. I define the bear market if the 2-year excess market return is negative, or 0 otherwise. I define the down (up) market if the contemporaneous excess market return is negative (positive), or 0 otherwise. The loser and winner portfolios are the bottom and top quintile—or tercile if less than 300 stocks are available—portfolios. This table only includes 21 MSCI developed countries - other emerging and unclassified countries are also available in the appendix.

	Total Liabilities/Shareholder's Equity									Current Liabilities/Shareholder's Equity								
	$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$			$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$		
Country	W	L	Diff	W	L	Diff	W	L	Diff	W	L	Diff	W	L	Diff	W	L	Diff
Japan	4.96	4.93	0.03	4.60	4.99	-0.39	4.54	5.34	-0.80***	1.87	1.89	-0.02	1.74	2.12	-0.38***	1.62	2.17	-0.54***
United Kingdom	4.86	4.12	0.75***	4.01	5.40	-1.38*	4.18	5.19	-1.00	1.13	1.24	-0.11**	1.05	1.11	-0.06	0.95	1.32	-0.38***
Australia	2.36	2.06	0.32**	2.56	1.77	0.79**	2.18	1.84	0.34*	0.59	0.57	0.02	0.54	0.54	0.00	0.58	0.57	0.01
France	6.05	5.29	0.75**	5.38	6.47	-1.09	4.53	6.41	-1.88***	2.33	2.01	0.32***	2.07	2.34	-0.27**	1.92	2.55	-0.63***
Germany	8.74	9.96	-1.22*	11.35	11.59	0.24	10.21	9.96	0.58	1.99	2.19	-0.23*	1.38	2.20	-0.36*	1.42	2.62	-0.80***
Hong Kong	2.07	1.98	0.05	1.64	2.55	-0.90*	1.79	2.47	-0.68**	0.78	0.70	0.09***	0.62	0.76	-0.14*	0.77	0.71	0.07
Singapore	1.66	1.32	0.34***	1.37	1.32	0.06	1.59	1.42	0.17	0.65	0.60	0.06***	0.61	0.57	0.03	0.72	0.60	0.11***
Sweden	6.13	3.93	2.20***	2.93	9.76	-6.83***	2.87	10.04	-7.17***	1.36	1.24	0.14*	1.25	1.86	-0.75***	1.17	1.68	-0.44***
Israel	5.56	4.43	1.13***	4.42	6.29	-1.87	4.36	5.37	-0.96	1.40	1.17	0.23**	1.31	1.99	-0.68*	1.27	1.41	-0.13
Italy	5.81	5.60	0.32	4.88	5.35	-0.41	4.44	5.74	-1.23***	1.17	1.74	-0.62***	1.61	1.56	0.02	1.42	1.51	-0.08
Switzerland	3.71	4.41	-0.65**	3.14	7.33	-4.14***	2.51	5.08	-2.54***	0.80	0.79	0.02	0.67	2.66	-2.16***	0.69	1.74	-1.05***
Norway	7.84	6.77	1.06	6.81	5.82	0.99	6.48	6.65	-0.09	1.40	1.11	0.25**	1.03	0.80	0.23**	1.64	2.02	-0.44
Netherlands	4.42	4.87	-0.52	2.44	3.89	-1.27**	2.64	5.60	-2.87***	1.32	1.32	0.02	1.09	1.23	-0.13	1.15	1.46	-0.27***
Denmark	3.71	3.78	-0.13	3.21	3.74	-0.52	3.70	3.51	0.22	1.05	1.11	-0.05	0.99	1.24	-0.21	0.79	1.35	-0.44***
Belgium	13.30	10.09	2.68**	5.65	7.70	-2.61	6.59	8.14	-1.30	1.46	0.91	0.32***	1.39	1.16	0.20	1.31	1.29	0.08
Spain	6.42	6.10	0.32	5.56	4.76	-0.61	6.42	5.45	0.33	1.26	1.21	0.05	1.23	0.91	0.13	1.55	0.88	0.60***
New Zealand	1.17	1.36	-0.19***	1.76	4.19	-2.43***	1.44	3.39	-1.95***	0.50	0.55	-0.05***	0.46	0.55	-0.09	0.47	0.60	-0.13***
Finland	3.22	2.63	0.58**	3.47	3.60	-0.14	3.10	4.23	-1.13***	0.94	0.93	0.01	1.13	1.31	-0.18**	1.18	1.28	-0.10
Austria	6.32	6.76	-0.45	7.88	9.06	-1.27	5.58	6.51	-0.93	1.06	1.12	-0.06	0.96	1.27	-0.29**	0.96	1.24	-0.27***
Ireland	4.17	6.10	-1.65**	2.69	2.09	0.83	2.95	2.54	0.66	0.95	1.44	-0.49***	0.88	0.90	-0.03	0.94	0.89	0.05
Portugal	5.84	7.34	-1.70***	6.44	4.86	1.63**	7.82	7.21	0.53	1.63	1.44	0.47***	1.42	1.43	0.01	2.16	1.92	0.27
Total	2.72***	3.06***	-0.38*	2.84***	4.36***	-1.52***	2.98***	5.07***	-2.12***	0.55***	0.61***	-0.06**	0.59***	0.89***	-0.31***	0.65***	0.95***	-0.30***
	(11.63)	(10.43)	(-1.98)	(8.12)	(8.00)	(-3.25)	(10.80)	(8.77)	(-4.20)	(11.46)	(9.62)	(-2.31)	(9.99)	(7.21)	(-3.32)	(12.20)	(9.42)	(-3.93)

Table 1.9. Daily Return Correlations Between Stock and Bond Markets

This table displays the daily return correlations between each country's stock market and bond market. The daily return data for stock markets (TOT**MK) and bond markets (BM**10Y) are from Datastream return indices. The table contains the statistics from MSCI developed countries except the United States, Canada, Hong Kong, and Israel. The start and end dates of each country's time-series are in the second and third columns, respectively. The first four partitions are for the (1) non-bear, (2) bear, (3) bear-up, and (4) bear-down markets, respectively. Each partition exhibits the (i) average stock market returns, (ii) average bond market returns, (iii) stock standard deviations, (iv) bond standard deviations, and (v) the stock-bond correlation coefficients. The last partition is for the decrease in the correlations during the (a) bear, (b) bear-up, and (c) bear-down markets compared to the non-bear markets. Each statistic displays the respective t-statistic in the parenthesis below estimated from 5,000 bootstraps.

			$I_B = 0$			$I_{BD} = 1$			$I_{BU} = 1$			Correlation Difference	
Country	Start	Finish	$\bar{r}_{\text{Stock}}$	$\bar{r}_{\text{Bond}}$	Corr	$\bar{r}_{\text{Stock}}$	$\bar{r}_{\text{Bond}}$	Corr	$\bar{r}_{\text{Stock}}$	$\bar{r}_{\text{Bond}}$	Corr	$I_{BD} = 1$	$I_{BU} = 1$
United States	19800101	20181231	0.05	0.03	0.04	-0.24	0.09	-0.39	0.22	-0.01	-0.27	0.43	0.31
			(5.52)	(5.65)	(2.19)	(-3.14)	(4.03)	(-6.56)	(3.66)	(-0.19)	(-5.50)	(6.83)	(5.81)
Canada	19850101	20181231	0.04	0.03	-0.02	-0.07	0.05	-0.37	0.14	0.00	-0.24	0.35	0.22
			(3.78)	(7.11)	(-0.99)	(-1.54)	(3.68)	(-10.62)	(4.96)	(0.34)	(-4.83)	(8.43)	(3.95)
Japan	19840102	20150731	0.04	0.02	-0.08	-0.23	0.03	-0.03	0.16	0.02	-0.09	-0.04	0.01
			(2.68)	(3.58)	(-2.44)	(-5.65)	(3.81)	(-1.02)	(5.18)	(2.49)	(-3.24)	(-0.98)	(0.35)
United Kingdom	19800101	20150731	0.05	0.03	0.08	-0.15	0.05	-0.22	0.17	0.05	0.04	0.30	0.05
			(5.15)	(6.55)	(3.54)	(-2.46)	(2.84)	(-5.10)	(5.71)	(3.98)	(0.96)	(6.27)	(1.11)
Australia	19870302	20150731	0.03	0.04	0.03	-0.09	0.04	-0.08	0.15	0.04	-0.10	0.11	0.13
			(2.24)	(5.80)	(0.90)	(-2.21)	(1.47)	(-1.66)	(5.62)	(2.33)	(-2.86)	(1.78)	(2.56)
France	19850201	20150630	0.05	0.03	0.09	-0.23	0.05	-0.27	0.17	0.04	-0.04	0.36	0.13
			(2.93)	(5.00)	(4.26)	(-4.08)	(3.80)	(-7.16)	(6.89)	(4.94)	(-0.97)	(8.49)	(2.92)
Germany	19800101	20150630	0.04	0.02	-0.03	-0.16	0.04	-0.21	0.15	0.04	-0.10	0.18	0.07
			(3.17)	(4.82)	(-1.22)	(-4.21)	(4.26)	(-5.46)	(7.85)	(4.93)	(-2.76)	(4.08)	(1.85)
Singapore	20090101	20150731	0.02	0.01	-0.10	-0.36	0.04	-0.16	0.19	0.04	-0.03	0.06	-0.06
			(0.88)	(1.21)	(-2.28)	(-1.92)	(0.35)	(-0.84)	(3.81)	(1.30)	(-0.55)	(0.34)	(-0.81)
Sweden	19890102	20150630	0.06	0.03	0.02	-0.24	0.04	-0.04	0.23	0.05	-0.10	0.07	0.12
			(3.56)	(5.24)	(1.23)	(-3.33)	(2.13)	(-0.66)	(4.29)	(2.84)	(-1.46)	(0.92)	(1.75)
Italy	19910401	20150630	0.03	0.02	0.26	-0.19	0.03	0.19	0.16	0.07	0.26	0.07	0.00
			(1.78)	(3.27)	(9.10)	(-4.56)	(2.04)	(4.38)	(4.61)	(5.51)	(7.96)	(1.41)	(0.08)
Switzerland	19810101	20150630	0.05	0.02	-0.11	-0.12	0.02	-0.15	0.10	0.03	-0.17	0.05	0.06
			(3.78)	(5.29)	(-4.53)	(-2.86)	(1.61)	(-2.96)	(5.64)	(3.90)	(-6.01)	(0.90)	(1.83)
Norway	19921201	20150630	0.04	0.02	-0.03	-0.24	0.08	-0.30	0.18	0.03	-0.06	0.27	0.03
			(2.44)	(3.97)	(-1.49)	(-2.60)	(4.28)	(-4.48)	(4.85)	(2.73)	(-1.66)	(3.60)	(0.64)
Netherlands	19880101	20150731	0.05	0.02	-0.10	-0.21	0.05	-0.41	0.13	0.03	-0.28	0.32	0.18
			(3.85)	(5.37)	(-4.19)	(-3.60)	(5.13)	(-8.36)	(4.70)	(4.19)	(-9.17)	(5.62)	(4.69)
Denmark	19890201	20150630	0.04	0.03	0.02	-0.15	0.05	-0.13	0.14	0.03	-0.06	0.15	0.08
			(3.04)	(6.47)	(1.12)	(-2.14)	(2.75)	(-2.49)	(5.63)	(3.21)	(-1.91)	(2.62)	(2.15)
Belgium	19890703	20150630	0.05	0.03	0.00	-0.12	0.03	-0.11	0.12	0.04	-0.08	0.12	0.09
			(4.15)	(4.74)	(0.16)	(-2.26)	(2.34)	(-2.98)	(4.40)	(4.64)	(-2.67)	(2.73)	(2.22)
Spain	19901203	20121130	0.05	0.02	0.02	-0.20	0.03	0.08	0.15	0.06	0.10	-0.06	-0.08
			(2.46)	(2.83)	(0.78)	(-3.47)	(1.61)	(1.34)	(5.16)	(5.78)	(3.12)	(-0.93)	(-1.82)
New Zealand	19910401	20150630	0.04	0.03	0.03	-0.16	0.06	-0.02	0.13	0.03	-0.07	0.05	0.10
			(3.87)	(4.63)	(0.82)	(-3.65)	(3.99)	(-0.34)	(6.23)	(3.03)	(-2.52)	(0.68)	(2.03)
Finland	19910902	20150630	0.08	0.02	-0.06	-0.33	0.04	-0.25	0.24	0.04	-0.20	0.19	0.14
			(2.49)	(4.04)	(-2.91)	(-4.20)	(3.16)	(-6.50)	(6.10)	(4.86)	(-6.34)	(4.03)	(3.82)
Austria	19850101	20150630	0.04	0.03	-0.11	-0.17	0.03	-0.09	0.15	0.03	-0.06	-0.03	-0.06
			(2.74)	(5.17)	(-5.45)	(-4.89)	(3.27)	(-2.13)	(9.20)	(5.14)	(-2.22)	(-0.58)	(-1.86)
Ireland	19850101	20150731	0.07	0.04	0.07	-0.27	0.03	-0.02	0.18	0.04	0.01	0.09	0.06
			(4.92)	(6.16)	(4.95)	(-4.19)	(1.93)	(-0.19)	(5.58)	(3.44)	(0.36)	(0.96)	(1.34)
Portugal	19930802	20150630	0.03	0.03	0.16	-0.21	-0.02	0.22	0.11	0.08	0.11	-0.06	0.05
			(1.94)	(2.88)	(4.75)	(-3.24)	(-0.46)	(2.45)	(3.73)	(4.94)	(3.03)	(-0.60)	(0.91)
Total	19800101	20181231	0.05	0.03	0.02	-0.19	0.04	-0.09	0.16	0.04	-0.04	0.10	0.06
			(15.40)	(21.07)	(3.04)	(-15.49)	(10.67)	(-5.08)	(24.34)	(15.27)	(-4.63)	(5.42)	(6.23)

Table 1.10. Flights to Quality and Daily Momentum Returns in the US

This table reports daily return regressions of value-weighted US momentum decile portfolios on Flight to Quality (FTQ) and Flight to Risk (FTR) events as well as other factors. Consistent with Boudry et al. (2019), daily FTQ events are identified as $I\left\{r_t^b > \kappa\sigma_t^b\right\} \times I\left\{r_t^e < -\kappa\sigma_t^e\right\}$ where r_t^b and r_t^e are daily bond and stock market returns, respectively. κ is 1.5 following also Boudry et al. (2019). σ_ts are estimated with a one-sided normal kernel density with a bandwidth of 250 days that skips nearest 5 days to exclude potential influence by the FTQ events. FTR events are identified similarly with opposite inequalities. The daily benchmark stock and bond (10-year government) market returns are from Datastream. Daily momentum decile (and the winner-minus-loser) returns are regressed on FTQ and FTR events without other factors (first partition) and with Fama and French (2015) five factors (second and third partitions) from Kenneth French. Both estimates and their Newey–West t-statistics (parentheses) are reported.

Variable	Lose	2	3	4	5	6	7	8	9	Win	WML
Intercept	0.018	0.043	0.048	0.051	0.046	0.048	0.049	0.056	0.051	0.067	0.048
	(0.97)	(2.99)	(3.96)	(4.73)	(4.64)	(5.12)	(5.33)	(6.00)	(5.15)	(5.19)	(2.86)
FTQ	-4.249	-3.679	-3.281	-3.163	-3.034	-2.920	-2.804	-2.829	-2.956	-3.363	0.887
	(-11.18)	(-11.16)	(-12.49)	(-12.32)	(-12.63)	(-11.58)	(-11.94)	(-11.35)	(-11.75)	(-11.40)	(2.60)
FTR	3.989	3.524	2.988	2.670	2.539	2.367	2.177	2.112	2.140	2.674	-1.315
	(9.99)	(10.88)	(12.43)	(15.37)	(13.29)	(17.40)	(18.00)	(17.98)	(19.62)	(14.07)	(-3.32)
Intercept	-0.035	-0.015	-0.010	-0.007	-0.010	-0.007	-0.006	0.001	-0.002	0.026	0.062
	(-3.34)	(-2.04)	(-1.83)	(-1.51)	(-2.62)	(-2.19)	(-1.70)	(0.38)	(-0.43)	(4.06)	(4.12)
MKTRF	1.244	1.145	1.058	1.028	0.999	0.980	0.982	0.992	1.016	1.077	-0.167
	(51.19)	(54.40)	(72.14)	(88.13)	(69.31)	(114.60)	(153.49)	(100.67)	(75.58)	(54.66)	(-4.13)
SMB	0.371	0.147	0.067	0.021	0.010	-0.009	-0.011	-0.009	0.072	0.276	-0.096
	(10.25)	(5.16)	(2.72)	(1.18)	(0.70)	(-0.56)	(-1.18)	(-0.56)	(4.65)	(11.16)	(-1.81)
HML	0.890	0.614	0.445	0.320	0.258	0.133	0.037	-0.040	-0.120	-0.406	-1.296
	(9.08)	(10.85)	(11.16)	(11.07)	(8.32)	(6.59)	(2.46)	(-2.09)	(-5.43)	(-11.64)	(-10.49)
RMW	-0.530	-0.039	0.139	0.199	0.255	0.251	0.235	0.242	0.191	-0.171	0.359
	(-6.34)	(-0.71)	(3.61)	(6.18)	(7.88)	(10.47)	(12.45)	(11.66)	(7.45)	(-3.60)	(3.07)
CMA	-0.915	-0.466	-0.174	-0.022	0.047	0.188	0.278	0.344	0.219	-0.103	0.813
	(-9.33)	(-7.64)	(-4.14)	(-0.67)	(1.03)	(6.61)	(12.79)	(12.66)	(6.31)	(-1.85)	(5.82)
FTQ	-0.353	-0.256	-0.465	-0.081	-0.122	0.315	0.113	0.222	0.052	0.090	0.444
	(-1.18)	(-0.76)	(-1.65)	(-0.58)	(-1.29)	(2.55)	(0.72)	(1.82)	(0.56)	(0.60)	(1.13)
FTQ×MKTRF	-0.143	-0.094	-0.155	-0.010	-0.022	0.123	0.026	0.071	0.022	0.058	0.201
	(-1.35)	(-0.73)	(-1.43)	(-0.19)	(-0.65)	(2.85)	(0.44)	(1.57)	(0.75)	(1.03)	(1.40)
FTR	-0.870	-0.987	-0.724	-0.162	-0.494	0.114	0.407	0.337	0.394	0.357	1.227
	(-2.02)	(-2.51)	(-2.61)	(-0.61)	(-1.19)	(0.64)	(3.24)	(1.70)	(2.46)	(1.36)	(2.08)
FTR×MKTRF	0.446	0.544	0.391	0.113	0.235	-0.027	-0.205	-0.204	-0.260	-0.210	-0.656
	(2.36)	(3.04)	(3.16)	(0.93)	(1.23)	(-0.37)	(-4.09)	(-2.36)	(-3.86)	(-1.78)	(-2.52)

Table 1.11. Flights to Quality and Monthly Momentum Returns Around the World

This table reports monthly return regressions of momentum portfolios on Flight to Quality (FTQ) and Flight to Risk (FTR) variables by country. Consistent with the daily FTQ events identified following Boudry et al. (2019), monthly FTQ (FTR) variables are one in month t if one or more daily FTQ (FTR) events are identified in month t. Similarly, monthly NFTQ (NFTR) variables in month t count the number of daily FTQ (FTR) events in month t. Monthly winner-minus-loser returns from momentum quintiles or terciles (†) are regressed on monthly FTQ and FTR (first partition) or NFTQ and NFTR (second partition) variables. Countries have different starting points as the FTQ and FTR identification requires Datastream stock and bond (10-year government) market returns simultaneously. Both estimates and their t-statistics (parentheses) are reported.

Country	Start	Intercept		FTQ		FTR		Intercept		NFTQ		NFTR	
United States	195007	0.007	(1.92)	0.025	(2.67)	-0.019	(-1.85)	0.008	(2.37)	0.011	(2.03)	-0.016	(-2.08)
Canada	195101	0.021	(3.04)	-0.010	(-0.54)	0.002	(0.08)	0.019	(2.77)	0.003	(0.23)	0.002	(0.11)
Japan	198312	0.002	(0.55)	-0.003	(-0.28)	0.002	(0.19)	0.002	(0.60)	0.002	(0.26)	-0.005	(-0.55)
United Kingdom	197912	0.006	(1.80)	0.026	(2.64)	-0.014	(-1.30)	0.007	(2.20)	0.011	(1.94)	-0.011	(-1.36)
Australia	198702	0.010	(2.87)	0.002	(0.20)	-0.011	(-0.92)	0.010	(2.99)	0.008	(1.26)	-0.022	(-2.20)
France	198501	0.002	(0.50)	0.020	(1.68)	-0.032	(-2.43)	0.001	(0.34)	0.013	(2.01)	-0.021	(-2.26)
Germany	198012	0.007	(1.46)	0.017	(1.36)	-0.013	(-0.91)	0.008	(1.68)	0.009	(1.52)	-0.014	(-1.38)
Singapore	200812	0.006	(0.93)	0.002	(0.10)	-0.030	(-0.60)	0.006	(0.93)	0.002	(0.13)	-0.030	(-0.60)
Sweden	198012	0.012	(1.48)	-0.007	(-0.32)	-0.034	(-1.47)	0.012	(1.56)	0.016	(1.23)	-0.058	(-2.96)
Italy	199103	0.006	(0.95)	0.024	(1.16)	-0.023	(-1.08)	0.005	(0.83)	0.021	(1.47)	-0.015	(-0.91)
Switzerland	198012	0.005	(1.32)	0.022	(1.76)	-0.016	(-1.33)	0.006	(1.60)	0.019	(2.97)	-0.028	(-2.99)
Norway	199211	0.008	(1.28)	0.032	(2.01)	-0.032	(-1.85)	0.007	(1.19)	0.022	(2.17)	-0.020	(-1.50)
Netherlands	198712	0.009	(1.40)	0.019	(1.17)	-0.025	(-1.40)	0.006	(1.06)	0.011	(1.26)	-0.005	(-0.39)
Denmark	198901	0.010	(1.56)	0.017	(0.90)	0.006	(0.25)	0.011	(1.84)	0.004	(0.34)	0.004	(0.23)
Belgium†	199006	-0.001	(-0.14)	-0.005	(-0.33)	0.010	(0.58)	0.000	(0.06)	-0.012	(-1.34)	0.011	(0.98)
Spain†	199011	0.003	(0.48)	0.001	(0.07)	-0.020	(-1.17)	0.002	(0.41)	0.004	(0.38)	-0.015	(-1.36)
New Zealand†	199103	0.007	(2.04)	0.002	(0.21)	-0.017	(-1.19)	0.007	(2.05)	0.005	(0.70)	-0.021	(-1.63)
Finland†	199108	0.010	(1.54)	-0.001	(-0.06)	-0.004	(-0.26)	0.013	(2.02)	0.000	(0.02)	-0.017	(-1.34)
Austria†	198412	0.002	(0.49)	0.020	(1.79)	-0.024	(-1.98)	0.001	(0.37)	0.011	(1.77)	-0.013	(-1.41)
Ireland†	198412	0.012	(2.29)	-0.001	(-0.03)	0.004	(0.24)	0.012	(2.22)	0.003	(0.25)	0.004	(0.32)
Portugal†	199307	0.013	(1.90)	0.039	(1.93)	-0.037	(-1.58)	0.014	(2.03)	0.018	(1.58)	-0.025	(-1.36)
All		0.008	(6.69)	0.012	(3.62)	-0.015	(-4.07)	0.008	(6.94)	0.009	(4.61)	-0.014	(-5.11)

Table 1.12. Difference Tests for Time-Varying Betas of Momentum Portfolios

This table reports the F-statistics that test the difference among the time-varying betas estimated in Table 1.3. The time-series regression specification by Daniel and Moskowitz (2016) is used for the quintile or tercile winner-minus-loser portfolios by country. The non-bear beta β, the bear-down beta β_{BD}, and the bear-up beta β_{BU} are estimated using the bear-down and bear-up indicators I_{BD} and I_{BU} and the significance of their difference is tested. The asterisks indicate the significance at the 10% (*), 5% (**), and 1% (***), respectively.

Country	$\beta = \beta_{BD}$	$\beta = \beta_{BU}$	$\beta_{BD} = \beta_{BU}$
United States	9.93***	40.36***	6.32**
Canada	39.54***	39.54***	7.16***
Japan	37.47***	70.88***	2.71
United Kingdom	2.83*	15.22***	4.15**
Australia	6.77***	26.50***	2.08
France	6.00**	28.86***	7.31***
Germany	1.90	9.00***	2.30
Hong Kong	29.58***	36.39***	0.01
Singapore	12.91***	35.63***	1.52
Sweden	27.26***	52.71***	1.50
Israel	0.17	15.75***	19.23***
Italy	0.30	0.91	0.14
Switzerland	20.54***	16.19***	1.39
Norway	6.59**	3.55*	1.02
Netherlands	11.33***	25.40***	2.36
Denmark	0.36	9.92***	6.53**
Belgium	32.33***	32.62***	0.27
Spain	4.15**	2.48	0.53
New Zealand	2.49	10.61***	3.29*
Finland	14.74***	27.99***	1.00
Austria	2.96*	15.22***	4.69**
Ireland	6.26**	17.05***	1.59
Portugal	1.00	9.99***	9.65***

Table 1.13. Further Backward Market Betas of Momentum Portfolios

This table, in addition to Table 1.4, reports the equal-weighted average of market beta (β) of winner-minus-loser (WML) portfolios for three different time periods: (i) normal periods, (ii) bear down market periods, and (iii) bear up market periods. As in Table 1.4, market beta is measured from rolling regressions but three additional horizons for each stock: from -59 to -48 months ($\beta_{(-59,-48)}$), from -47 to -36 months ($\beta_{(-47,-36)}$), from -35 to -24 months ($\beta_{(-35,-24)}$), relative to the portfolio formation period. The winner and the loser portfolios are the top and bottom quintile—or tercile if less than 300 stocks are available—portfolios. The asterisks indicate the significance at 10% (*), 5% (**), and 1% (***), respectively.

	$\beta_{(-59,-48)}$			$\beta_{(-47,-36)}$			$\beta_{(-35,-24)}$		
Country	$I_B = 0$	$I_{BD} = 1$	$I_{BU} = 1$	$I_B = 0$	$I_{BD} = 1$	$I_{BU} = 1$	$I_B = 0$	$I_{BD} = 1$	$I_{BU} = 1$
Japan	-0.0162	-0.0830**	-0.0779***	0.0368**	-0.0617**	-0.1194***	-0.0273	-0.0886***	-0.0935***
United Kingdom	-0.0245**	-0.1176**	-0.0156	-0.0264**	-0.0497	0.0062	0.0428***	0.0132	-0.0513
Australia	-0.0604***	-0.0598	-0.1364***	-0.0074	-0.1341***	-0.1284***	0.0084	-0.1445***	-0.0701**
France	0.0108	0.0210	-0.0226	0.0284	-0.1224**	-0.0902***	-0.0094	-0.1300**	-0.1596***
Germany	-0.0307	-0.1541***	-0.2346***	-0.1152***	-0.2366***	-0.3781***	-0.0860***	-0.2188***	-0.2756***
Hong Kong	0.0639***	-0.1317*	0.0136	-0.0173	-0.1143*	-0.0626	0.0174	-0.1302**	-0.0682
Singapore	-0.1336***	-0.0918*	0.0313	-0.0368	-0.1659**	-0.0070	-0.1270***	-0.1224**	-0.0253
Sweden	0.0222	-0.0681	-0.1157**	-0.0852**	-0.3840***	-0.2026***	0.0958***	-0.3549***	-0.1289
Israel	-0.0692***	-0.0931	-0.0804**	-0.0200	-0.1411**	-0.1067**	-0.0692***	-0.1469***	-0.2092***
Italy	-0.0699*	0.1239*	-0.0213	-0.1170***	0.0454	-0.1388***	-0.1058***	0.0167	-0.0546
Switzerland	-0.0370	-0.2108***	-0.1485***	-0.0091	-0.1421**	-0.1646***	-0.0118	-0.0308	0.0373
Norway	0.1553***	-0.1837*	-0.3427***	0.0662**	-0.0877	-0.2322***	0.0762***	-0.4191***	-0.3560***
Netherlands	0.0372	-0.0789	-0.0086	0.1956	-0.0424	0.0416	0.1500***	-0.2429***	-0.1505***
Denmark	-0.0161	-0.0560	-0.0489	0.0994***	-0.1979**	-0.1649***	-0.0976***	-0.0133	-0.1504***
Belgium	0.1353***	-0.1425***	-0.0993**	0.1478***	0.0949*	-0.0016	-0.0220	-0.1589**	-0.1705***
Spain	-0.0933***	0.0289	-0.1602**	-0.0585	-0.1901**	-0.2158***	-0.1595***	-0.2754***	-0.1290**
New Zealand	0.0115	-0.0382	0.0009	-0.0240	0.1293**	-0.0134	-0.0284	0.0166	-0.0689
Finland	0.0428	-0.1985***	-0.0724**	0.0755***	-0.3443***	-0.2081***	0.0659**	-0.3079***	-0.1707***
Austria	0.2209***	-0.1523	-0.0320	0.0438	-0.2177	0.1132*	0.1498***	0.0367	-0.0013
Ireland	-0.0540	-0.2679***	-0.1116**	-0.0610**	-0.1938**	-0.2805***	-0.0479	-0.0887	-0.1630***
Portugal	-0.0596	-0.0287	-0.0580	0.1258***	-0.1265	-0.0928**	0.0765*	-0.3077***	-0.1671**

Table 1.14. Flights to Quality and Monthly Momentum Returns Around the World

This table reports monthly return regressions of momentum portfolios on Flight to Quality (FTQ), Flight to Risk (FTR), and their respective interaction terms with the bear market indicator (I_B) by country. Consistent with the daily FTQ events identified following Boudry et al. (2019), monthly FTQ (FTR) variables are one in month t if one or more daily FTQ (FTR) events are identified in month t. Similarly, monthly NFTQ (NFTR) variables in month t count the number of daily FTQ (FTR) events in month t. The bear market indicator I_B is 1 if the 2-year excess market return is negative or 0 otherwise. Monthly winner-minus-loser returns from momentum quintiles or terciles (†) are regressed on monthly FTQ, FTR, and their I_B-interaction terms (first partition) or NFTQ, NFTR, and their I_B-interaction terms (second partition). Countries have different starting points as the FTQ and FTR identification requires Datastream stock and bond (10-year government) market returns simultaneously. Both estimates and their t-statistics (parentheses) are reported.

Country	Intercept	FTQ	FTR	FTQ · I_B	FTR · I_B	Intercept	NFTQ	NFTR	NFTQ · I_B	NFTR · I_B
United States	0.007 (1.95)	0.021 (2.03)	-0.012 (-1.03)	0.019 (0.80)	-0.023 (-1.05)	0.008 (2.39)	0.008 (1.34)	-0.012 (-1.29)	0.026 (1.11)	-0.021 (-0.95)
Canada	0.021 (3.03)	0.003 (0.13)	0.012 (0.46)	-0.036 (-0.96)	-0.036 (-0.75)	0.020 (2.93)	0.010 (0.73)	0.009 (0.43)	-0.044 (-1.28)	-0.036 (-0.77)
Japan	0.002 (0.64)	-0.022 (-1.60)	0.018 (1.08)	0.047 (2.29)	-0.038 (-1.67)	0.002 (0.61)	-0.008 (-0.80)	0.001 (0.05)	0.035 (1.78)	-0.020 (-0.93)
United Kingdom	0.006 (1.83)	0.017 (1.43)	0.012 (0.91)	0.027 (1.38)	-0.076 (-3.41)	0.007 (2.25)	0.005 (0.89)	0.005 (0.51)	0.034 (1.92)	-0.072 (-3.41)
Australia	0.009 (2.85)	0.016 (1.27)	-0.013 (-0.99)	-0.040 (-1.86)	-0.012 (-0.37)	0.010 (3.12)	0.014 (2.07)	-0.024 (-2.35)	-0.040 (-2.08)	-0.009 (-0.28)
France	0.003 (0.69)	-0.004 (-0.25)	0.011 (0.63)	0.063 (2.67)	-0.099 (-3.89)	0.002 (0.52)	0.004 (0.60)	0.006 (0.55)	0.051 (2.29)	-0.097 (-4.03)
Germany	0.009 (1.74)	0.012 (0.74)	0.016 (0.91)	0.011 (0.43)	-0.113 (-3.44)	0.010 (2.00)	0.005 (0.68)	0.007 (0.58)	0.015 (0.66)	-0.110 (-3.27)
Sweden	0.012 (1.54)	-0.016 (-0.63)	0.007 (0.25)	0.038 (0.77)	-0.109 (-2.37)	0.012 (1.67)	0.012 (0.91)	-0.038 (-1.64)	0.011 (0.25)	-0.064 (-1.46)
Italy	0.007 (1.14)	-0.008 (-0.27)	0.017 (0.52)	0.056 (1.44)	-0.066 (-1.61)	0.006 (1.02)	0.007 (0.39)	0.012 (0.53)	0.037 (1.06)	-0.065 (-1.72)
Switzerland	0.005 (1.35)	-0.010 (-0.72)	0.001 (0.07)	0.113 (4.36)	-0.068 (-2.50)	0.006 (1.48)	0.005 (0.72)	-0.017 (-1.58)	0.092 (3.61)	-0.042 (-1.48)
Norway	0.008 (1.41)	0.007 (0.38)	0.001 (0.03)	0.074 (2.16)	-0.067 (-2.00)	0.008 (1.39)	0.006 (0.46)	0.005 (0.30)	0.070 (2.02)	-0.073 (-2.27)
Netherlands	0.009 (1.44)	0.020 (0.99)	-0.022 (-0.91)	-0.006 (-0.17)	-0.007 (-0.20)	0.007 (1.18)	0.009 (0.91)	0.006 (0.40)	-0.001 (-0.02)	-0.035 (-1.03)
Denmark	0.011 (1.74)	-0.031 (-1.30)	0.062 (1.93)	0.120 (3.18)	-0.114 (-2.56)	0.011 (1.83)	-0.024 (-1.85)	0.051 (1.97)	0.128 (3.50)	-0.113 (-2.55)
Belgium†	0.000 (-0.06)	-0.027 (-1.46)	0.026 (1.07)	0.063 (2.04)	-0.033 (-0.99)	0.000 (0.09)	-0.028 (-2.65)	0.023 (1.60)	0.081 (2.69)	-0.038 (-1.27)
Spain†	0.004 (0.62)	-0.009 (-0.40)	0.004 (0.18)	0.012 (0.40)	-0.053 (-1.59)	0.003 (0.60)	0.002 (0.16)	-0.005 (-0.37)	0.001 (0.02)	-0.042 (-1.40)
New Zealand†	0.007 (2.07)	-0.011 (-0.97)	-0.003 (-0.21)	0.048 (2.27)	-0.058 (-1.66)	0.007 (1.96)	-0.002 (-0.22)	-0.010 (-0.72)	0.040 (1.97)	-0.046 (-1.29)
Finland†	0.011 (1.61)	-0.002 (-0.07)	0.013 (0.58)	-0.004 (-0.11)	-0.027 (-0.81)	0.013 (2.13)	-0.001 (-0.11)	-0.014 (-0.98)	-0.001 (-0.04)	0.000 (0.01)
Austria†	0.002 (0.53)	0.006 (0.47)	-0.002 (-0.16)	0.046 (2.01)	-0.064 (-2.63)	0.002 (0.42)	0.005 (0.69)	0.003 (0.25)	0.044 (1.99)	-0.070 (-3.02)
Ireland†	0.013 (2.39)	-0.006 (-0.28)	0.004 (0.16)	0.013 (0.37)	-0.002 (-0.05)	0.013 (2.31)	0.001 (0.10)	0.004 (0.27)	0.004 (0.14)	-0.003 (-0.08)
Portugal†	0.015 (2.09)	-0.015 (-0.43)	0.016 (0.41)	0.103 (2.27)	-0.087 (-1.71)	0.015 (2.11)	-0.003 (-0.17)	0.002 (0.08)	0.093 (2.28)	-0.073 (-1.73)
All	0.008 (6.91)	0.000 (0.03)	0.006 (1.40)	0.032 (4.89)	-0.055 (-7.62)	0.008 (7.14)	0.004 (1.67)	-0.001 (-0.43)	0.027 (4.38)	-0.047 (-6.86)

Figure 1.1. Monthly Market Betas of Daily Momentum and Other Factors

This figure shows the market betas of the momentum, size, value, profitability, and investment factors from June 1964 to January 2020. Each daily factor is regressed on the daily market factor in each month. The one-year moving averages of these betas are plotted. Shaded areas indicate US recessions by NBER.

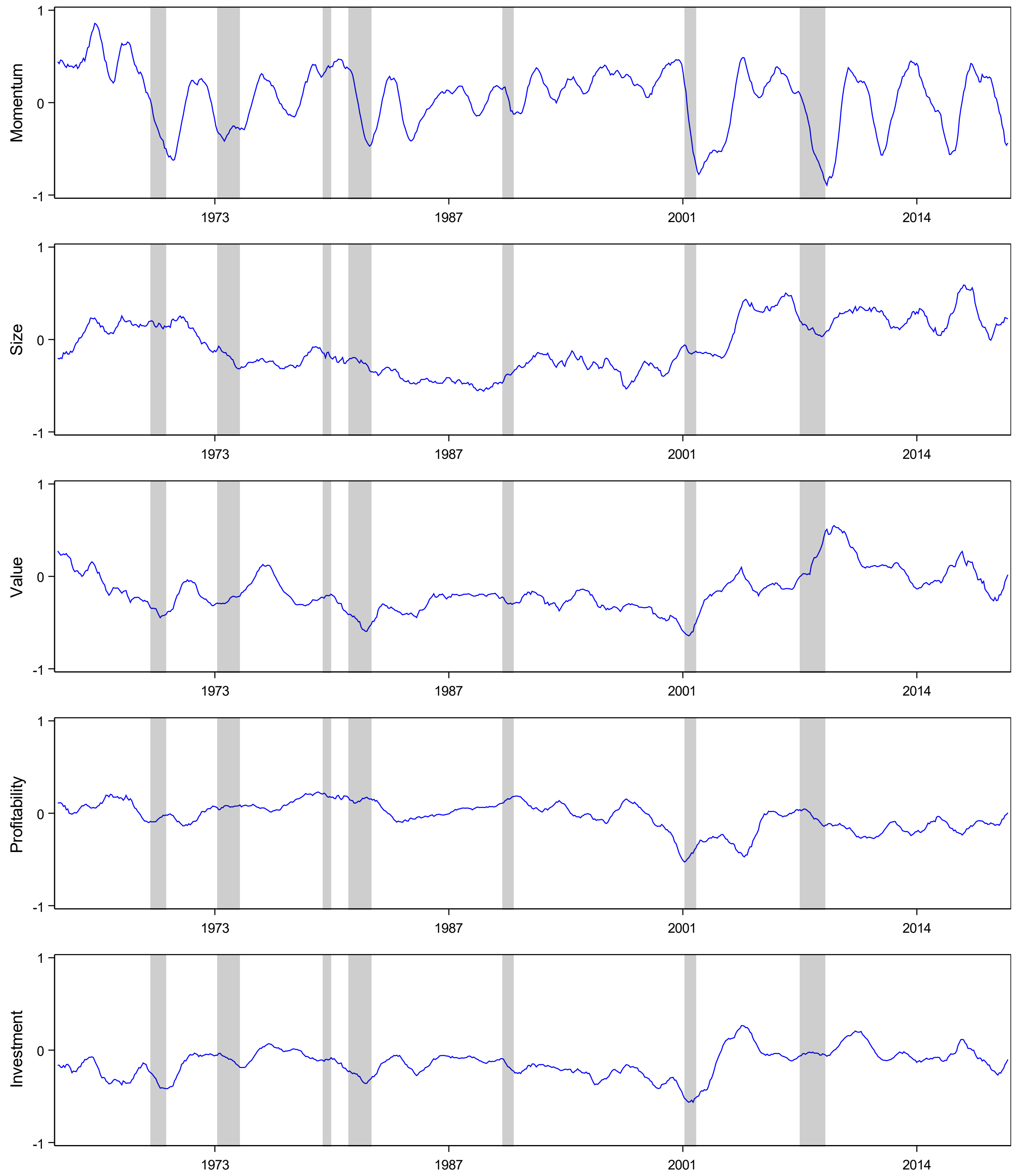


Figure 1.2. Backward, Central, Forward Betas of WML Portfolios

This figure reports equal-weighted average of market beta (β) of winner-minus-loser (WML) portfolios for three different time periods: (i) normal periods, (ii) bear down market periods, and (iii) bear up market periods. Market beta (β) is measured from rolling regressions for three different horizons for each stock: from -59 to -48 months $\left(\beta_{(-59,-48)}\right)$ (upper left), from -23 to -12 months $\left(\beta_{(-23,-12)}\right)$ (upper right), from -5 to 6 months $\left(\beta_{(-5,6)}\right)$ (lower left), and from 13 to 24 months $\left(\beta_{(13,24)}\right)$ (lower right) relative to the portfolio formation period. I define the bear market if the 2-year excess market return is negative, or 0 otherwise. I define the down (up) market if the contemporaneous excess market return is negative (positive), or 0 otherwise. The loser and winner portfolios are the bottom and top quintile—or tercile if less than 300 stocks are available—portfolios. This figure includes 21 MSCI developed countries.

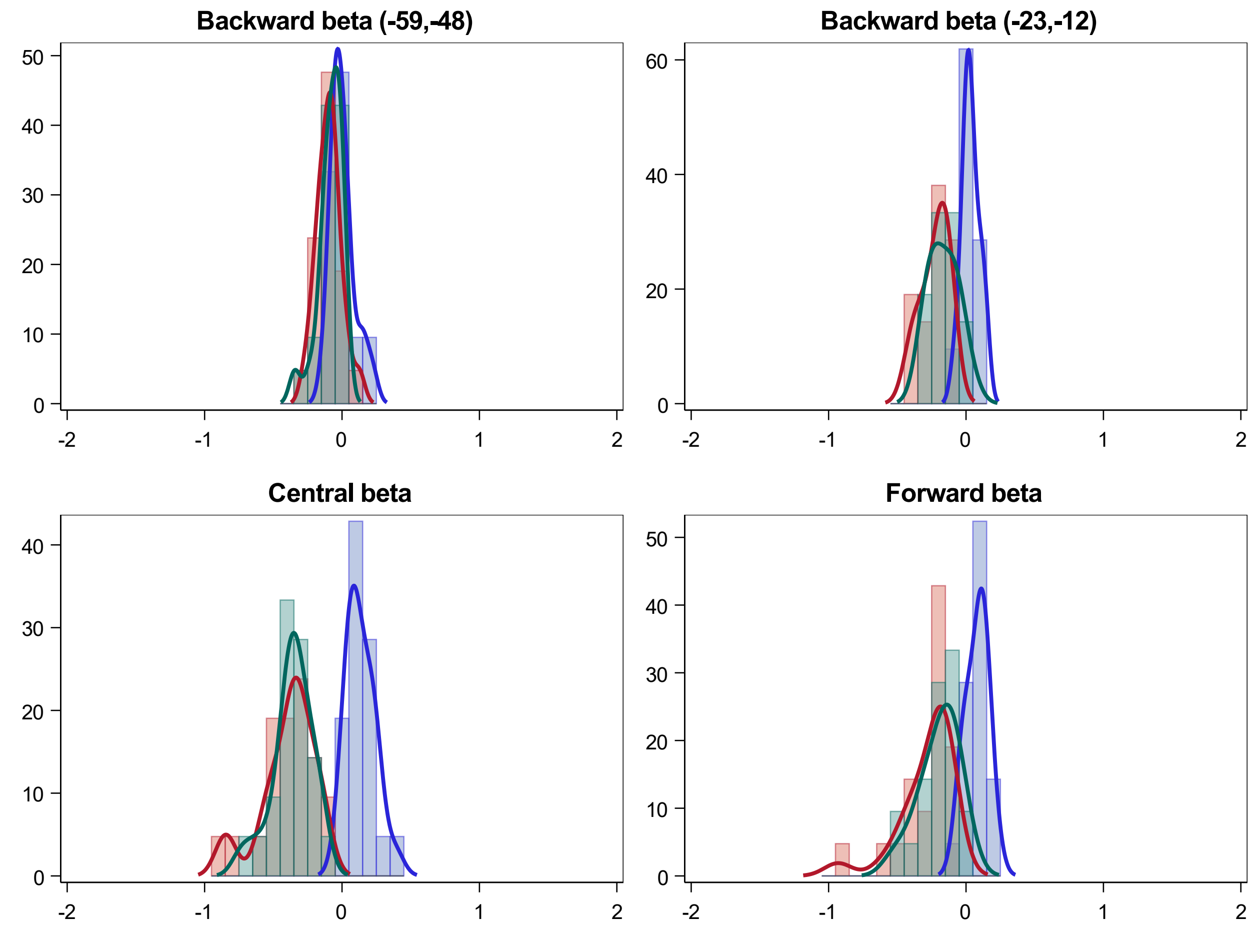
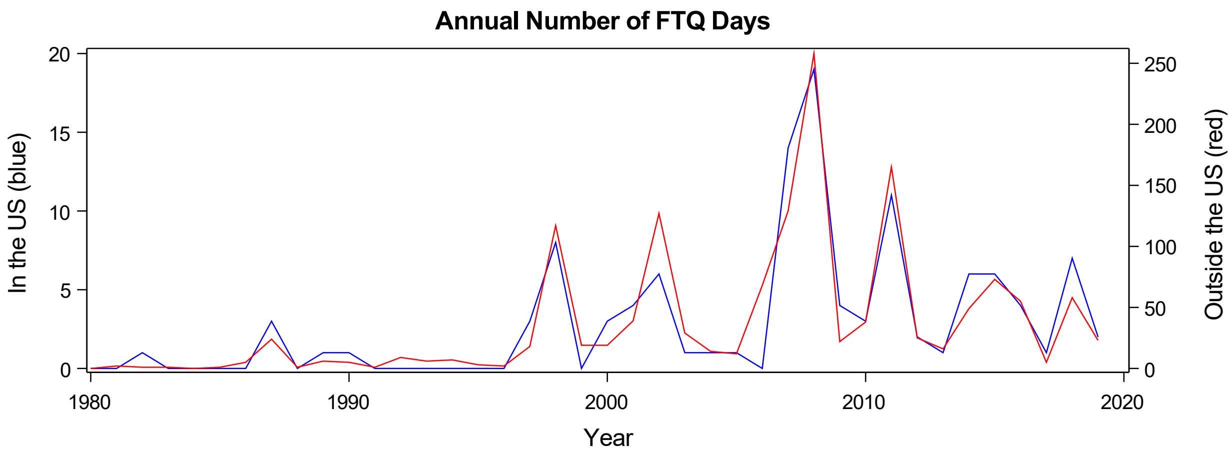
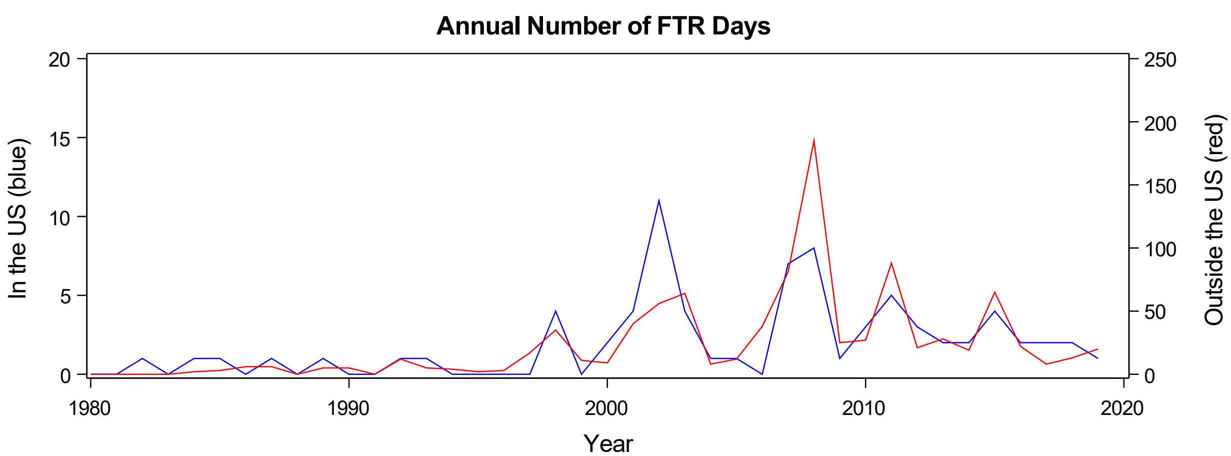


Figure 1.3. Annual Number of FTQs and FTRs from Stock–Bond Market Disagreement

This figure reports the annual distribution of FTQ and FTR days that are defined using the disagreement between stock and bond markets, following the literature (Baele et al., 2019; Boudry et al., 2019). FTQ (FTR) days in the first (second) plot are identified as the periods with large negative (positive) stock market returns and large positive (negative) bond market returns together. Each plot displays the annual number of FTQ or FTR days in the United States (blue) and outside the United States (red).





Chapter 2

Which Volatility Drives the Anomaly? Cash Flow Versus Discount Rate

The idiosyncratic volatility puzzle in the cross-section has been investigated over the past decade. Economic theories as well as our intuitions predict zero or positive association between returns and idiosyncratic volatilities. Surprisingly, however, most findings are the opposite—returns and idiosyncratic volatilities apparently exhibit a negative relationship. According to a substantial body of empirical finance literature, this counterintuitive anomaly is not only consistent after controlling other factors, but also robust to a variety of research designs (Ang et al., 2006, 2009; Jiang et al., 2009; Chen et al., 2012; Fink et al., 2012; Guo et al., 2014). In particular, it has attracted a lot of interest from researchers because it is hard to reconcile these findings with traditional asset pricing paradigms—that is to say, this anomaly violates a high-risk high-return principle. Does the market compensate investors for low idiosyncratic volatilities?

To fit this puzzle, previous papers have emphasized the importance of behavioral motives such as investor sentiment, conditional heteroskedasticity, return reversal, and skewness preference (Fu, 2009; Huang et al., 2010; Bali et al., 2011; Stambaugh et al., 2015; Cao and Han, 2016; Egginton and Hur, 2018). However, none of the papers has paid attention to the role of cash flow and discount rate news, which is a fundamental determinant of both returns and volatilities. Studies concerning news decomposition methods likewise have paid no attention to the idiosyncratic volatility anomaly notwithstanding its importance in the realm of asset pricing. Instead, they have highlighted the role of systematic risks to account for the cross-section of returns (Campbell and Vuolteenaho, 2004; Campbell et al., 2010, 2013; Yeh et al., 2015).

Many studies have employed news decomposition methods to dissect stock returns into cash flow and discount rate news at an aggregate level since Campbell and Shiller (1988), but one of the most important papers is Vuolteenaho (2002), which decomposes stock returns at an individual level rather than at an aggregate level using an accounting identity framework and a panel VAR model. Unlike other papers that emphasize the role of discount rate news in determining stock returns at an aggregate level, the paper concludes that stock returns at a firm level are mostly determined by cash flow news. Chen et al. (2013) disaggregate returns by altering a predictive regression method with an implied cost of capital approach, but their findings also indicate that cash flow news play a significant role, which is consistent with the former paper as well. As these two news components fundamentally determine both stock returns and volatilities, one is able to better figure out the structural forces that drive the aforementioned idiosyncratic volatility puzzle by scrutinizing the news components.

In this paper, I reexamine the existence of the idiosyncratic volatility anomaly at the quarterly horizon and investigate the relative importance of cash flow and discount rate news in driving this anomaly. I estimate quarterly idiosyncratic volatilities using daily returns. I also estimate a panel VAR model with quarterly CRSP and Compustat data obtained from WRDS. Following the volatility literature (Fu, 2009; Guo et al., 2014), both discount rate and cash flow news volatilities are computed using an out-of-sample EGARCH model. Two decile portfolios sorted on respective volatilities are created. While the average return of the arbitrage portfolio constructed with discount rate news volatilities is insignificant, the average return of the arbitrage portfolio constructed with cash flow news volatilities is positive and significant after controlling the market factor (α=1.52%/quarter, t-statistic=2.24) or Fama–French factors (α=1.21%/quarter, t-statistic=1.81). These results indicate that cash flow news volatilities rather than discount rate news volatilities mainly drive the idiosyncratic volatility anomaly.

Firstly, these findings are consistent with the results of Vuolteenaho (2002) as cash flow news components rather than discount rate news counterparts mostly drive the cross-section of returns. Secondly, the findings are similar to the results of Campbell and Vuolteenaho (2004) as only the systematic or idiosyncratic cash flow risks are priced in the cross-section, while the systematic or idiosyncratic discount rate risks are not. Lastly but not leastly, these findings are simultaneously consistent with existing papers about the volatility anomaly (Ang et al., 2006, 2009) and those about the skewness preference (Boyer et al., 2010; Bali et al., 2011). Cash flow news components are right-skewed, while discount rate news counterparts are not—correspondingly, idiosyncratic cash flow volatilities are priced negatively, while idiosyncratic discount rate counterparts are not. The findings suggest that the idiosyncratic volatilities are priced in a meaningful way only when they contain information about skewness. As idiosyncratic cash flow volatilities convey more information about return skewness, they tend to be priced in the cross-section. Unlike these volatilities, idiosyncratic discount rate volatilities less deliver such information, so they tend not to be priced. Collectively, this evidence supports the argument concerning investors' lottery preference.

Subsequent sections are organized in the following manner. Section I recalls the literature related to the idiosyncratic volatility anomaly and the news decomposition methodology. Section 2.2 describes economic models employed. Section 2.3 illustrates the data analyzed. Section 2.4 demonstrates major findings. Section 2.5 concludes this paper.

2.1 Literature Review

Many asset pricing theories articulate that the relation between idiosyncratic risks and subsequent returns should be insignificant or at least positive (Merton, 1987; Xu and Malkiel, 2004). Early findings based on time-series data display the insignificant or positive relation between them at the aggregate level (Longstaff, 1989; Lehmann, 1990; Goyal and Santa-Clara, 2003). However, recent cross-sectional studies exhibit the opposite relation between them at the individual level. In the first subsection, I introduce several papers investigating the idiosyncratic volatility anomaly, which is the main issue I explore. In the second subsection, I introduce another group of papers employing news decomposition methods with which I dissect stock returns.

2.1.1 Cross-section of return and volatility

Fama and MacBeth (1973) use idiosyncratic volatilities as well as market betas in their return regression models and exhibit that the coefficients for idiosyncratic volatilities are insignificant in almost every specification. However, their research concentrates on market betas rather than idiosyncratic volatilities hence constructs market beta-sorted portfolios to mitigate measurement errors in market beta estimates. As a result, their coefficients for idiosyncratic volatilities are biased and inconsistent as idiosyncratic volatility estimates are subject to a measurement error problem.

In contrast, Ang et al. (2006) concentrate on idiosyncratic volatilities instead of market betas and show the negative relation between idiosyncratic volatilities and subsequent returns based on cross-sectional analyses. They use daily excess returns and Fama–French three-factor model to compute monthly idiosyncratic volatilities and investigate the performance of idiosyncratic volatility-sorted portfolios. The authors confirm that the arbitrage portfolio constructed with idiosyncratic volatilities outperforms even after considering both risks and characteristics. In addition, the outperformance survives in different L/M/N specifications and over various subsamples. This phenomenon is globally observed in the financial markets of G7 countries as well (Ang et al., 2009).

Since monthly idiosyncratic volatilities are positively autocorrelated, Fu (2009) introduces an EGARCH model instead to estimate idiosyncratic volatilities and uses monthly returns rather than daily counterparts. He demonstrates the positive relation between expected idiosyncratic volatilities and subsequent returns. This positive relation is consistent with the theoretical prediction that suggests the positive risk premium for an idiosyncratic volatility under an underdiversification problem, which is more realistic according to empirical findings (Campbell et al., 2001). However, other subsequent papers point out a potential look-ahead bias and demonstrate the opposite relation between returns and out-of-sample EGARCH volatilities (Fink et al., 2012; Guo et al., 2014), which are free from the aforementioned bias.

In order to explain this counterintuitive relation between idiosyncratic volatilities and returns, previous research has adopted skewness preference (Barberis and Huang, 2008; Boyer et al., 2010; Bali et al., 2011; Egginton and Hur, 2018), liquidity cost (Han and Lesmond, 2011), return reversal (Huang et al., 2010), January effect (Huang et al., 2011), arbitrage asymmetry and investor sentiment (Stambaugh et al., 2015; Cao and Han, 2016), etc. Hou and Loh (2016) assess the explanatory power of these candidates and confirm that, though some explanations such as skewness preference and market friction partly justify the volatility puzzle, a significant portion of this anomaly remains unexplained.

2.1.2 News decomposition of return and volatility

Inherently, stock returns are driven by both cash flow news components and discount rate news counterparts. This is intuitive since the price is the expected value of discounted payoff (i.e. $p = \mathrm{E}[mx]$). Campbell and Shiller (1988) firstly propose the way to disentangle these two components by applying both log-linearized dividend-price ratio model and a VAR model. They relate returns to dividend-price ratios as well as dividend growths and use annual time-series data at the aggregate level. Though the authors confirm the relative importance of cash flow and discount rate news in their findings, they focus not on returns but on the behavior of dividend-price ratios.

Other early studies also investigate cash flow and discount rate news at the aggregate level (Campbell, 1991; Campbell and Ammer, 1993). These studies report that stock returns are largely determined by discount rate news components rather than cash flow news counterparts at the aggregate level. On the other hand, (Vuolteenaho, 2002) suggest another way to decompose by adopting log-linearized book-to-market ratio model and a VAR model. The author relates returns to book-to-market ratio as well as return on equity and use annual panel at the individual level. Unlike the former evidence from the aggregate level, the firm-level result indicates that stock returns are mainly driven by cash flow news components instead of discount rate news counterparts. Subsequent papers applying different methods such as a Feltham–Ohlson clean surplus relation and an implied cost of capital approach provide consistent results as well (Callen and Segal, 2004; Chen et al., 2013).

Since both aggregate level returns and firm-level counterparts can be disaggregated, the interaction among the components has also been examined. Campbell and Vuolteenaho (2004) decompose the aggregate level data into cash flow and discount rate news, and introduce two respective systematic risks. For individual returns, they use the two components to estimate cash flow betas (i.e. "bad" betas) and discount rate betas (i.e. "good" betas) separately. The authors show that value stocks tend to have high cash flow betas, and growth stocks tend to have high discount rate betas. They also show that small stocks tend to have high discount rate betas, while both small and large stocks tend to have comparable cash flow betas. Overall, this justifies the failure of CAPM after 1963 because "bad" cash flow betas are compensated more than "good" discount rate betas. Campbell et al. (2013) adopt a similar framework and compare the downturn of the early 2000s, which is largely driven by discount rate news, and that of the late 2000s, which is mainly driven by cash flow news.

Campbell et al. (2010) further disaggregate the firm-level data into cash flow and discount rate news. They use two aggregate level components CFm, DRm and two firm-level components CFi, DRi to estimate four different betas (i.e. CFi-CFm betas, DRi-CFm betas, CFi-DRm betas, DRi-DRm betas). The authors demonstrate that, while value stocks tend to have high CFi-CFm betas, growth stocks tend to have high CFi-DRm betas. In contrast, they show that two firm-level discount rate betas of value stocks and growth counterparts are not significantly different from each other.

Other details of news decomposition methods have also been studied by previous literature. Chen and Zhao (2009) point out several vulnerabilities of these VAR-based decomposition methods and explore some methodological remedies. Engsted et al. (2012) suggest another technical way to circumvent these issues. Cenedese and Mallucci (2016) decompose aggregate level international stock returns and report that the international returns are largely driven by cash flow shocks rather than discount rate counterparts. Lochstoer and Tetlock (2019) decompose size, book-to-market, profitability, investment, and momentum portfolio returns and show that cash flow news components drive these anomalies rather than discount rate news counterparts.

2.2 Economic Model

2.2.1 Idiosyncratic volatility

Firstly, I estimate idiosyncratic volatilities using Fama–French model to check whether the idiosyncratic volatility puzzle is consistent at a quarterly horizon or not. In detail, I estimate the following time-series regression repeatedly.

$$r_{itd} - r_{ftd} = \alpha_{it} + \beta_{it}\left(r_{mtd} - r_{ftd}\right) + s_{it}SMB_{td} + h_{it}HML_{td} + \varepsilon_{itd}.$$

The subscripts i, t and d stand for firm, quarter (or month) and day respectively. Overall, I follow the details of Fama and French (1993) for regression variables and Ang et al. (2006) for idiosyncratic volatilities. Each quarter (month), I estimate this time-series regression using daily return data and compute $\sqrt{\widehat{\mathrm{Var}}\left[\varepsilon_{itd}\right]}$ on a quarterly (monthly) basis recursively. I exclude idiosyncratic volatilities that are computed with less than 31 (11) daily observations to address an errors-in-variables issue.

Secondly, I sort stocks based on these idiosyncratic volatility estimates at the end of quarter (month) t and construct value-weighted quintile portfolios for the quarter (month) $t+1$, i.e. 3/3 (1/1) strategy of Jegadeesh and Titman (1993b). Portfolios are rebalanced each quarter (month). In addition, I construct zero investment portfolio by buying the first (i.e. the least volatile) portfolio and selling the fifth (i.e. the most volatile) one.

Thirdly, I measure the performance of those portfolios based on their historical returns.

$$r_{pt} - r_{ft} = \alpha_p + \beta_p \left(r_{mt} - r_{ft}\right) + s_p SMB_t + h_p HML_t + \varepsilon_{pt}.$$

For the portfolio p, I compute (i) sample statistics, (ii) CAPM statistics and (iii) Fama–French model statistics by using its time-series.

2.2.2 News decomposition

To decompose firm-level stock returns, I adopt the framework of Vuolteenaho (2002). Unlike the method of Campbell and Shiller (1988), this framework incorporates book-to-market ratio, return on equity and clean-surplus relation to disaggregate firm-level returns.

$$\begin{aligned}
\theta_t \approx & \sum_{j=0}^{\infty} \rho^j r_{t+1+j} - \sum_{j=0}^{\infty} \rho^j \left(e_{t+1+j} - f_{t+j}\right) \\
\Rightarrow r_t - \mathrm{E}_{t-1}\left[r_t\right] =& \Delta \mathrm{E}_t \left[\sum_{j=0}^{\infty} \rho^j \left(e_{t+j} - f_{t+j}\right)\right] + \kappa_t - \Delta \mathrm{E}_t \left[\sum_{j=1}^{\infty} \rho^j r_{t+j}\right] \\
=& N_{cf,t} - N_{r,t}.
\end{aligned}$$

For simplicity, I omit firm subscripts. The function $\mathrm{E}_t[\cdot]$ represents the expected value subject to the information set available at time t, i.e. $\mathrm{E}\left[\cdot|\Omega_t\right]$. The function $\Delta\mathrm{E}_t[\cdot]$ denotes the change of expectation at time t, i.e. $\mathrm{E}_t[\cdot] - \mathrm{E}_{t-1}[\cdot]$. The variables θ, r, e and f stand for log book-to-market ratio, log excess return, log return on equity and log interest rate, respectively. The coefficient ρ and the term κ stand for discount factor and approximation, respectively. Therefore, the returns are decomposed into two components, i.e. cash flow news N_{cf} and discount rate news N_r. In addition, this return decomposition implies the following variance decomposition simultaneously.

$$\mathrm{Var}\left[r_t - \mathrm{E}_{t-1}\left[r_t\right]\right] = \mathrm{Var}\left[N_{cf,t}\right] + \mathrm{Var}\left[N_{r,t}\right] - 2\mathrm{Cov}\left[N_{cf,t}, N_{r,t}\right].$$

In practice, one is able to decompose the returns by assuming VAR process for state variables. In particular, I assume the first-order VAR process rather than others.

$$\mathbf{z}_t = \mathbf{\Gamma}\mathbf{z}_{t-1} + \mathbf{u}_t.$$

The first element of the state vector $\mathbf{z}$ is r, i.e. $\mathbf{z} = \begin{pmatrix} r & \cdots \end{pmatrix}^\top$. This linear process implies the recursive structure hence the change of expectation can be obtained as well.

$$\Delta\mathrm{E}_t\left[\mathbf{z}_{t+j}\right] = \mathbf{\Gamma}^j\mathbf{u}_t.$$

Both discount rate news and cash flow news can be obtained by combining the change of expectation with the news decomposition above.

$$\begin{aligned} N_{r,t} =& \Delta\mathrm{E}_t\left[\sum_{j=1}^{\infty}\rho^j r_{t+j}\right] \\ =& \Delta\mathrm{E}_t\left[\sum_{j=1}^{\infty}\rho^j \mathbf{e1}^\top\mathbf{\Gamma}^j\mathbf{u}_t\right] \\ =& \mathbf{e1}^\top(\mathbf{I} - \rho\mathbf{\Gamma})^{-1}\rho\mathbf{\Gamma}\mathbf{u}_t \\ =& \boldsymbol{\lambda}^\top\mathbf{u}_t, \end{aligned}$$

and

$$
\begin{aligned}
N_{cf,t} =& r_t - \mathrm{E}_{t-1}\left[r_t\right] + N_{r,t} \\
=& \mathbf{e1}^\top \mathbf{u}_t + \boldsymbol{\lambda}^\top \mathbf{u}_t \\
=& (\mathbf{e1} + \boldsymbol{\lambda})^\top \mathbf{u}_t .
\end{aligned}
$$

The vector $\mathbf{e1}$ contains 1 only for the first element and 0 for the others, i.e. $\mathbf{e1} = \begin{pmatrix} 1 & \mathbf{0}^\top \end{pmatrix}^\top$. By defining $\boldsymbol{\Sigma}$ as the variance of $\mathbf{u}_t$, i.e. $\boldsymbol{\Sigma} = \mathrm{E}\left[\mathbf{u}_t \mathbf{u}_t^\top\right]$, one can rewrite the variance decomposition above.

$$
\begin{aligned}
\mathrm{Var}\left[N_{r,t}\right] =& \boldsymbol{\lambda}^\top \boldsymbol{\Gamma} \boldsymbol{\lambda} \\
\mathrm{Var}\left[N_{cf,t}\right] =& (\mathbf{e1} + \boldsymbol{\lambda})^\top \boldsymbol{\Sigma} (\mathbf{e1} + \boldsymbol{\lambda}) \\
\mathrm{Cov}\left[N_{r,t}, N_{cf,t}\right] =& \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} (\mathbf{e1} + \boldsymbol{\lambda}).
\end{aligned}
$$

The coefficient matrix $\boldsymbol{\Gamma}$ and the variance matrix $\boldsymbol{\Sigma}$ are estimated using panel data. In order to consider the time effect in the state vector $\mathbf{z}$, I demean the observations cross-section by cross-section and estimate both $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ using WLS with the weight $1/N_t$ following Fama and MacBeth (1973). N_t stands for the number of firms at time t. In addition, I adopt a time-clustered standard error for both $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ because it is robust to the time effect (Petersen, 2009). Following Vuolteenaho (2002) and Callen and Segal (2004), I employ 0.967 ($0.967^{1/4}$) as the annual (quarterly) discount factor ρ.

2.2.3 EGARCH

Ang et al. (2006) exploit daily data to compute idiosyncratic volatilities. Since the news decomposition proposed by Vuolteenaho (2002) requires accounting data, neither daily nor monthly news data are available. In addition, the aforementioned variance decomposition is static so cannot alter traditional idiosyncratic volatilities. Instead, I apply an EGARCH model to estimate news volatilities since many papers studying idiosyncratic volatilities adopt this model (Fu, 2009; Fink et al., 2012). These papers use monthly data to calculate EGARCH idiosyncratic volatilities.

$$\begin{aligned} N_{it} - r_{ft} =& \alpha_i + \beta_i \left(r_{mt} - r_{ft} \right) + s_i SMB_t + h_i HML_t + \varepsilon_{it} \\ \sigma_{it}^2 =& \exp\left(a_i + \sum_{l=1}^{p} b_{il} \ln \sigma_{it-l}^2 + \sum_{k=1}^{q} c_{ik} \left[\theta_i \left(\frac{\varepsilon_{it-k}}{\sigma_{it-k}} \right) + \gamma_i \left(\left| \frac{\varepsilon_{it-k}}{\sigma_{it-k}} \right| - \sqrt{\frac{2}{\pi}} \right) \right] \right). \end{aligned}$$

Since volatilities are correlated serially, an EGARCH model better reflects the time-varying property. In order to avoid the look-ahead bias mentioned by Guo et al. (2014), I compute out-of-sample EGARCH idiosyncratic volatilities recursively. Following above researches, I combine an EGARCH model together with Fama–French model. However, unlike these researches, I only consider the case $p = q = 1$, i.e. EGARCH(1,1), because quarterly data provide less available observations than monthly data. Both maximum likelihood and Normal distribution are employed to estimate this model. The log likelihood function is maximized using TR (trust region) method in SAS. I set 32,767 ($2^{15} - 1$) as the maximum number of iterations.

2.3 Data Description

2.3.1 Raw data

I obtain all CRSP and Compustat data from WRDS. Firstly, I employ CRSP daily stock file to calculate monthly (quarterly) idiosyncratic volatilities and CRSP monthly stock file to construct monthly (quarterly) quintile portfolios sorted on past idiosyncratic volatilities. CRSP daily stock file is from December 31, 1925 to December 31, 2015 and CRSP monthly stock file is from December 1925 to December 2015, respectively. Secondly, I use both Compustat fundamentals annual and Compustat fundamentals quarterly to attain the relevant accounting information such as book-to-market ratio and return on equity. Compustat fundamentals annual is from January 1950 to November 2016 and Compustat fundamentals quarterly is from January 1961 to November 2016, respectively. Thirdly, I merge CRSP and Compustat data by using the linking table of CRSP/Compustat merged. Fourthly, I exploit daily and monthly Fama–French factors to apply CAPM and Fama–French model, which enable to (i) compute an idiosyncratic volatility and (ii) measure the excess performance of a portfolio after considering risk factors.

2.3.2 Volatility and news

Firstly, I estimate both monthly and quarterly idiosyncratic volatilities with daily data. In depth, I regress firm-level excess returns $(r_i - r_f)$ on the market factor $(r_m - r_f)$ and Fama–French factors (SMB, HML) recursively and compute the sample standard deviation of residuals $\left(\sqrt{\widehat{\mathrm{Var}}\left[\varepsilon_i\right]}\right)$. I exclude monthly volatilities estimated with less than 11 observations and quarterly volatilities estimated with less than 31 observations. CRSP stocks are sorted on one-month-lagged or one-quarter-lagged volatilities, but excluded if the volatilities are unavailable. Value-weighted quintile portfolios are constructed and rebalanced each month or each quarter.

Secondly, I estimate both annual and quarterly VAR models with annual and quarterly data, respectively. For annual data, I only include the observations at time t with (i) a book equity available at $t-1$, $t-2$, $t-3$, (ii) a net income available at $t-1$, $t-2$, (iii) a long-term debt available at $t-1$, $t-2$, (iv) a December fiscal-year end month, (v) a market equity more than \$10M and (vi) a log book-to-market ratio bigger than 1/100 but smaller than 100. In order to compare the results conveniently, I follow these requirements imposed by Vuolteenaho (2002) in the annual case. In contrast, I impose only two restrictions in the quarterly case, i.e. an observation must have (i) a log excess return (r), a log book-to-market ratio (θ) and a log excess return on equity (e) available at $t-1$ and (ii) a December fiscal-year end month. Since idiosyncratic volatilities only require CRSP data, their availabilities are more sufficient than the availabilities of news components that require both CRSP and Compustat data. By relaxing the requirements instead, more returns are able to be decomposed into news components.

Thirdly, I estimate the idiosyncratic volatilities of both discount rate news and cash flow news with the data above using an EGARCH model. In detail, I obtain out-of-sample EGARCH volatilities firm by firm with all historical data available at that time, but only include the volatilities computed with more than or equal to 12 quarterly observations (3 years). Since the estimation involves numerical procedures, one cannot be fully apart from the threat of outliers. Following Fu (2009), I winsorize the smallest and biggest 2.5% of news volatilities quarter by quarter. Panel A and Panel B of Figure 2.1 display the distributions of discount rate and cash flow news idiosyncratic volatilities, respectively.

2.4 Main Result

2.4.1 Idiosyncratic volatility

Table 2.1 shows the month by month (1/0/1) performance of quintile portfolios sorted on lagged idiosyncratic volatilities. The first column is the portfolio with lowest volatility and vice versa. In addition, the sixth column is the zero cost portfolio formed by selling the most volatile and buying the least volatile. The first three rows contain sample means, corresponding t-statistics and standard deviations of quintile portfolios, respectively. While the average return of the least volatile quintile is positive (0.67%/month) and significant (t-statistic=4.71), that of the most volatile quintile is marginal (0.15%/month) and insignificant (t-statistic=0.53). The average return of the arbitrage portfolio is positive (0.53%/month) and significant (t-statistic=2.65). Though this return is smaller than what is reported by Ang et al. (2006) (1.06%/month), its significance is close enough (Newey–West t-statistic=3.10). Though unreported, the result from matching subsample is consistent (0.94%/month, Newey–West t-statistic=2.89).

The second four rows and last eight rows include the results from CAPM and Fama–

French model, respectively. While the CAPM α of quintile 1 is positive (0.13%/month) and significant (t-statistic=3.97), that of quintile 5 is negative (−0.74%/month) and significant (t-statistic=−4.69). The CAPM α of the 1−5 portfolio is positive (0.87%/month) and significant (t-statistic=4.84). Since the CAPM β of this portfolio is negative (−0.53), the abnormal performance cannot be justified by the market risk. Furthermore, this pattern is obvious with Fama–French model as well. The Fama–French model α of the arbitrage portfolio is positive (0.97%/month) and significant (t-statistic=6.90). Risk loadings are also negative (β=−0.31, s=−1.19) or insignificant ($t(h)$-statistic=0.59) so cannot justify the abnormal return. In a nutshell, this confirms the consistency of the results and implies the existence of the idiosyncratic volatility anomaly.

Table 2.2 shows the performance of quintile portfolios quarter by quarter (3/0/3) instead. The format of this table is identical to that of Table 2.1. Unlike the case of Table 2.1, the average return of the 1−5 portfolio is negative (−0.81%/quarter). However, this average return is insignificant (t-statistic=−0.78). In contrast, the results from both models indicate that the abnormal return is positive (α_{CAPM}=1.27%/quarter, α_{FF}=1.17%/quarter) and more significant ($t\left(\alpha_{\text{CAPM}}\right)$-statistic=1.45, $t\left(\alpha_{\text{FF}}\right)$-statistic=1.61) after controlling other factors. Though unreported, this abnormal performance is even more significant with the subsample after 1963 (Ang et al., 2006). With the average return 1.90 percent, both CAPM α and FF α are positive (α_{CAPM}=4.19%/quarter, α_{FF}=4.09%/quarter) and significant ($t\left(\alpha_{\text{CAPM}}\right)$-statistic=3.51, $t\left(\alpha_{\text{FF}}\right)$-statistic=4.01). The signs of risk loadings are consistent with those in Table 2.1 and the magnitudes are bigger than them. In particular, the CAPM β of the quarterly 1−5 portfolio is −0.99 (−1.44 with the post-1963 subsample), which is about twice bigger than that of the monthly counterpart (−0.54). Likewise, the coefficients of Fama–French model from quarterly data (β=−0.54, s=−1.74, h=0.23) are bigger than those from monthly data (β=−0.31, s=−1.19, h=0.02). In short, this implies that the idiosyncratic volatility anomaly is consistent in quarterly data as well.

2.4.2 News decomposition

Table 2.3 displays the descriptive statistics computed from annual data. In order to compare the results conveniently, I record both all sample statistics and subsample counterparts. In detail, all sample is from 1954 to 2015 with 58,554 firm-years and subsample is from 1954 to 1996 with 33,302 firm-years. As aforementioned, observations are demeaned cross-section by cross-section to address time fixed effects. Panel A and Panel B exhibit the descriptive statistics obtained from all sample raw data and all sample demeaned data, respectively. Panel C and Panel D present the descriptive statistics calculated from subsample raw data and subsample demeaned data, respectively. All statistics are calculated from pooled data. For three variables r (log excess return), θ (log book-to-market ratio) and e (log excess return on equity), I estimate sample mean, standard deviation, maximum, minimum and three quartiles. While the demeaning reduces the variations in r and θ significantly (0.48 versus 0.44, 0.94 versus 0.91), the reduced portion of the variation in e is marginal (0.41 versus 0.41).

Table 2.4 contains the estimates of annual VAR models. In depth, I analyze both all sample (1954–2015) and subsample (1954–1996). Panel A and Panel B report the results obtained from all sample. Panel C and Panel D present the results attained from subsample. The first 3-by-3 square of Panel A displays the VAR(1) coefficient matrix. Firstly, three coefficients in the first row suggest the positive and significant relation between log return (r_{it}) and three state variables lagged one year (r_{it-1}, θ_{it-1}, e_{it-1}). (1,1), (1,2), and (1,3) coefficients are respectively 0.0543, 0.0519, and 0.0660 and all significant. These coefficients are consistent with a momentum effect (Carhart, 1997), a high book-to-market effect (Fama and French, 1992), and a high profitability effect (Fama and French, 2016), respectively. Secondly, the (2,2) and (3,3) coefficients imply the positive autocorrelation of log book-to-market ratio and log return on equity. The second 3-by-3 square of Panel A demonstrates the variance matrix. Log return shocks are negatively correlated to log book-to-market ratio shocks, but positively correlated to log return on equity. Panel B shows the result of static news variance decomposition. The ratio of discount rate news variance to cash flow news variance is about 11.35% (0.0157/0.1383). This result indicates that firm-level returns are mainly driven not by discount rate news but by cash flow news. The format of Panel C and Panel D is identical to that of Panel A and Panel B. By and large, the sign and significance of the estimates are comparable. Only the (1,3) and (3,2) coefficients have different results (0.0660 versus −0.0104, 0.0133 versus −0.0049). The ratio between discount rate news variance and cash flow news variance is about 9.91% (0.0077/0.0777). In summary, these findings are consistent with those of Vuolteenaho (2002) and Callen and Segal (2004).

Table 2.5 exhibits the descriptive statistics computed from quarterly data instead. The format of this table is identical to that of Table 2.3. The sample is from March 1972 to December 2015 (176 quarters) with 235,704 firm-quarters. To address time fixed effects, observations are demeaned cross-section by cross-section. Panel A displays the descriptive statistics calculated from raw data and Panel B demonstrates the descriptive statistics obtained from demeaned data, respectively. All statistics are calculated from pooled data. I estimate sample mean, standard deviation, maximum, minimum and three quartiles for three variables r, θ and e. Similar to the case of Table 2.3, the demeaning reduces the variations in r and θ (0.29 versus 0.26, 0.98 versus 0.93), but does not reduce the variation in e (0.18 versus 0.18).

Table 2.6 contains the estimates of a quarterly VAR model. Panel A exhibits the estimates of both coefficient matrix and variance matrix and Panel B displays the result of static news variance decomposition. In general, the differences are marginal compared to Table 2.4, and (1,1), (1,2), and (1,3) coefficients still exhibit momentum, book-to-market, and profitability effects, respectively. In the quarterly VAR model, the (1,1) and (1,2) coefficients are smaller than the counterparts in the annual VAR model (0.0543 versus 0.0288, 0.0519 versus 0.0084). In contrast, the (1,3) coefficient is bigger than the counterpart in the annual VAR model (0.0660 versus 0.0975). This is natural since this VAR model requires quarterly returns instead of annual returns. By and large, both momentum and book-to-market effects are weaker in the short run and researchers often use a past 11-month return from $t-2$ to $t-12$ months for a momentum effect and a book-to-market ratio from $t-1$ accounting year. On the other hand, the profitability effect is clearer in the short run (Hou et al., 2015). Panel B shows the result of static news variance decomposition. The ratio of discount rate news variance to cash flow news variance is about 7.60% (0.0040/0.0526), which is smaller than the ratio computed using annual data (11.35%) and the ratio from the subsample (9.91%). Again, this result emphasizes the role of cash flow news in determining quarterly returns. Overall, this result is consistent with what Table 2.4 exhibits and justifies the use of quarterly data in decomposing quarterly firm-level returns.

2.4.3 News volatility

Table 2.7 shows the performance of quarterly decile portfolios sorted on EGARCH idiosyncratic volatilities of discount rate news. The format of this table is identical to that of Table 2.2. The average return of the 1—10 portfolio is negative (—0.04%/quarter) but insignificant (t-statistic=—0.06). Likewise, both CAPM and Fama–French alphas of this portfolio are positive (α_{CAPM}=0.58%/quarter, α_{FF}=0.09%/quarter) but insignificant ($t\,(\alpha_{\text{CAPM}})$-statistic=0.88, $t\,(\alpha_{\text{FF}})$-statistic=0.14). Like the case of Table 2.2, both CAPM and Fama–French betas are negative (β_{CAPM}=—0.32, β_{FF}=—0.26) and significant ($t\,(\beta_{\text{CAPM}})$-statistic=—4.20, $t\,(\beta_{\text{FF}})$-statistic=—3.03). Unlike the case of Table 2.2, however, the HML coefficient instead the SMB coefficient is positive (0.36) and significant (t-statistic=3.27). This implies that the stocks in the first portfolio and those in the fifth counterpart are not much different in terms of size. This is not a surprising result as I impose more restrictions here. In contrast, traditional idiosyncratic volatility portfolios impose relatively less restrictions, so smaller firms are included more. In short, this indicates the difference between the portfolios sorted on traditional idiosyncratic volatilities and the portfolios sorted on discount rate news idiosyncratic volatilities.

Table 2.8 shows the performance of quarterly decile portfolios sorted on EGARCH idiosyncratic volatilities of cash flow news. The format of this table is identical to that of Table 2.2. The average return of the zero investment portfolio is positive (0.76%/quarter) but insignificant (t-statistic=1.08). However, both CAPM and Fama–French alphas of this portfolio are positive (α_{CAPM}=1.52%/quarter, α_{FF}=1.21%/quarter) and significant ($t\left(\alpha_{\text{CAPM}}\right)$-statistic=2.24, $t\left(\alpha_{\text{FF}}\right)$-statistic=1.81). Like the case of Table 2.2, both CAPM and Fama–French betas are negative (β_{CAPM}=−0.39, β_{FF}=−0.23) and significant ($t\left(\beta_{\text{CAPM}}\right)$-statistic=−4.95, $t\left(\beta_{\text{FF}}\right)$-statistic=−2.69). The sign and significance of Fama–French coefficients are identical as well. The SMB coefficient is negative (−0.46, t-statistic=−3.18) and the HML coefficient is positive (0.30, t-statistic=2.74). Unlike the case of Table 2.7, the portfolios sorted on cash flow news idiosyncratic volatilities are comparable with those sorted on discount rate news idiosyncratic volatilities. Therefore, one can confirm that the behavior of the arbitrage portfolio using cash flow news idiosyncratic volatilities is similar to that of the arbitrage portfolio using traditional idiosyncratic volatilities, while the behavior of the arbitrage portfolio using discount rate news idiosyncratic volatilities is different from the other two.

This result suggests that not all idiosyncratic volatilities are priced in the cross-section, and that investors price cash flow news idiosyncratic volatilities and discount rate news idiosyncratic volatilities differently due to some reasons such as skewness preference and distress preference. According to the result, investors tend to underprice less volatile stocks but overprice more volatile counterparts in terms of either traditional idiosyncratic volatilities or cash flow news idiosyncratic volatilities. In contrast, such a tendency dissapears when one introduces discount rate news idiosyncratic volatilities instead.

2.5 Conclusion

Throughout this paper, I revisit the idiosyncratic volatility anomaly of Ang et al. (2006) using quarterly data and compare the relative importance of discount rate and cash flow news volatilities in driving the volatility puzzle based on the predictive regression-based approach proposed by Vuolteenaho (2002). There have been many suggestions concerning why investors price the idiosyncratic volatilities in a counterintuitive way in the cross-section, but researchers have seldom paid attention to the respective role of cash flow and discount rate volatilities in explaining the anomaly.

First, I estimate monthly and quarterly idiosyncratic volatilities using daily data and construct respective quintile portfolios sorted on these volatilities. The arbitrage portfolio collects 0.9 percent (1.3 percent) alpha returns per month (quarter) on average after considering the market factor and does 1.0 percent (1.2 percent) alpha returns per month (quarter) on average after considering Fama–French factors, respectively. The result shows that the anomaly is consistent in both monthly and quarterly data. Second, I estimate annual and quarterly VAR models to decompose firm-level stock returns into discount rate news components and cash flow news counterparts. Overall, the estimates from annual and quarterly data are consistent with each other and suggest that cash flow news components rather than discount rate news counterparts play a more critical role in driving returns.

To see if both cash flow and discount rate idiosyncratic volatilities are priced, I estimate these volatilities using an out-of-sample EGARCH model on a quarterly basis and investigate the cross-section of returns following Fu (2009). I create two decile portfolios sorted on discount rate and cash flow news volatilities, correspondingly. The average return of the $1-10$ portfolio from discount rate news volatilities is insignificant, but the corresponding portfolio from cash flow news volatilities acquires about 1.5 percent (1.2 percent) alpha returns per quarter on average after controlling the market factor (Fama–French factors). These findings indicate that cash flow news volatilities rather than discount rate news counterparts are the main driving force of the volatility anomaly. In other words, investors do not equally value cash flow and discount rate idiosyncratic volatilities, but cherry-pick and overvalue the former volatilities, while do not price the latter volatilities in the cross-section.

Skewness preference partly vindicates this tendency. If cash flow shocks are more skewed then discount rate counterparts, then cash flow idiosyncratic volatilities will deliver some information about how much returns are skewed, while discount rate idiosyncratic volatilities will not convey such information. Another candidate explanation regarding the volatility anomaly is distress preference. Because overcoming financial distress is a positive signal for investors, they tend to consider stocks under distress as unscratched lottery tickets and correspondingly overvalue them. Further discussions and investigations will be worthwhile as the volatility anomaly is prevalent apparently, while economic theories supporting the volatility anomaly are relatively weaker yet.

Table 2.1. Monthly Idiosyncratic Volatility-Sorted Portfolio

This table reports the performance of monthly idiosyncratic volatility-sorted quintile portfolios. I construct these quintile portfolios by using lagged monthly idiosyncratic volatilities. The volatilities are estimated from daily returns and Fama–French model, i.e.

$$r_i - r_f = \alpha_i + \beta_i (r_m - r_f) + s_i SMB + h_i HML + \varepsilon_i.$$

For notational convenience, I omit time subscripts. I estimate $\sqrt{\mathrm{Var}[\varepsilon_i]}$ for all firms month by month and construct quintile portfolios recursively (1/0/1). Idiosyncratic volatilities computed with less than 11 daily observations are excluded. The sample is from December 31, 1925 to December 31, 2015. 5 portfolios are from August 1926 to December 2015 (1,073 months). The first row displays quintiles column by column. The last column is the zero investment portfolio (i.e. constructed by buying the first and selling the fifth). The second partition include sample means and standard deviations. The third and fourth partitions contain both CAPM and Fama–French model estimates, respectively. Corresponding t-statistics are reported by using round brackets.

Quintile	1	2	3	4	5	1–5
Mean	0.0067	0.0071	0.0076	0.0052	0.0015	0.0053
	(4.7129)	(3.8419)	(3.4983)	(2.1295)	(0.5280)	(2.6515)
St. dev.	0.0466	0.0606	0.0715	0.0804	0.0900	0.0648
CAPM α	0.0013	0.0000	−0.0006	−0.0036	−0.0074	0.0087
	(3.9674)	(0.0032)	(−1.1564)	(−3.6459)	(−4.6909)	(4.8436)
CAPM β	0.8430	1.1020	1.2830	1.3670	1.3720	−0.5294
	(143.2530)	(159.6910)	(125.0570)	(75.2020)	(47.1820)	(−16.0576)
FF α	0.0013	−0.0004	−0.0013	−0.0043	−0.0084	0.0097
	(4.5894)	(−1.2072)	(−2.7654)	(−5.8466)	(−6.6785)	(6.9000)
FF β	0.8660	1.0810	1.2070	1.2290	1.1730	−0.3074
	(159.4050)	(158.7280)	(131.4350)	(84.1420)	(47.0550)	(−11.0391)
FF s	−0.1687	−0.0092	0.2923	0.7003	1.0214	−1.1902
	(−18.9814)	(−0.8220)	(19.4516)	(29.2910)	(25.0326)	(−26.1115)
FF h	0.0550	0.1466	0.1306	0.0309	0.0308	0.0243
	(6.8963)	(14.6702)	(9.6842)	(1.4401)	(0.8399)	(0.5932)

Table 2.2. Quarterly Idiosyncratic Volatility-Sorted Portfolio

This table reports the performance of quarterly idiosyncratic volatility-sorted quintile portfolios. I construct these quintile portfolios by using lagged quarterly idiosyncratic volatilities. The volatilities are estimated from daily returns and Fama–French model, i.e.

$$r_i - r_f = \alpha_i + \beta_i (r_m - r_f) + s_i SMB + h_i HML + \varepsilon_i.$$

For notational convenience, I omit time subscripts. I estimate $\sqrt{\mathrm{Var}[\varepsilon_i]}$ for all firms quarter by quarter and construct quintile portfolios recursively (3/0/3). Idiosyncratic volatilities computed with less than 31 daily observations are excluded. The sample is from December 31, 1925 to December 31, 2015. 5 portfolios are from September 1926 to December 2015 (358 quarters). The first row displays quintiles column by column. The last column is the zero investment portfolio (i.e. constructed by buying the first and selling the fifth). The second partition include sample means and standard deviations. The third and fourth partitions contain both CAPM and Fama–French model estimates, respectively. Corresponding t-statistics are reported by using round brackets.

Quintile	1	2	3	4	5	1–5
Mean	0.0204	0.0227	0.0273	0.0248	0.0285	−0.0081
	(4.1454)	(3.0555)	(2.9126)	(2.3107)	(2.1512)	(−0.7767)
St. dev.	0.0933	0.1406	0.1773	0.2029	0.2507	0.1965
CAPM α	0.0034	−0.0030	−0.0047	−0.0105	−0.0093	0.0127
	(3.2583)	(−1.9531)	(−1.8782)	(−2.6013)	(−1.1681)	(1.4525)
CAPM β	0.8170	1.2314	1.5314	1.6875	1.8102	−0.9933
	(89.8724)	(90.0240)	(69.3692)	(47.5046)	(25.7498)	(−12.9130)
FF α	0.0029	−0.0049	−0.0062	−0.0110	−0.0087	0.0117
	(3.1553)	(−3.8043)	(−3.0571)	(−3.5278)	(−1.3149)	(1.6066)
FF β	0.8478	1.1589	1.3680	1.4232	1.3849	−0.5372
	(86.6026)	(86.1318)	(63.7162)	(43.4302)	(19.8181)	(−7.0312)
FF s	−0.1495	0.0970	0.4530	0.9106	1.5917	−1.7412
	(−8.6259)	(4.0700)	(11.9160)	(15.6914)	(12.8622)	(−12.8691)
FF h	0.0617	0.2203	0.1609	0.0053	−0.1721	0.2337
	(4.6777)	(12.1592)	(5.5649)	(0.1199)	(−1.8281)	(2.2713)

Table 2.3. Annual Sample Descriptive Statistic

This table displays the descriptive statistics obtained from annual data. I report both all sample and subsample statistics for convenient comparison. All sample is from 1954 to 2015 (58,554 firm-years) and subsample is from 1954 to 1996 (33,302 firm-years). The variables are demeaned each year to address time fixed effects, i.e.

$$X_{it}^{\text{demeaned}} = X_{it}^{\text{raw}} - \sum_{j=1}^{N_t} X_{jt}/N_t.$$

The subscripts i and t stand for firm and year, respectively. The variables r, θ and e are log excess return, log book-to-market ratio and log excess return on equity, respectively. Panel A–D contain the descriptive statistics of raw data from all sample, those of demeaned data from all sample, those of raw data from subsample and those of demeaned data from subsample, respectively. To be included in the data, an observation must have (i) a book equity available at $t-1$, $t-2$, $t-3$, (ii) a net income available at $t-1$, $t-2$, (iii) a long-term debt available at $t-1$, $t-2$, (iv) a December fiscal-year end month, (v) a market equity more than \$10M and (vi) a log book-to-market ratio bigger than 1/100 but smaller than 100.

Variable	Mean	St. dev.	Minimum	1Q	Median	3Q	Maximum
			Panel A. All sample raw data				
r_{it}^{raw}	−0.0076	0.4831	−3.9068	−0.2297	0.0284	0.2568	3.3388
θ_{it}^{raw}	−0.3640	0.9402	−4.6024	−0.8689	−0.3232	0.1679	4.5943
e_{it}^{raw}	−0.0490	0.4132	−2.3026	−0.0426	0.0328	0.0907	4.3625
			Panel B. All sample demeaned data				
r_{it}^{demeaned}	0	0.4435	−3.8314	−0.2047	0.0181	0.2362	3.1295
$\theta_{it}^{\text{demeaned}}$	0	0.9061	−4.5725	−0.4605	0.0457	0.4633	5.1649
e_{it}^{demeaned}	0	0.4078	−2.3617	−0.0260	0.0582	0.1549	4.3760
			Panel C. Subsample raw data				
r_{it}^{raw}	0.0349	0.4071	−3.2730	−0.1755	0.0409	0.2524	3.3388
θ_{it}^{raw}	−0.2778	0.8149	−4.5156	−0.7087	−0.1926	0.2383	4.5803
e_{it}^{raw}	−0.0032	0.2940	−2.3026	−0.0180	0.0364	0.0843	4.3625
			Panel D. Subsample demeaned data				
r_{it}^{demeaned}	0	0.3646	−3.1734	−0.1904	0.0024	0.1926	3.1295
$\theta_{it}^{\text{demeaned}}$	0	0.7684	−4.1723	−0.3815	0.0804	0.4546	5.0759
e_{it}^{demeaned}	0	0.2907	−2.3617	−0.0292	0.0266	0.0923	4.3760

Table 2.4. Annual VAR Model Estimate

Panel A and Panel C of this table report the estimates of annual firm-level VAR models, i.e.

$$\mathbf{z}_{it} = \boldsymbol{\Gamma}\mathbf{z}_{it-1} + \mathbf{u}_{it}$$
$$\boldsymbol{\Sigma} = \mathrm{E}\left[\mathbf{u}_{it}\mathbf{u}_{it}^{\top}\right].$$

Subscripts i and t stand for firm and year, respectively. Above $\mathbf{z}_{it}$ is the vector of three state variables r_{it}, θ_{it} and e_{it}, which are log excess return, log book-to-market ratio and log excess return on equity. To address time fixed effects, all state variables are demeaned year by year. I report both all sample and subsample estimates in Panel A and Panel C. All sample is from 1954 to 2015 (58,554 firm-years) and subsample is from 1954 to 1996 (33,302 firm-years). I estimate both $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ by using WLS with the weight $1/N_t$. The first and second 3-by-3 squares include the estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, respectively. Corresponding t-statistics are computed with time-clustered standard errors (Petersen, 2009) and reported by using round brackets.

Panel A. All sample VAR model

	$\boldsymbol{\Gamma}$				$\boldsymbol{\Sigma}$		
	r_{it-1}	θ_{it-1}	e_{it-1}		r_{it}	θ_{it}	e_{it}
r_{it}	0.0543	0.0519	0.0660	r_{it}	0.1516	−0.1242	0.0225
	(3.1081)	(6.2696)	(2.9213)		(11.5505)	(−11.5470)	(8.2013)
θ_{it}	0.1412	0.8278	0.0633	θ_{it}	−0.1242	0.2043	0.0147
	(7.3666)	(59.7793)	(2.4752)		(−11.5470)	(11.2931)	(4.9005)
e_{it}	0.1178	0.0133	0.4993	e_{it}	0.0225	0.0147	0.0775
	(7.1741)	(2.3366)	(19.5019)		(8.2013)	(4.9005)	(8.4908)

Panel B. All sample variance decomposition

$\mathrm{Var}[N_r]$	$\mathrm{Var}[N_{cf}]$	$-2 \times \mathrm{Cov}[N_r, N_{cf}]$	$\mathrm{Corr}[N_r, N_{cf}]$
0.0157	0.1383	−0.0024	0.0259
[0.1035]	[0.9124]	[−0.0159]	

Panel C. Subsample VAR model

	$\boldsymbol{\Gamma}$				$\boldsymbol{\Sigma}$		
	r_{it-1}	θ_{it-1}	e_{it-1}		r_{it}	θ_{it}	e_{it}
r_{it}	0.0434	0.0567	−0.0104	r_{it}	0.1007	−0.0844	0.0121
	(2.5414)	(5.4237)	(−0.4520)		(10.6608)	(−10.3151)	(6.9688)
θ_{it}	0.1147	0.8150	0.0921	θ_{it}	−0.0844	0.1373	0.0073
	(5.1928)	(35.4021)	(2.6917)		(−10.3151)	(8.6025)	(2.6708)
e_{it}	0.0868	−0.0049	0.3846	e_{it}	0.0121	0.0073	0.0388
	(4.1457)	(−0.6027)	(7.5094)		(6.9688)	(2.6708)	(5.8756)

Panel D. Subsample variance decomposition

$\mathrm{Var}[N_r]$	$\mathrm{Var}[N_{cf}]$	$-2 \times \mathrm{Cov}[N_r, N_{cf}]$	$\mathrm{Corr}[N_r, N_{cf}]$
0.0077	0.0777	0.0153	−0.3117
[0.0769]	[0.7713]	[0.1518]	

Table 2.5. Quarterly Sample Descriptive Statistic

This table displays the descriptive statistics obtained from quarterly data. The sample is from March 1972 to December 2015 (176 quarters, 235,704 firm-quarters). The variables are demeaned each quarter to address time fixed effects, i.e.

$$X_{it}^{\text{demeaned}} = X_{it}^{\text{raw}} - \sum_{j=1}^{N_t} X_{jt}/N_t.$$

The subscripts i and t stand for firm and quarter, respectively. The variables r, θ and e are log excess return, log book-to-market ratio and log excess return on equity, respectively. Panel A, B contain the descriptive statistics of raw data and those of demeaned data, respectively. In order for an observation to be included in the data, here I impose two requirements, i.e. an observation must have (i) r, θ and e available at $t-1$ and (ii) a December fiscal-year end month.

Variable	Mean	St. dev.	Minimum	1Q	Median	3Q	Maximum
			Panel A. Raw data				
r_{it}^{raw}	−0.0196	0.2938	−4.6771	−0.1353	0.0021	0.1242	2.5846
θ_{it}^{raw}	−0.3422	0.9777	−8.4217	−0.8697	−0.3024	0.2052	9.8809
e_{it}^{raw}	−0.0228	0.1842	−2.3026	−0.0146	0.0075	0.0233	4.9245
			Panel B. Demeaned data				
r_{it}^{demeaned}	0	0.2649	−4.3816	−0.1113	0.0082	0.1280	2.5366
$\theta_{it}^{\text{demeaned}}$	0	0.9259	−8.3662	−0.4717	0.0431	0.4796	10.5509
e_{it}^{demeaned}	0	0.1826	−2.3052	−0.0046	0.0223	0.0511	4.9665

Table 2.6. Quarterly VAR Model Estimate

Panel A of this table reports the estimates of a quarterly firm-level VAR model, i.e.

$$\mathbf{z}_{it} = \mathbf{\Gamma}\mathbf{z}_{it-1} + \mathbf{u}_{it}.$$

Subscripts i and t stand for firm and quarter, respectively. Above $\mathbf{z}_{it}$ is the vector of three state variables r_{it}, θ_{it} and e_{it}, which are log excess return, log book-to-market ratio and log excess return on equity, respectively, i.e. $\mathbf{z}_{it} = \begin{pmatrix} r_{it} & \theta_{it} & e_{it} \end{pmatrix}^\top$. To address time fixed effects, all state variables are demeaned quarter by quarter. The sample is from March 1972 to December 2015 (176 quarters, 235,704 firm-quarters). I estimate both $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ by using WLS with the weight $1/N_t$. The first 3-by-3 square includes the estimate of $\mathbf{\Gamma}$. The second 3-by-3 square contains the estimate of $\mathbf{\Sigma}$. Corresponding t-statistics are computed with time-clustered standard errors (Petersen, 2009) and reported by using round brackets. Panel B of this table states the variance decomposition, i.e.

$$\begin{aligned}
\mathrm{Var}\left[N_r\right] =& \boldsymbol{\lambda}^\top \mathbf{\Sigma} \boldsymbol{\lambda} \\
\mathrm{Var}\left[N_{cf}\right] =& (\mathbf{e1} + \boldsymbol{\lambda})^\top \mathbf{\Sigma} (\mathbf{e1} + \boldsymbol{\lambda}) \\
\mathrm{Cov}\left[N_r, N_{cf}\right] =& \boldsymbol{\lambda}^\top \mathbf{\Sigma} (\mathbf{e1} + \boldsymbol{\lambda}).
\end{aligned}$$

Note that $\mathbf{e1} = \begin{pmatrix} 1 & \mathbf{0}^\top \end{pmatrix}^\top$ and $\boldsymbol{\lambda} = \rho \mathbf{\Gamma}^\top \left((\mathbf{I} - \rho \mathbf{\Gamma})^{-1} \right)^\top \mathbf{e1}$. For notational convenience, I omit time subscripts. N_r and N_{cf} stand for discount rate news and cash flow news, respectively. The first row contains the estimates of $\mathrm{Var}\left[N_r\right]$, $\mathrm{Var}\left[N_{cf}\right]$, $-2 \times \mathrm{Cov}\left[N_r, N_{cf}\right]$ and $\mathrm{Corr}\left[N_r, N_{cf}\right]$. The ratio of each component to the total variance is reported in the second row with a square bracket.

Panel A. VAR model

	$\mathbf{\Gamma}$				$\mathbf{\Sigma}$		
	r_{it-1}	θ_{it-1}	e_{it-1}		r_{it}	θ_{it}	e_{it}
r_{it}	0.0288	0.0084	0.0975	r_{it}	0.0605	-0.0557	0.0046
	(1.8182)	(2.8660)	(6.6187)		(17.6341)	(-17.1722)	(9.0500)
θ_{it}	0.0421	0.9525	0.0426	θ_{it}	-0.0557	0.1003	0.0089
	(2.4363)	(245.0869)	(1.8873)		(-17.1722)	(19.8868)	(11.0361)
e_{it}	0.0648	0.0057	0.4597	e_{it}	0.0046	0.0089	0.0233
	(10.2844)	(3.4727)	(21.6617)		(9.0500)	(11.0361)	(15.5855)

Panel B. Variance decomposition

$\mathrm{Var}\left[N_r\right]$	$\mathrm{Var}\left[N_{cf}\right]$	$-2 \times \mathrm{Cov}\left[N_r, N_{cf}\right]$	$\mathrm{Corr}\left[N_r, N_{cf}\right]$
0.0040	0.0526	0.0039	-0.1331
[0.0657]	[0.8706]	[0.0637]	

Table 2.7. Quarterly Discount Rate News Volatility-Sorted Portfolio

This table reports the performance of quarterly discount rate news (N_r) volatility-sorted decile portfolios. I construct these decile portfolios by using lagged quarterly discount rate news volatilities. Following Fu (2009), these volatilities are estimated from quarterly discount rate news data, Fama–French model and an EGARCH model, i.e.

$$N_{rit} - r_{ft} = \alpha_i + \beta_i (r_{mt} - r_{ft}) - s_i SMB_t + h_i HML_t + \varepsilon_{it}$$
$$\sigma_{it}^2 = \exp\left(a_i + \sum_{l=1}^{p} b_{il} \ln \sigma_{it-l}^2 + \sum_{k=1}^{q} c_{ik}\left[\theta_i\left(\frac{\varepsilon_{it-k}}{\sigma_{it-k}}\right) + \gamma_i\left(\left|\frac{\varepsilon_{it-k}}{\sigma_{it-k}}\right| - \sqrt{\frac{2}{\pi}}\right)\right]\right).$$

In order to avoid the look-ahead bias mentioned by Guo et al. (2014), I estimate out-of-sample EGARCH idiosyncratic volatilities recursively. Idiosyncratic volatilities computed with less than 12 quarterly observations are excluded. 10 portfolios are from December 1974 to December 2015 (165 quarters). The first row displays deciles column by column. The last column is the zero investment portfolio (i.e. constructed by buying the first and selling the tenth). The second partition include sample means and standard deviations. The third and fourth partitions contain both CAPM and Fama–French model estimates, respectively. Corresponding t-statistics are reported by using round brackets.

Decile	1	2	3	4	5	6	7	8	9	10	1–10
Mean	0.0213	0.0229	0.0220	0.0222	0.0245	0.0166	0.0245	0.0115	0.0233	0.0217	−0.0004
	(2.9876)	(3.3858)	(3.1723)	(3.2408)	(3.3829)	(2.1033)	(3.0763)	(1.2655)	(2.6248)	(2.2628)	(−0.0572)
St. dev.	0.0906	0.0865	0.0886	0.0871	0.0924	0.1016	0.1013	0.1164	0.1133	0.1218	0.0855
CAPM α	0.0040	0.0060	0.0041	0.0052	0.0059	−0.0043	0.0047	−0.0123	0.0007	−0.0018	0.0058
	(0.9742)	(1.5841)	(1.0101)	(1.3921)	(1.5631)	(−0.9454)	(1.0934)	(−2.3993)	(0.1459)	(−0.3424)	(0.8841)
CAPM β	0.9008	0.8687	0.8605	0.8873	0.9533	0.9959	1.0285	1.1491	1.1626	1.2230	−0.3223
	(18.8992)	(19.4372)	(18.4553)	(20.4961)	(21.4052)	(18.9358)	(20.2686)	(19.4492)	(20.9381)	(19.5381)	(−4.1973)
FF α	0.0012	0.0043	0.0016	0.0038	0.0034	−0.0080	0.0031	−0.0135	0.0003	0.0002	0.0009
	(0.2883)	(1.1003)	(0.4000)	(0.9991)	(0.8760)	(−1.7753)	(0.7044)	(−2.6179)	(0.0669)	(0.0432)	(0.1433)
FF β	0.9504	0.9211	0.9235	0.8880	0.9778	1.0347	1.0075	1.0742	1.1125	1.2058	−0.2555
	(18.1093)	(18.5468)	(17.8189)	(18.2290)	(19.8866)	(17.9530)	(17.7547)	(16.2990)	(17.8864)	(17.1152)	(−3.0288)
FF s	−0.0391	−0.1026	−0.1118	0.0711	0.0499	0.0461	0.1690	0.3272	0.2142	−0.0462	0.0071
	(−0.4428)	(−1.2295)	(−1.2903)	(0.8683)	(0.6039)	(0.4854)	(1.7716)	(2.9697)	(2.0496)	(−0.3902)	(0.0504)
FF h	0.2133	0.1510	0.2019	0.0893	0.1873	0.2777	0.0930	0.0327	−0.0057	−0.1458	0.3591
	(3.1255)	(2.3379)	(2.9875)	(1.4090)	(2.9284)	(3.7102)	(1.2605)	(0.3807)	(−0.0700)	(−1.5910)	(3.2740)

Table 2.8. Quarterly Cash Flow News Volatility-Sorted Portfolio

This table reports the performance of quarterly cash flow news (N_{cf}) volatility-sorted decile portfolios. I construct these decile portfolios by using lagged quarterly discount rate news volatilities. Following Fu (2009), these volatilities are estimated from quarterly discount rate news data, Fama–French model and an EGARCH model, i.e.

$$N_{cfit} - r_{ft} = \alpha_i + \beta_i (r_{mt} - r_{ft}) - s_i SMB_t + h_i HML_t + \varepsilon_{it}$$
$$\sigma_{it}^2 = \exp\left(a_i + \sum_{l=1}^{p} b_{il} \ln \sigma_{it-l}^2 + \sum_{k=1}^{q} c_{ik} \left[\theta_i \left(\frac{\varepsilon_{it-k}}{\sigma_{it-k}}\right) + \gamma_i \left(\left|\frac{\varepsilon_{it-k}}{\sigma_{it-k}}\right| - \sqrt{\frac{2}{\pi}}\right)\right]\right).$$

In order to avoid the look-ahead bias mentioned by Guo et al. (2014), I estimate out-of-sample EGARCH idiosyncratic volatilities recursively. Idiosyncratic volatilities computed with less than 12 quarterly observations are excluded. 10 portfolios are from December 1974 to December 2015 (165 quarters). The first row displays deciles column by column. The last column is the zero investment portfolio (i.e. constructed by buying the first and selling the tenth). The second partition include sample means and standard deviations. The third and fourth partitions contain both CAPM and Fama–French model estimates, respectively. Corresponding t-statistics are reported by using round brackets.

Decile	1	2	3	4	5	6	7	8	9	10	1–10
Mean	0.0215	0.0245	0.0186	0.0210	0.0235	0.0210	0.0249	0.0181	0.0299	0.0139	0.0076
	(2.8606)	(3.4521)	(2.6668)	(3.1173)	(3.1258)	(2.7191)	(3.3580)	(1.9856)	(2.8415)	(1.3073)	(1.0807)
St. dev.	0.0956	0.0905	0.0891	0.0856	0.0960	0.0990	0.0942	0.1169	0.1342	0.1349	0.0899
CAPM α	0.0032	0.0077	0.0001	0.0034	0.0040	−0.0001	0.0056	−0.0053	0.0078	−0.0119	0.0152
	(0.7520)	(1.7340)	(0.0308)	(1.0936)	(1.0517)	(−0.0209)	(1.6110)	(−0.9676)	(1.0134)	(−1.9682)	(2.2378)
CAPM β	0.9507	0.8646	0.8884	0.9143	1.0021	0.9997	1.0041	1.1274	1.1357	1.3427	−0.3920
	(18.9083)	(16.7097)	(20.1506)	(25.0356)	(22.4598)	(20.9885)	(24.8295)	(18.0526)	(12.6414)	(18.9538)	(−4.9501)
FF α	0.0015	0.0038	−0.0037	0.0008	0.0018	−0.0020	0.0060	−0.0045	0.0083	−0.0106	0.0121
	(0.3421)	(0.8705)	(−1.0210)	(0.2417)	(0.4554)	(−0.4829)	(1.6988)	(−0.8106)	(1.0591)	(−1.7935)	(1.8135)
FF β	0.9554	0.9479	0.9765	0.9352	1.0035	1.0116	0.9753	1.0548	1.0630	1.1848	−0.2293
	(16.9066)	(17.1053)	(20.9427)	(23.5492)	(20.2934)	(18.8455)	(21.3704)	(14.9396)	(10.5075)	(15.6762)	(−2.6870)
FF s	0.0753	−0.1042	−0.1376	0.0646	0.1207	0.0563	0.0853	0.2265	0.2504	0.5312	−0.4559
	(0.7927)	(−1.1197)	(−1.7650)	(0.9676)	(1.4524)	(0.6365)	(1.1123)	(1.9188)	(1.4730)	(4.1819)	(−3.1784)
FF h	0.1156	0.3100	0.3082	0.1871	0.1507	0.1419	−0.0487	−0.0955	−0.0805	−0.1888	0.3044
	(1.5727)	(4.3018)	(5.0707)	(3.6217)	(2.3437)	(2.0360)	(−0.8201)	(−1.0375)	(−0.6122)	(−1.9207)	(2.7423)

Figure 2.1. Distributions of Discount Rate and Cash Flow News Volatilities

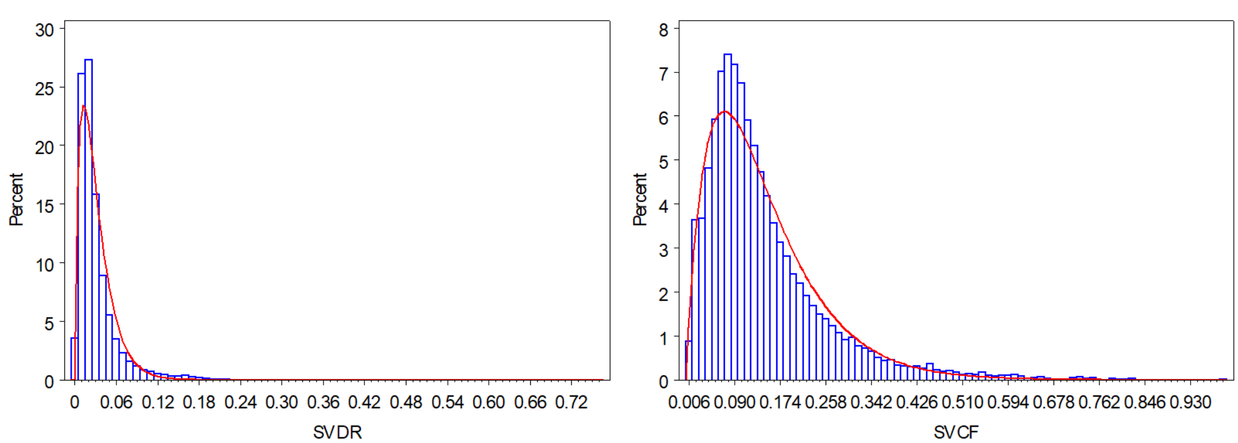
This figure displays the distribution of discount rate news volatilities ($\sigma_{it}[N_{r,it}]$) and that of cash flow news volatilities ($\sigma_{it}[N_{cf,it}]$). I first estimate the news data using a panel VAR model with accounting variables following Vuolteenaho (2002) and second estimate the volatilities using an EGARCH model with Fama–French factors following Fu (2009). For each distribution, I estimate a shape parameter k and a scale parameter θ of a gamma distribution using the following probability density function.

$$f_X(x) = \frac{1}{\Gamma(x)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right) 1_{\mathbb{R}_+}(x).$$

The mean and variance of the distribution are defined as $k\theta$ and $k\theta^2$, respectively. Idiosyncratic volatilities computed with less than 12 quarterly observations are excluded. The smallest and biggest 2.5% of respective news volatilities are winsorized quarter by quarter to remedy potential measurement errors from intense numerical processes. The sample is from December 1974 to December 2015 (165 quarters, 152,099 firm-quarters).

Panel A. Discount rate news volatilities
($\hat{k} = 1.7096$, $\hat{\theta} = 0.0181$)

Panel B. Cash flow news volatilities
($\hat{k} = 1.9618$, $\hat{\theta} = 0.0735$)



Chapter 3

Multiway Clustered Standard Errors in Finite Samples

3.1 Econometric Model

Throughout this paper, I assume that neither multicollinearity ($\mathrm{Cov}\left[x_k, x_{k'}\right] = 0\ \forall k \neq k'$) nor endogeneity ($\mathrm{E}\left[u|x_k\right] = 0\ \forall k$) exists and focus only on the issues related to clustered standard errors in panel data. Suppose that there exists a panel regression below with N firms, T times, and K regressors.

$$y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it} \tag{3.1}$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i \tag{3.2}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{3.3}$$

where i and t are firm and time, respectively. (3.1)–(3.3) express the same regression and are interchangeable. (3.1) denotes the model at the observation level. (3.2) aggregates the observations firm by firm and (3.3) further aggregates them. In many cases, clustering issues are closely related to (3.3). Suppose that both regressors and errors are clustered by firm.

$$\operatorname{Cov}\left[\mathbf{x}_{it}, \mathbf{x}_{i't'}\right] \in \begin{cases} \mathbb{R}^{K\times K} & i = i' \\ \\ \{\mathbf{0}\} & i \neq i' \end{cases} \tag{3.4}$$

$$\operatorname{Cov}\left[u_{it}, u_{i't'}\right] \in \begin{cases} \mathbb{R} & i = i' \\ \\ \{0\} & i \neq i' \end{cases}. \tag{3.5}$$

(3.4) and (3.5) impose no parametric assumption and hence allow heterogeneity across clusters.

What is important in estimating clustered standard errors is the number of nonzero elements in the expected value of the outer product of $\mathbf{u}$. When the errors are clustered at the firm level, the number of the nonzero elements to be considered is equal to $f(N, T) = NT^2$. Hence f increases linearly with N since $\partial f/\partial N = T^2$ and quadratically with T since $\partial^2 f/\partial T^2 = 2N$. In addition, f subject to $g(N, T) = NT$ is minimized at $N^* = g$ and $T^* = 1$. That is, one is able to deal with this clustering by maximizing N and minimizing T, a cross-sectional regression.

Figure 3.1 visualizes four different assumptions regarding regression errors using $N = 2$ and $T = 5$. As shown in the first panel, $\operatorname{E}\left[\mathbf{u}\mathbf{u}^\top\right]$ is diagonal when the observations are clustered by neither firm nor time. When they are clustered at the firm level, however, the expected value of the error outer product is instead block diagonal as shown in the second panel. The expected value, in contrast, is diagonal block by block as shown in the third panel when the data are subject to time effects.

When there exists a clustering by firm, the OLS estimator $\hat{\boldsymbol{\beta}}$ has the limiting distribution below.

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{y} \tag{3.6}$$

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \text{Avar}\left[\hat{\boldsymbol{\beta}}\right]\right) \tag{3.7}$$

$$\text{Avar}\left[\hat{\boldsymbol{\beta}}\right] = \underset{N\to\infty}{\text{plim}}\, N\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{u}\mathbf{u}^\top\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \tag{3.8}$$

$$\begin{aligned} &= \lim_{N\to\infty} N\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left[\mathbf{x}_{it}\mathbf{x}_{it}^\top\right]\right)^{-1} \\ &\quad\times\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left[u_{it}^2\mathbf{x}_{it}\mathbf{x}_{it}^\top\right] + \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'\neq t}^{T}\mathrm{E}\left[u_{it}u_{it'}\mathbf{x}_{it}\mathbf{x}_{it'}^\top\right]\right) \\ &\quad\times\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left[\mathbf{x}_{it}\mathbf{x}_{it}^\top\right]\right)^{-1}. \end{aligned} \tag{3.9}$$

One is able to further simplify (3.9) by assuming higher moments for $\mathbf{x}$ and u.

When the errors are not clustered at the firm level, the second filling of the sandwich $\text{Avar}\left[\hat{\boldsymbol{\beta}}\right]$ above is equal to zero. Here one can consider the OLS variance estimator $s^2(\mathbf{X}^\top\mathbf{X})^{-1}$ when the errors are homoskedastic and the heteroskedasticity-consistent variance estimator otherwise (White, 1980). White estimator is $\sqrt{N}$-consistent but biased when N is finite. There exist several alternatives that adjust the estimator using degrees of freedom (Hinkley, 1977), leverages (Horn et al., 1975) or the jackknife (Efron, 1982) to address this bias issue. MacKinnon and White (1985) compare the alternatives and conclude that the jackknife adjustment better performs than the other two adjustments in estimating the variance.

When the second term is nonzero, however, one must modify the estimator for $\text{Avar}\left[\hat{\boldsymbol{\beta}}\right]$ to reflect the second filling. One can consider the variance estimator below, which is consistent under the clustering (Rogers, 1993; Petersen, 2009). Here I define $\mathbf{1}$ as the matrix whose ijth element is equal to 1 for all is and js and $\mathbf{1}^N$ as the matrix whose ijth element is equal to 1 when both ith and jth observations come from the same firm, that is, the observations are clustered by firm.

$$
\begin{aligned}
\widehat{\mathrm{Var}}_{\mathrm{NCR}}\left[\hat{\boldsymbol{\beta}}\right] &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}^{\top}\right)^{-1} \\
&\quad \times \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{i'=1}^{N}\sum_{t'=1}^{T}\hat{u}_{it}\hat{u}_{i't'}1_{iti't'}^{N}\mathbf{x}_{it}\mathbf{x}_{i't'}^{\top}\right) \\
&\quad \times \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}\mathbf{x}_{it}^{\top}\right)^{-1} && (3.10) \\
&= \left(\sum_{i=1}^{N}\mathbf{X}_i^{\top}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i^{\top}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^{\top}\mathbf{X}_i\right)\left(\sum_{i=1}^{N}\mathbf{X}_i^{\top}\mathbf{X}_i\right)^{-1} && (3.11) \\
&= \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^{\top}\circ\mathbf{1}^{N}\right)\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}, && (3.12)
\end{aligned}
$$

where $\widehat{\mathrm{Var}}_{\mathrm{NCR}}[\hat{\boldsymbol{\beta}}]$ is the clustered variance estimator by firm and $\circ$ is the Hadamard product operator. (3.10)–(3.12) express the same estimator and are interchangeable. (3.10) denotes the estimator at the observation level using a scalar indicator and corresponds to (3.1). (3.12) aggregates this estimator using residual vectors at the firm level and corresponds to (3.2). (3.11) further aggregates this using the indicator matrix $\mathbf{1}^{N}$ and corresponds to (3.3).

Figure 3.2 visualizes four different standard error estimators using the last panel of Figure 3.1. Each sandwich estimator uses $\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}$ and $\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}$ as its bun and the highlighted elements among the residual outer product $\hat{\mathbf{u}}\hat{\mathbf{u}}^{\top}$ as its fillings. While the heteroskedasticity-consistent estimator in the first panel includes the diagonal elements only, the cluster-robust estimator by firm in the second panel includes the block diagonal elements together.

This cluster-robust estimator is $\sqrt{N}$-consistent so one is able to employ asymptotic theory when N is large. When N is small, however, one must change the fillings of the sandwich because $\mathrm{E}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^{\top}\circ\mathbf{1}^{N}\right]$ is the biased estimator of $\mathrm{E}\left[\mathbf{u}\mathbf{u}^{\top}\right]$. (Bell and McCaffrey, 2002) address this bias issue using the bias reduced linearization. Resampling techniques such as the jackknife (Shao and Rao, 1993) or the bootstrap (Cameron et al., 2008) have also been employed.

Suppose, more generally, that both $\mathbf{x}$ and u are doubly clustered by both firm and time.

$$\operatorname{Cov}\left[\mathbf{x}_{it}, \mathbf{x}_{i't'}\right] \in \begin{cases} \mathbb{R}^{K\times K} & i = i' \cup t = t' \\ \{\mathbf{0}\} & i \neq i' \cap t \neq t' \end{cases} \tag{3.13}$$

$$\operatorname{Cov}\left[u_{it}, u_{i't'}\right] \in \begin{cases} \mathbb{R} & i = i' \cup t = t' \\ \{0\} & i \neq i' \cap t \neq t' \end{cases}. \tag{3.14}$$

That is, two different $\mathbf{x}$s (or us) are uncorrelated only when they come from different firms ($i \neq i'$) and ($\cap$) different times ($t \neq t'$). When the data are subject to both firm and time effects, $\operatorname{E}\left[\mathbf{u}\mathbf{u}^\top\right]$ is not only block diagonal but also diagonal block by block as shown in the last panel of Figure 3.1.

When the observations are clustered at both dimensions, the number of nonzero elements to be considered in $\operatorname{E}\left[\mathbf{u}\mathbf{u}^\top\right]$ is equal to $f(N,T) = NT(N+T-1)$. Therefore, f increases quadratically with both N and T since $\partial^2 f/\partial N^2 = 2T$ and $\partial^2 f/\partial T^2 = 2N$. Unlike the singly clustered environment above, f subject to g is minimized at $N^* = T^* = \sqrt{g}$. That is, one is not able to handle this two-way clustering by maximizing N and minimizing T.

Figure 3.3 compares the coverage of two one-way clustered standard error estimators under two-way clustering. When $T \gg N$, a one-way clustered standard error estimator by firm covers most of the residuals so performs well. In contrast, a one-way clustered standard error estimator by time covers the residuals mostly when $N \gg T$. In order for the one-way estimators to be able to alter the the two-way estimators under two-way clustering, the panel regression must have an extreme $N \times T$ combination. That is, the one-way estimator underperforms the two-way estimator significantly when N and T are moderate, so the two-way estimator is necessary as long as there exists two-way clustering.

A lot of finance data are unfortunately closer to an extreme combination—for example, $N \gg T$ in corporate finance and $T \gg N$ in asset pricing—rather than a moderate combination (Skoulakis, 2008). The variance estimator below is still consistent when the observations are clustered by both firm and time (Cameron et al., 2011; Thompson, 2011).

$$
\begin{aligned}
\widehat{\mathrm{Var}}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right] &= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{N\cup T}\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} && (3.15)\\
&= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{N} + \hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{T} - \hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{N\cap T}\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} && (3.16)\\
&= \widehat{\mathrm{Var}}_{\mathrm{NCR}}\left[\hat{\boldsymbol{\beta}}\right] + \widehat{\mathrm{Var}}_{\mathrm{TCR}}\left[\hat{\boldsymbol{\beta}}\right] - \widehat{\mathrm{Var}}_{\mathrm{HC}}\left[\hat{\boldsymbol{\beta}}\right], && (3.17)
\end{aligned}
$$

where $\widehat{\mathrm{Var}}_{\mathrm{TCR}}\left[\hat{\boldsymbol{\beta}}\right]$ is the clustered variance estimator by time and $\widehat{\mathrm{Var}}_{\mathrm{HC}}\left[\hat{\boldsymbol{\beta}}\right]$ is the heteroskedasticity-consistent variance estimator. The first, second, third, and last panels of Figure 3.2 visualizes $\widehat{\mathrm{Var}}_{\mathrm{HC}}\left[\hat{\boldsymbol{\beta}}\right]$, $\widehat{\mathrm{Var}}_{\mathrm{NCR}}\left[\hat{\boldsymbol{\beta}}\right]$, $\widehat{\mathrm{Var}}_{\mathrm{TCR}}\left[\hat{\boldsymbol{\beta}}\right]$, and $\widehat{\mathrm{Var}}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right]$, respectively.

So the two-way clustered variance estimator $\widehat{\mathrm{Var}}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right]$ has the sandwich representation (3.15) and the formula (3.17) is its easy recipe. As shown in the last panel of Figure 3.2, the indicator matrix $\mathbf{1}^{N\cup T} = \mathbf{1}^{N} + \mathbf{1}^{T} - \mathbf{1}^{N\cap T}$ is symmetric but neither diagonal nor block diagonal and hence this two-way clustered variance estimator is always symmetric but not always positive semi-definite unlike its one-way counterpart above. One is able to address this issue using eigendecomposition (Cameron et al., 2011; Politis, 2011).

$$
\begin{aligned}
\widehat{\mathrm{Var}}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right] &= \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\\
&= \mathbf{V}\mathrm{diag}\left(\lambda_1, \cdots, \lambda_K\right)\mathbf{V}^{-1} && (3.18)\\
\widehat{\mathrm{Var}}^{+}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right] &= \mathbf{V}\boldsymbol{\Lambda}^{+}\mathbf{V}^{-1}\\
&= \mathbf{V}\mathrm{diag}\left(\max\left(\lambda_1, 0\right), \cdots, \max\left(\lambda_K, 0\right)\right)\mathbf{V}^{-1}, && (3.19)
\end{aligned}
$$

where $\mathbf{V}$ and $\boldsymbol{\Lambda}$ are the orthonormal matrix of the eigenvectors of $\widehat{\mathrm{Var}}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right]$ and the diagonal matrix of the eigenvalues of $\widehat{\mathrm{Var}}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right]$, respectively. Since $\boldsymbol{\Lambda}^{+}$ is non-negative and diagonal, $\widehat{\mathrm{Var}}^{+}_{\mathrm{NTCR}}\left[\hat{\boldsymbol{\beta}}\right]$ is always positive semi-definite.

Though this two-way clustered variance estimator is consistent, it is biased since $\mathrm{E}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{N\cup T}\right] \neq \mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$. This problem is mainly due to the difference between the regression errors $\mathbf{u}$ and the corresponding residuals $\hat{\mathbf{u}}$. The residuals are neither independent nor identically distributed even with independent and identically distributed errors since $\hat{\mathbf{u}} = \left(\mathbf{I}_{NT} - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\right)\mathbf{u}$ consists of both $\mathbf{X}$ and $\mathbf{u}$. One is able to adjust the clustered estimator using degrees of freedom when the errors are homoskedastic. When the errors are heteroskedastic, however, one has to consider either leverages, the jackknife, or the bootstrap instead to adjust the estimator.

MacKinnon and White (1985) compare the performance of three heteroskedasticity-consistent variance estimators that are adjusted using degrees of freedom (HC1), leverages (HC2), and the jackknife (HC3), respectively. Cameron et al. (2008) likewise compare the performance of several cluster-robust variance estimators including the jackknife estimator (CR3) proposed by Bell and McCaffrey (2002). Following these studies, I hereafter define CR0 as the cluster-robust variance estimator with no adjustment, CR1 as the estimator adjusted using degrees of freedom, CR2 as the estimator adjusted using leverages, and CR3 as the estimator adjusted using the jackknife, respectively.

Using these single clustered variance estimators, one is able to consider either firm effects or time effects but not both (Petersen, 2009; Thompson, 2011). I further attach the prefix N (or the superscript N) if the single clustered variance estimator considers firm effects or the prefix T (or the superscript T) if the estimator considers time effects, respectively. These two estimators are equivalent to each other fundamentally hence only the first one is discussed hereafter. Table 3.1 demonstrates the recipes for heteroskedasticity-consistent variance estimators (HC0–HC3) and cluster-robust ones (CR0–CR3) in detail.

Table 3.1 exhibits the variance estimators together. The first column is the lineup of the estimators—OLS, HC, NCR, and NTCR, respectively. The second column shows the assumptions correspond to the estimators. The third column contains the formulae of the estimators. The fourth column introduces the ways to adjust regression residuals. I employ $\mathbf{\Sigma}^N$ and $\mathbf{\Sigma}^T$ to simplify firm and time effects, respectively, but they can be generalized when there exists heterogeneity among clusters.

Table 3.1 consists of three partitions. The first partition of Table 3.1 displays OLS variance estimator. In detail, OLS estimator assumes that the errors are homoskedastic and uncorrelated with each other. Due to this assumption, OLS estimator does not have a sandwich form anymore. In addition, OLS estimator employs s^2, which is a consistent and unbiased estimator of σ^2.

The second partition displays heteroskedasticity-consistent variance estimators. These estimators still assume that the errors are uncorrelated with each other, but allow the existence of heteroskedasticity. Therefore, the filling inside the sandwich form is not cancelled. HC0 estimator employs the diagonal elements of the outer product $\hat{\mathbf{u}}\hat{\mathbf{u}}^\top$, which is a consistent but biased estimator of $\mathbf{\Sigma}$ (White, 1980). HC1 estimator adjusts this unbiasedness using degrees of freedom. HC2 and HC3 estimators adjust this using leverages. In particular, the way HC3 estimator adjusts the unbiasedness is similar to that of the jackknife estimator (MacKinnon and White, 1985; Davidson and MacKinnon, 2004).

The third partition displays firm clustered variance estimators. These estimators not only allow heteroskedastic errors, but also assume that the errors are clustered due to firm effects. I assume both balanced panel and homogeneous covariance across clusters for simplicity but these assumptions can be easily generalized. Like HC0 estimator, CR0 estimator employs the block diagonal elements of the outer product $\hat{\mathbf{u}}\hat{\mathbf{u}}^\top$ (Rogers, 1993) hence it is consistent but biased (Cameron et al., 2008). CR1, CR2 and CR3 estimators respectively adjust this unbiasedness using degrees of freedom, leverages and the jackknife but they adjust block diagonal elements rather than just diagonal elements.

The last partition displays double clustered variance estimators. These estimators assume that there exist both firm and time effects. Figure 6 of Petersen (2009) visualizes this assumption. Like single CR0 estimator, double CR0 estimator employs (i) all elements in diagonal blocks and (ii) diagonal elements in off-diagonal blocks of the outer product $\hat{\mathbf{u}}\hat{\mathbf{u}}^\top$ (Cameron et al., 2011) hence they are fundamentally equivalent. This sandwich estimator can be rewritten conveniently using two single CR0 estimators and one HC0 estimator (Thompson, 2011; Cameron and Miller, 2015). This recipe is practically useful because one cannot directly adjust $\hat{\mathbf{u}}$ using $\mathbf{I}-\mathbf{P}\circ\mathbf{1}^{N\cup T}$ here since it is not always invertible.

Overall, these estimators rely on the modified residual $\tilde{\mathbf{u}}$ rather than the residual $\hat{\mathbf{u}}$. The following NCR3 estimator, for instance, adjusts the residual using the jackknife approach (Cameron and Miller, 2015).

$$\widehat{\mathrm{Var}}^{N}_{\mathrm{CR3}}\left[\hat{\boldsymbol{\beta}}\right]=\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i^\top\tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i^\top\mathbf{X}_i\right)\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \tag{3.20}$$

$$\tilde{\mathbf{u}}_i=\sqrt{C_N}\left(\mathbf{I}-\mathbf{P}_i\right)^{-1}\hat{\mathbf{u}}_i \tag{3.21}$$

$$C_N=\frac{N}{N-1} \tag{3.22}$$

$$\mathbf{P}_i=\mathbf{X}_i\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}_i^\top. \tag{3.23}$$

Where $\mathbf{P}=\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$ is the projection matrix and $\mathbf{P}_i$ is its ith diagonal entry, respectively. Without such correction, the estimator tends to be biased downward (Davidson and MacKinnon, 2004; Cameron et al., 2008) because the outer product $\hat{\mathbf{u}}\hat{\mathbf{u}}^\top$ is also biased. As aforementioned, one is able to further aggregate the filling of the sandwich estimator (3.20) using Hadamard product as follows.

$$\widehat{\mathrm{Var}}^{N}_{\mathrm{CR3}}\left[\hat{\boldsymbol{\beta}}\right]=\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\tilde{\mathbf{u}}^N\tilde{\mathbf{u}}^{N\top}\circ\mathbf{1}^N\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \tag{3.24}$$

$$\tilde{\mathbf{u}}^N=\sqrt{C_N}\left(\mathbf{I}-\mathbf{P}\circ\mathbf{1}^N\right)^{-1}\hat{\mathbf{u}}. \tag{3.25}$$

And its T-clustered counterpart likewise can be written as follows.

$$\widehat{\mathrm{Var}}^{T}_{\mathrm{CR3}}\left[\hat{\boldsymbol{\beta}}\right]=\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\tilde{\mathbf{u}}^T\tilde{\mathbf{u}}^{T\top}\circ\mathbf{1}^T\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1} \tag{3.26}$$

$$\tilde{\mathbf{u}}^T=\sqrt{C_T}\left(\mathbf{I}-\mathbf{P}\circ\mathbf{1}^T\right)^{-1}\hat{\mathbf{u}}. \tag{3.27}$$

In addition, one can consider a double clustered variance estimator using these two when there exist both firm and time effects. Here one must add these two and subtract their intersection to avoid double counting. The intersection can be written as follows.

$$\widehat{\mathrm{Var}}_{\mathrm{CR3}}^{N\cap T}\left[\hat{\boldsymbol{\beta}}\right] = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\left(\tilde{\mathbf{u}}^{N\cap T}\tilde{\mathbf{u}}^{N\cap T\top}\circ\mathbf{1}^{N\cap T}\right)\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \tag{3.28}$$

$$\tilde{\mathbf{u}}^{N\cap T} = \sqrt{C_{N\cap T}}\left(\mathbf{I}-\mathbf{H}\circ\mathbf{1}^{N\cap T}\right)^{-1}\hat{\mathbf{u}}. \tag{3.29}$$

Because there exist only two clusters N and T here, $\mathbf{1}^{N\cap T}$ and $C_{N\cap T}$ are equal to $\mathbf{I}_{NT}$ and $NT/(NT-1)$, respectively. The following estimator is therefore identical to HC3 estimator. Similar to the estimator (3.17) above, the double clustered variance estimator can be written as follows.

$$\begin{aligned}\widehat{\mathrm{Var}}_{\mathrm{CR3}}^{N\cup T}\left[\hat{\boldsymbol{\beta}}\right] =& \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{N}\left[\hat{\boldsymbol{\beta}}\right] + \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{T}\left[\hat{\boldsymbol{\beta}}\right] - \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{N\cap T}\left[\hat{\boldsymbol{\beta}}\right] &(3.30)\\ =& \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\\ &\times\left(\tilde{\mathbf{u}}^{N}\tilde{\mathbf{u}}^{N\top}\circ\mathbf{1}^{N}+\tilde{\mathbf{u}}^{T}\tilde{\mathbf{u}}^{T\top}\circ\mathbf{1}^{T}-\tilde{\mathbf{u}}^{N\cap T}\tilde{\mathbf{u}}^{N\cap T\top}\circ\mathbf{1}^{N\cap T}\right)\\ &\times\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}. &(3.31)\end{aligned}$$

The cluster-robust estimator (3.30) not only considers both firm and time effects, but also addresses the underestimation issue above using the jackknife method. This estimator is thus superior to both the single clustered CR3 estimator (Bell and McCaffrey, 2002; Cameron et al., 2008) and the double clustered CR0 estimator Cameron et al. (2011); Thompson (2011). It is impossible to further aggregate the filling inside the sandwich estimator (3.31) above because three modified residuals $\tilde{\mathbf{u}}^{N}$, $\tilde{\mathbf{u}}^{T}$ and $\tilde{\mathbf{u}}^{N\cap T}$ are different with each other.

Unlike single clustered variance estimators such as (3.24), double clustered variance estimators such as (3.30) adjust themselves cluster by cluster. Instead of this pesky expression, one can directly generalize single clustered variance estimators to derive double clustered variance estimators as follows.

$$\widehat{\mathrm{Var}}_{\mathrm{CR5}}^{N\cup T}\left[\hat{\boldsymbol{\beta}}\right] = \left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1}\mathbf{X}^{\top}\left(\tilde{\mathbf{u}}^{N\cup T}\tilde{\mathbf{u}}^{N\cup T\top}\circ\mathbf{1}^{N\cup T}\right)\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X}\right)^{-1} \tag{3.32}$$

$$\tilde{\mathbf{u}}^{N\cup T} = \sqrt{NT/((N-1)(T-1))}\left(\mathbf{I}-\mathbf{P}\circ\mathbf{1}^{N\cup T}\right)^{-1}\hat{\mathbf{u}}. \tag{3.33}$$

Unlike the estimator (3.31), which adjusts the original estimator cluster by cluster, the estimator (3.32) tunes the residual based on one multiplication, so is a more natural generalization, which alters the indicator matrix $\mathbf{1}^{N}$ by $\mathbf{1}^{N\cup T}$. However, this version has some practical concerns. First, the inverse of the matrix $\mathbf{I}-\mathbf{P}\circ\mathbf{1}^{N\cup T}$ does not always exist. In practice, one can use eigendecomposition similar to the method in (3.19), but the finite sample properties can be more complicated. Second, inverting an $NT \times NT$ matrix is numerically expensive as it requires significant computing time. On the other hand, estimating CR3 is cheaper because the computational complexity of matrix inversion is $\mathcal{O}\left(n^{3}\right)$. Therefore, I compare the performance of these cluster-by-cluster and multiple-cluster estimators in this paper.

3.1.1 The Model

Here I impose additional assumptions to derive several asymptotic and finite sample properties. I assume a single regression below in particular with N firms, T times, and $K = 1$ regressor to pursue simplicity.

$$y_{it} = x_{it}\beta + u_{it} \tag{3.34}$$

$$\mathbf{y}_{i} = \mathbf{x}_{i}\beta + \mathbf{u}_{i} \tag{3.35}$$

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{u}, \tag{3.36}$$

where

$$x_{it} = \nu_i + \nu_t + \nu_{it} \tag{3.37}$$

$$u_{it} = \varepsilon_i + \varepsilon_t + \varepsilon_{it}. \tag{3.38}$$

(3.34)–(3.36) express the same regression and are interchangeable. νs and εs are assumed to be independent and identically distributed with zero mean without loss of generality. I assume εs and νs to have finite second and fourth moments, respectively. The regressor x and the error u are clustered by firm and time together according to this specification. First, ν_i (ε_i) is the source of firm effects since both x_{it} (u_{it}) and $x_{it'}$ ($u_{it'}$) consist of it. Second, ν_t (ε_t) is the source of time effects since both x_{it} (u_{it}) and $x_{i't}$ ($u_{i't}$) consist of it. Third, ν_{it} (ε_{it}) is the source of idiosyncratic effects. In detail, the covariance of x and that of u are

$$\mathrm{Cov}\left[x_{it}, x_{i't'}\right] = \begin{cases} \mathrm{E}\left[\nu_{i=1}^2\right] + \mathrm{E}\left[\nu_{t=1}^2\right] + \mathrm{E}\left[\nu_{it=11}^2\right] & i = i' \cap t = t' \\ \mathrm{E}\left[\nu_{i=1}^2\right] & i = i' \cap t \neq t' \\ \mathrm{E}\left[\nu_{t=1}^2\right] & i \neq i' \cap t = t' \\ 0 & i \neq i' \cap t \neq t' \end{cases} \tag{3.39}$$

$$\mathrm{Cov}\left[u_{it}, u_{i't'}\right] = \begin{cases} \mathrm{E}\left[\varepsilon_{i=1}^2\right] + \mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\varepsilon_{it=11}^2\right] & i = i' \cap t = t' \\ \mathrm{E}\left[\varepsilon_{i=1}^2\right] & i = i' \cap t \neq t' \\ \mathrm{E}\left[\varepsilon_{t=1}^2\right] & i \neq i' \cap t = t' \\ 0 & i \neq i' \cap t \neq t' \end{cases}, \tag{3.40}$$

and correspondingly the correlation of x and that of u are

$$\operatorname{Corr}\left[x_{it}, x_{i't'}\right] = \begin{cases} 1 & i = i' \cap t = t' \\ \frac{\mathrm{E}\left[\nu_{i=1}^2\right]}{\mathrm{E}\left[\nu_{i=1}^2\right]+\mathrm{E}\left[\nu_{t=1}^2\right]+\mathrm{E}\left[\nu_{it=11}^2\right]} & i = i' \cap t \neq t' \\ \frac{\mathrm{E}\left[\nu_{t=1}^2\right]}{\mathrm{E}\left[\nu_{i=1}^2\right]+\mathrm{E}\left[\nu_{t=1}^2\right]+\mathrm{E}\left[\nu_{it=11}^2\right]} & i \neq i' \cap t = t' \\ 0 & i \neq i' \cap t \neq t' \end{cases} \tag{3.41}$$

$$\operatorname{Corr}\left[u_{it}, u_{i't'}\right] = \begin{cases} 1 & i = i' \cap t = t' \\ \frac{\mathrm{E}\left[\varepsilon_{i=1}^2\right]}{\mathrm{E}\left[\varepsilon_{i=1}^2\right]+\mathrm{E}\left[\varepsilon_{t=1}^2\right]+\mathrm{E}\left[\varepsilon_{it=11}^2\right]} & i = i' \cap t \neq t' \\ \frac{\mathrm{E}\left[\varepsilon_{t=1}^2\right]}{\mathrm{E}\left[\varepsilon_{i=1}^2\right]+\mathrm{E}\left[\varepsilon_{t=1}^2\right]+\mathrm{E}\left[\varepsilon_{it=11}^2\right]} & i \neq i' \cap t = t' \\ 0 & i \neq i' \cap t \neq t' \end{cases}. \tag{3.42}$$

While νs and εs are independent and identically distributed, xs and us are neither independent nor identically distributed, so this model violates the classical assumptions. (3.41) and (3.42) govern the intensity of firm and time effects of xs and us, respectively. This model is clustered by firm when both $\mathrm{E}\left[\nu_{i=1}^2\right]$ and $\mathrm{E}\left[\varepsilon_{i=1}^2\right]$ are nonzero, clustered by time when both $\mathrm{E}\left[\nu_{t=1}^2\right]$ and $\mathrm{E}\left[\varepsilon_{t=1}^2\right]$ are nonzero, and clustered by both firm and time when these four second moments are collectively nonzero.

This model allows xs and us to be clustered differently as well. For example, the regression can simultaneously have x and u that are exclusively clustered by either firm or time. When there exists such a misalignment, both firm and time effects do not inflate the standard error of the estimator of β (Petersen, 2009). In practice, however, the cacophony rarely happens because (i) many dependent variables that appear in finance literature are clustered by both firm and time, and (ii) most finance researchers employ regression models with multiple independent variables (Thompson, 2011).

Though fixed and random effects models are applicable, I only examine clustered standard errors in this study following the literature. The OLS estimator of β is

$$\hat{\beta} = \left(\mathbf{x}^\top \mathbf{x}\right)^{-1} \mathbf{x}^\top \mathbf{y} \tag{3.43}$$

$$= \left(\sum_{i=1}^{N} \mathbf{x}_i^\top \mathbf{x}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{x}_i^\top \mathbf{y}_i \tag{3.44}$$

$$= \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2\right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} y_{it}. \tag{3.45}$$

3.1.2 Asymptotic Property

Since νs and εs inside x and u are independent and identically distributed, one can apply some asymptotic theorems such as the law of large numbers and the central limit theorem.

Proposition 3.1 (Consistency). *$\hat{\beta} \xrightarrow{p} \beta$ as $(N,T) \to (\infty,\infty)$.*

Proof. See Appendix 3.A.1.

Proposition 3.1 is not surprising since there is no endogeneity concern in this model. However, the proposition is important as the consistency of the estimator separately demands $N \to \infty$ and $T \to \infty$. In other words, just $NT \to \infty$ does not guarantee the estimator to be consistent since it is not a sufficient but a necessary condition of $(N,T) \to (\infty,\infty)$. This is problematic when N and T are disproportionate, and such an imbalance is common in many cases as aforementioned (Skoulakis, 2008).

Proposition 3.2 (Asymptotic normality). *As $(N,T) \to (\infty,\infty)$,*

$$\frac{\sqrt{NT}\left(\hat{\beta} - \beta\right)}{\sqrt{\widehat{\text{Avar}}\left[\hat{\beta}\right]}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\widehat{\text{Avar}}\left[\hat{\beta}\right] = NT\left(\mathbf{x}^\top \mathbf{x}\right)^{-1} \mathbf{x}^\top \left(\hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{N \cup T}\right) \mathbf{x} \left(\mathbf{x}^\top \mathbf{x}\right)^{-1}$.

Proof. See Appendix 3.A.2

It is noteworthy that $\widehat{\text{Avar}}\left[\hat{\beta}\right]$ is not a consistent estimator of $\text{Avar}\left[\hat{\beta}\right]$ in this model. Though most econometric studies articulate asymptotic normality by demonstrating $\sqrt{N}\left(\hat{\theta}-\theta\right)\xrightarrow{d}\mathcal{N}\left(0,\text{Avar}\left[\hat{\theta}\right]\right)$, the asymptotic variance of $\hat{\beta}$ diverges to infinity as N and T separately diverge to infinity and hence $\hat{\beta}$ per se does not directly converge in distribution to a normal distribution. Such behavior is distinct from how one-way clustered standard error estimators behave because a standard error estimator clustered by firm is $\sqrt{N}$-consistent and its counterpart clustered by time is $\sqrt{T}$-consistent, respectively (Petersen, 2009). Fortunately, however, its asymptotic normality is well preserved after scaling by the seemingly asymptotic variance estimator.

3.1.3 Finite Sample Property

Though the variance estimator is inconsistent, asymptotic inference still works based on Propositions 3.1 and 3.2. The variance estimator is also biased in finite samples as $\text{E}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top\right]$ is a biased estimator of $\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]$.

Proposition 3.3. $\text{Bias}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top,\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]\right]=\text{E}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top-\mathbf{u}\mathbf{u}^\top\right]=\text{E}\left[\mathbf{P}\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]\mathbf{P}\right]-\text{E}\left[\mathbf{P}\right]\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]-\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]\text{E}\left[\mathbf{P}\right]$, *where* $\mathbf{P}=\mathbf{x}\left(\mathbf{x}^\top\mathbf{x}\right)^{-1}\mathbf{x}^\top$.

Proof. $\hat{\mathbf{u}}=\mathbf{y}-\mathbf{x}\hat{\beta}=\mathbf{y}-\mathbf{P}\mathbf{y}=\mathbf{x}\beta+\mathbf{u}-\mathbf{P}\mathbf{x}\beta-\mathbf{P}\mathbf{u}=\left(\mathbf{I}_{NT}-\mathbf{P}\right)\mathbf{u}$, so

$$\begin{aligned}\text{E}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top\right]&=\text{E}\left[\left(\mathbf{I}_{NT}-\mathbf{P}\right)\mathbf{u}\mathbf{u}^\top\left(\mathbf{I}_{NT}-\mathbf{P}\right)\right]\\&=\text{E}\left[\mathbf{u}\mathbf{u}^\top+\mathbf{P}\mathbf{u}\mathbf{u}^\top\mathbf{P}-\mathbf{P}\mathbf{u}\mathbf{u}^\top-\mathbf{u}\mathbf{u}^\top\mathbf{P}\right]\\&=\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]+\text{E}\left[\mathbf{P}\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]\mathbf{P}\right]-\text{E}\left[\mathbf{P}\right]\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]-\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]\text{E}\left[\mathbf{P}\right],\end{aligned}$$

as $\mathbf{u}$ and $\mathbf{x}$ are independent. □

$\text{E}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top\right]$ often underestimates $\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]$. One toy model with u and x that are independent and identically distributed standard normal demonstrates the downward bias. In this toy model, $\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]$, $\text{E}\left[\mathbf{P}\right]$, and $\text{E}\left[\mathbf{P}\text{E}\left[\mathbf{u}\mathbf{u}^\top\right]\mathbf{P}\right]$ are $\mathbf{I}_{NT}$, $\mathbf{I}_{NT}/NT$, and $\text{E}\left[\mathbf{P}\right]=\mathbf{I}_{NT}$, respectively, so $\operatorname{Bias}\left[\hat{\mathbf{u}}\hat{\mathbf{u}}^\top, \mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]\right] = -\mathbf{I}_{NT}/NT$, which converges to $\mathbf{0}_{NT}$ as either N or T diverges to infinity, is negative definite.

One remedy for this underestimation tunes $\hat{\mathbf{u}}$ by multiplying $\sqrt{NT/(NT-1)}$, so $\mathrm{E}\left[\tilde{\mathbf{u}}\tilde{\mathbf{u}}^\top\right]$ with $\tilde{\mathbf{u}} = \sqrt{NT/(NT-1)}\hat{\mathbf{u}}$ is unbiased when xs are balanced. This remedy works with $\sqrt{NT/(NT-K)}$ for a multiple regression but doesn't work for a regression with unbalanced xs. Another remedy employs leverages to tune the estimator. That is, this remedy divides $\hat{u}_{it}$ by $\sqrt{1 - x_{it}^2/\left(\mathbf{x}^\top\mathbf{x}\right)}$ or $\sqrt{1 - \mathbf{x}_{it}^\top\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{x}_{it}}$ for a multiple regression. In other words, the remedy tunes $\hat{\mathbf{u}}$ by multiplying $\operatorname{diag}(\mathbf{M})^{-1/2}$, where $\mathbf{M} = \mathbf{I}_{NT} - \mathbf{P}$, so $\mathrm{E}\left[\tilde{\mathbf{u}}\tilde{\mathbf{u}}^\top\right]$ with $\tilde{\mathbf{u}} = \operatorname{diag}(\mathbf{M})^{-1/2}\hat{\mathbf{u}}$. This estimator is unbiased even with unbalanced xs but biased under heteroskedasticity. It is an almost unbiased estimator instead as the bias disappears with correct weighting (Horn et al., 1975).

Though $\mathrm{E}\left[\tilde{\mathbf{u}}\tilde{\mathbf{u}}^\top\right]$ is almost unbiased, it is not always the best choice in estimating correct standard errors. For example, MacKinnon and White (1985) introduce not only the HC2 estimator based on $\operatorname{diag}(\mathbf{M})^{-1/2}\hat{\mathbf{u}}$ but also the HC3 counterpart, the jackknife estimator, based on $\operatorname{diag}(\mathbf{M})^{-1}\hat{\mathbf{u}}$, though the latter overcorrects $\hat{\mathbf{u}}$ and hence is further biased upward. Despite this overestimation, Long and Ervin (2000), Davidson and MacKinnon (2004), Hausman and Palmer (2012), MacKinnon (2013), as well as MacKinnon and White (1985) report that HC3 rather than HC2 exhibits more accurate simulated P-values when heteroskedasticity exists and hence outperforms HC2 in finite samples. Likewise, Cribari-Neto (2004), Cribari-Neto and Lima (2009), and Poirier (2011) propose the HC4 estimator, which is not almost unbiased but consistent and considers high leverage effects, and advocate the use of HC4 over HC2 based on its preferrable behavior in finite samples.

Similar to this heteroskedasticity literature, the research about clustered standard errors provides consistent estimators that asymptotically work first and finite sample adjustments that improve these estimators second. For example, Froot (1989), Rogers (1993), and Petersen (2009) offer a one-way clustered standard error estimator, or the CR0 estimator, that is consistent but biased, and demonstrate simulation evidence that supports the use of this estimator against other inconsistent estimators without clustering. In contrast, Shao and Rao (1993), Bell and McCaffrey (2002), Cameron et al. (2008), Webb (2008), and Cameron and Miller (2015) examine how biased CR0 is in finite samples, and propose a few remedies such as CR3, which adopts $\operatorname{diag}\left(\mathbf{M}_1, \cdots, \mathbf{M}_N\right)^{-1} \hat{\mathbf{u}}$ rather than $\operatorname{diag}(\mathbf{M})^{-1}\hat{\mathbf{u}}$ of HC3, where $\mathbf{M}_i = \mathbf{I}_T - \mathbf{X}_i \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}_i^\top$, or some nonparametric variants such as WB, the wild bootstrap, to minimize the bias. Cameron et al. (2008) report that these finite sample adjustments, especially the wild bootstrap, rather than CR0 exhibit more accurate simulated P-values.

Though the clustered standard error literature such as Petersen (2009), Cameron et al. (2011), Thompson (2011), and Cameron and Miller (2015) offer a non-nested multiway clustered standard error estimator and compare the performance of the estimator and other competitors by simulating multiway clustering, its finite sample properties have not been highlighted in detail and no further finite sample adjustment has been discussed yet. First, Figure 3.7 of Petersen (2009) displays the simulated P-values from a two-way clustered standard error estimator as well as two one-way clustered estimators by firm and time, respectively. Despite the 1% significance level, about 5% of the two-way clustered simulated statistics falsely reject the true null hypothesis when $N = 10$ and $T = 1,000$, and when $N = 1,000$ and $T = 10$, respectively. On the other hand, about 1% of the statistics reject the null hypothesis when $N = T = 100$. As $\hat{\mathbf{u}}$ with no adjustment understates $\mathbf{u}$ in finite samples, the consistent two-way clustered standard error estimator is biased downward and hence the true null hypothesis is falsely over-rejected. Second, Table 1 of Cameron et al. (2011) displays the simulated P-values from an OLS standard error estimator, a one-way clustered standard error estimator, a two-way random effects standard error estimator, three different two-way clustered standard error estimators, respectively. Consistent with the first result, about 18.4%, 13.7%, and 11.3% of the two-way simulated statistics regarding $\hat{\beta}_1$ falsely reject the true null hypothesis concerning β_1 when x_1s and us are one-way clustered by firm and two-way clustered by firm and time, respectively, and (N, T) is equal to $(10, 10)$, $(10, 50)$, and $(10, 100)$, respectively. As the asymptotic properties require $(N, T) \to (\infty, \infty)$ rather than

$NT \to \infty$, the underestimation and over-rejection problems of the two-way estimator even worsen in finite samples. Third, Table 1 of Thompson (2011) displays the simulated P-values from an OLS standard error estimator, two one-way clustered standard error estimators, two one-way clustered standard error estimators with another-way fixed effects, and two different two-way clustered standard error estimators, respectively. Consistent with the first and second cases, Panels B and C exhibit that, even with the two-way clustered standard error estimators, more than 5% of the simulated statistics falsely reject the true null hypothesis in finite samples. Among others, five worst case scenarios display 20.1%, 17.6%, 17.4%, 16.2%, and 14.5%, respectively, despite the 5% significance level.

Overall, these findings confirm that, despite the asymptotic properties of the existing multiway clustered standard error estimators, their finite sample problems can be more critical in finite samples as the asymptotic properties require more stringent conditions under multiway clustering. Following this literature, I suggest several multiway clustered standard error estimators in this paper by generalizing these heteroskedasticity-consistent estimators and cluster-robust counterparts introduced above.

3.2 Monte Carlo Simulation

As the main purpose of this paper is to provide a multiway clustered standard error estimator that preserves the existing asymptotic properties of the consistent estimators while presents a better finite sample property, I explain the details of the Monte Carlo simulations used in this paper. Using the outcomes of these Monte Carlo simulations, I compare the finite sample performance of the candidate estimators numerically. I experiment both two-way clustered model and three-way clustered model.

3.2.1 Two-Way Clustered Model

I start with a balanced panel regression model with non-nested two-way clustering first as there already exists a corresponding clustered standard error estimator that is consistent but biased downward on average. Following the literature, I include a regression error clustered by both firm and time to check whether the two-way estimators I suggest are accurate under two-way clustering. Unlike the literature, however, I include together (i) a constant, (ii) a regressor without clustering, (iii) a regressor clustered by firm, (iv) a regressor clustered by time, and (v) a regressor clustered by both firm and time to check whether the estimators are robust under misspecification. Overall, the regression I simulate is

$$
\begin{aligned}
y_{it} =& \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it}, \;\; i \in \{1, \cdots, N\}, \;\; t \in \{1, \cdots, T\}, && (3.46)\\
\text{where } \mathbf{x}_{it} =& \begin{pmatrix} 1 & x_{it1} & x_{it2} & x_{it3} & x_{it4} \end{pmatrix}^{\top} \\
\boldsymbol{\beta} =& \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}^{\top}, && (3.47)
\end{aligned}
$$

and

$$
\begin{aligned}
\text{where } x_{it1} =& \nu_{it1} \\
x_{it2} =& \sqrt{\rho_2}\nu_{i2} + \sqrt{1-\rho_2}\nu_{it3} \\
x_{it3} =& \sqrt{\rho_3}\nu_{t4} + \sqrt{1-\rho_3}\nu_{it5} \\
x_{it4} =& \sqrt{\rho_4^N}\nu_{i6} + \sqrt{\rho_4^T}\nu_{t7} + \sqrt{1-\rho_4^N-\rho_4^T}\nu_{it8} \\
u_{it} =& \sqrt{\rho_u^N}\varepsilon_{i1} + \sqrt{\rho_u^T}\varepsilon_{t2} + \sqrt{1-\rho_u^N-\rho_u^T}\varepsilon_{it3}, && (3.48)\\
\text{where } \nu, \varepsilon \sim& \mathcal{N}(0,1), && (3.49)
\end{aligned}
$$

where N is the number of firms and T is the number of years. I simulate the regression (3.46) $S = 5,000$ times for each of 28 combinations from $N \geq T \in \{10, 20, 30, 40, 50, 75, 100\}$ and 3 additional combinations from $\begin{pmatrix} N & T \end{pmatrix} \in \left\{ \begin{pmatrix} 1,000 & 5 \end{pmatrix}, \begin{pmatrix} 500 & 10 \end{pmatrix}, \begin{pmatrix} 250 & 20 \end{pmatrix} \right\}$. Using the simulated statistics, I test the null hypothesis that $\beta = 1$ as (3.47) exhibits. (3.48) displays the recipe of the regressors and the regression error based on the ingredients in (3.49).

As each regressor weights its standard normal building blocks correspondingly, the regressors and the regression error respectively are also standard normal and hence the regressand is $\mathcal{N}(1,5)$ as well. Moreover, there exists neither multicollinearity nor endogeneity as the building blocks are independent and identically distributed. Furthermore, one can better contrast the performance of the competing clustered standard error estimators as the five regressors are clustered in five different ways. First, the inference for β_0 and β_4 requires a two-way clustered standard error estimator as the corresponding regressors are clustered by both firm and time. Second, the inference for β_2 and β_3 demands a one-way clustered standard error estimator as the corresponding regressors are clustered by either firm or time. Third, the inference for β_1 does not require a clustered standard error estimator as the corresponding regressor is not clustered. In practice, the existing two-way clustered standard error estimators consider multiway clustering but ignore finite sample bias, while the existing one-way clustered standard error estimators consider finite sample bias but ignore multiway clustering. In the spirit of MacKinnon and White (1985) and Cameron et al. (2008), I suggest a few multiway clustered standard error estimators that improve finite sample properties and examine whether one estimator outperforms another.

The error u is a random variable and has both firm effects and fixed effects, i.e.

$$u_{it} \sim \mathcal{N}\left(0, \sigma_u^2\right) \tag{3.50}$$

$$\mathrm{Corr}\left[u_{it}, u_{js}\right] = \begin{cases} 1 & i = j \cap t = s \\ \rho_u^N & i = j \cap t \neq s \\ \rho_u^T & i \neq j \cap t = s \\ 0 & i \neq j \cap t \neq s \end{cases}. \tag{3.51}$$

I exploit $\rho_u^N = \rho_u^T = 1/3$ hence u_{it} is normal with zero mean and unit variance, and $\rho_2 = \rho_3 = 1/2$ and $\rho_4^N = \rho_4^T = 1/3$ hence each element except the first is normal with zero mean and unit variance.

I further aggregate the regression model using the following notation to sketch the big picture.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{3.52}$$

$$\mathbf{x}^k = \begin{pmatrix} x_{11k} & x_{12k} & \cdots & x_{NTk} \end{pmatrix}^\top \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_k\right) \tag{3.53}$$

$$\mathbf{u} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_u\right) \tag{3.54}$$

$$\boldsymbol{\Sigma}_u = \sigma_u^2 \left(\mathbf{I}_{NT} + \rho_u^N \mathbf{I}_N \otimes \left(\mathbf{1}_T - \mathbf{I}_T\right) + \rho_u^T \left(\mathbf{1}_N - \mathbf{I}_N\right) \otimes \mathbf{I}_T\right) \tag{3.55}$$

$$= \sigma_u^2 \begin{pmatrix} \mathbf{I}_T + \rho_u^N \left(\mathbf{1}_T - \mathbf{I}_T\right) & \rho_u^T \mathbf{I}_T & \cdots & \rho_u^T \mathbf{I}_T \\ \rho_u^T \mathbf{I}_T & \mathbf{I}_T + \rho_u^N \left(\mathbf{1}_T - \mathbf{I}_T\right) & \cdots & \rho_u^T \mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \rho_u^T \mathbf{I}_T & \rho_u^T \mathbf{I}_T & \cdots & \mathbf{I}_T + \rho_u^N \left(\mathbf{1}_T - \mathbf{I}_T\right) \end{pmatrix}. \tag{3.56}$$

Where $\otimes$ is Kronecker product operator. $\boldsymbol{\Sigma}_u$ in (3.55) summarizes (3.51) above and exhibits the variance structure of $\mathbf{u}$ entirely. In detail, the first term $\mathbf{I}_{NT}$ is diagonal hence implies homoskedasticity. The second term $\rho_u^N \mathbf{I}_N \otimes \left(\mathbf{1}_T - \mathbf{I}_T\right)$ is a block diagonal matrix with 0s in the diagonal hence implies firm effects. The last term $\rho_u^T \left(\mathbf{1}_N - \mathbf{I}_N\right) \otimes \mathbf{I}_T$ is the matrix of diagonal matrices in off-diagonal blocks hence implies time effects. (3.56) provides more details by further expanding (3.55).

In addition, $\mathbf{x}^k$ in (3.53) summarizes the regressors variable by variable hence exhibits the variance structure of each independent variable. This variance structure can be written as follows.

$$\boldsymbol{\Sigma}_1 = \sigma_1^2 \mathbf{I}_{NT} \tag{3.57}$$

$$\boldsymbol{\Sigma}_2 = \sigma_2^2 \left(\mathbf{I}_{NT} + \rho_2 \mathbf{I}_N \otimes (\mathbf{1}_T - \mathbf{I}_T)\right) \tag{3.58}$$

$$\boldsymbol{\Sigma}_3 = \sigma_3^2 \left(\mathbf{I}_{NT} + \rho_3 (\mathbf{1}_N - \mathbf{I}_N) \otimes \mathbf{I}_T\right) \tag{3.59}$$

$$\boldsymbol{\Sigma}_4 = \sigma_4^2 \left(\mathbf{I}_{NT} + \rho_4^N \mathbf{I}_N \otimes (\mathbf{1}_T - \mathbf{I}_T) + \rho_4^T (\mathbf{1}_N - \mathbf{I}_N) \otimes \mathbf{I}_T\right). \tag{3.60}$$

Firstly, $\boldsymbol{\Sigma}_1$ in (3.57) is a diagonal matrix hence $\mathbf{x}^1$ has neither firm effects nor time effects. Secondly, $\boldsymbol{\Sigma}_2$ in (3.58) is a block diagonal matrix hence $\mathbf{x}^2$ only has firm effects. Thirdly, $\boldsymbol{\Sigma}_3$ in (3.59) is the matrix of diagonal matrices everywhere hence $\mathbf{x}^3$ only has time effects. Lastly, $\boldsymbol{\Sigma}_4$ in (3.60) is equivalent to $\boldsymbol{\Sigma}_u$ in (3.55) above hence $\mathbf{x}^4$ has both firm and time effects.

Instead of the variable by variable approach above, one is able to consider the variance structure $\mathrm{E}\left[\mathbf{x}_{it}\mathbf{x}_{i't'}^\top\right]$ across the regressors as well. Since they are independent of each other, there exists no multicollinearity hence the covariance among them is straightforward. Due to firm and time effects, however, there also exist both the covariance among firm clusters and the covariance among time clusters. These structures can be written as follows.

$$\mathrm{E}\left[\mathbf{x}_{it}\mathbf{x}_{i't'}^\top\right] = \begin{cases} \boldsymbol{\Sigma}^{NT} & i = i' \cup t = t' \\ \boldsymbol{\Sigma}^{N} & i = i' \cap t \neq t' \\ \boldsymbol{\Sigma}^{T} & i \neq i' \cup t = t' \\ \mathbf{0} & i \neq i' \cup t \neq t' \end{cases} \tag{3.61}$$

and

$$\mathbf{\Sigma}^{NT} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_4^2 \end{pmatrix} \tag{3.62}$$

$$\mathbf{\Sigma}^{N} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho_2\sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_4^N\sigma_4^2 \end{pmatrix} \tag{3.63}$$

$$\mathbf{\Sigma}^{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_3\sigma_3^2 & 0 \\ 0 & 0 & 0 & 0 & \rho_4^T\sigma_4^2 \end{pmatrix}. \tag{3.64}$$

Where $\mathbf{\Sigma}^{NT}$ is the covariance among regressors, $\mathbf{\Sigma}^{N}$ is the covariance among firm clusters and $\mathbf{\Sigma}^{T}$ is the covariance among time clusters, respectively. As aforementioned, these structures are diagonal and the asymptotic variance of the estimator $\hat{\boldsymbol{\beta}}$ is therefore diagonal as well. That is, according to Proposition 3.2, $\hat{\boldsymbol{\beta}}$ is asymptotically normal with the following distribution.

$$\begin{aligned} \hat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}\Bigg(\boldsymbol{\beta}, \frac{1}{NT}\sigma_u^2\left(\mathbf{\Sigma}^{NT}\right)^{-1} + \frac{T-1}{NT}\rho_u^N\sigma_u^2\left(\mathbf{\Sigma}^{NT}\right)^{-1}\mathbf{\Sigma}^{N}\left(\mathbf{\Sigma}^{NT}\right)^{-1} \\ + \frac{N-1}{NT}\rho_u^T\sigma_u^2\left(\mathbf{\Sigma}^{NT}\right)^{-1}\mathbf{\Sigma}^{T}\left(\mathbf{\Sigma}^{NT}\right)^{-1}\Bigg). \end{aligned} \tag{3.65}$$

Since the variance above is diagonal, this distribution can be rewritten separately as follows.

$$\hat{\beta}_0 \overset{a}{\sim} \mathcal{N}\left(\beta_0, \frac{\sigma_u^2}{NT}\left(1+(T-1)\rho_u^N+(N-1)\rho_u^T\right)\right) \tag{3.66}$$

$$\hat{\beta}_1 \overset{a}{\sim} \mathcal{N}\left(\beta_1, \frac{\sigma_u^2}{NT\sigma_1^2}\right) \tag{3.67}$$

$$\hat{\beta}_2 \overset{a}{\sim} \mathcal{N}\left(\beta_2, \frac{\sigma_u^2}{NT\sigma_2^2}\left(1+(T-1)\rho_u^N\rho_2\right)\right) \tag{3.68}$$

$$\hat{\beta}_3 \overset{a}{\sim} \mathcal{N}\left(\beta_3, \frac{\sigma_u^2}{NT\sigma_3^2}\left(1+(N-1)\rho_u^T\rho_3\right)\right) \tag{3.69}$$

$$\hat{\beta}_4 \overset{a}{\sim} \mathcal{N}\left(\beta_4, \frac{\sigma_u^2}{NT\sigma_4^2}\left(1+(T-1)\rho_u^N\rho_4^N+(N-1)\rho_u^T\rho_4^T\right)\right). \tag{3.70}$$

Using simulated panels, I estimate both βs and their standard errors. βs are estimated using OLS estimator and their standard errors are estimated using 19 different estimators including OLS, HC and CR. Then I compute 19 different t-statistics for each β to test the null hypothesis that it is equal to one. Whether the null hypothesis is rejected or not is determined with P-value equal to 10%, 5%, and 1%. First, I investigate if the size of each test statistic is accurate. Second, I test the normality of each sample using Kolmogorov–Smirnov D and Anderson–Darling A^2 statistics.

3.2.2 Three-Way Clustered Model

I also simulate a non-nested three-way panel regression to investigate if these improved clustered standard error estimators still work when higher order clustering exists. The regression is

$$
\begin{aligned}
y_{ijk} =& x_{ijk}\beta + u_{ijk},\ i \in \{1,2,\cdots,I\},\ j \in \{1,2,\cdots,J\},\ k \in \{1,2,\cdots,K\} && (3.71)\\
\mathrm{E}\left[x_{ijk}\right] =& \mathrm{E}\left[u_{ijk}\right] = 0,\ \forall i,\ j,\ k && (3.72)\\
\mathrm{E}\left[x_{ijk}x_{i'j'k'}\right] =& \begin{cases} \sigma_x^2 & i = i' \cap j = j' \cap k = k' \\ \rho_x \sigma_x^2 & (i = i' \cup j = j' \cup k = k') \setminus (i = i' \cap j = j' \cap k = k') \\ 0 & i \neq i' \cap j \neq j' \cap k \neq k' \end{cases} && (3.73)\\
\mathrm{E}\left[u_{ijk}u_{i'j'k'}\right] =& \begin{cases} \sigma_u^2 & i = i' \cap j = j' \cap k = k' \\ \rho_u \sigma_u^2 & (i = i' \cup j = j' \cup k = k') \setminus (i = i' \cap j = j' \cap k = k') \cdot \\ 0 & i \neq i' \cap j \neq j' \cap k \neq k' \end{cases} && (3.74)
\end{aligned}
$$

In other words, there exists three clustering effects due to the clusters with respect to I, J and K. That is, each observation has three indices i, j and k here and any pair of xs or us are clustered when they share at least one index. By defining $\mathbf{x} = \begin{pmatrix} x_{111} & \cdots & x_{IJK} \end{pmatrix}^\top$ and $\mathbf{u} = \begin{pmatrix} u_{111} & \cdots & u_{IJK} \end{pmatrix}^\top$, one can rewrite the second moments (3.73) and (3.74) as follows.

$$
\begin{aligned}
\mathrm{E}\left[\mathbf{x}\mathbf{x}^\top\right] =& \sigma_x^2 \Bigg(\mathbf{I}_{IJK} + \underbrace{\rho_x \mathbf{I}_I \otimes (\mathbf{1}_{JK} - \mathbf{I}_{JK})}_{I\text{-cluster effects}} \\
& + \underbrace{\rho_x (\mathbf{1}_I - \mathbf{I}_I) \otimes (\mathbf{I}_J \otimes \mathbf{1}_K)}_{J\text{-cluster effects}} + \underbrace{\rho_x (\mathbf{1}_I - \mathbf{I}_I) \otimes ((\mathbf{1}_J - \mathbf{I}_J) \otimes \mathbf{I}_K)}_{K\text{-cluster effects}} \Bigg) && (3.75)\\
\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right] =& \sigma_u^2 \Bigg(\mathbf{I}_{IJK} + \underbrace{\rho_u \mathbf{I}_I \otimes (\mathbf{1}_{JK} - \mathbf{I}_{JK})}_{I\text{-cluster effects}} \\
& + \underbrace{\rho_u (\mathbf{1}_I - \mathbf{I}_I) \otimes (\mathbf{I}_J \otimes \mathbf{1}_K)}_{J\text{-cluster effects}} + \underbrace{\rho_u (\mathbf{1}_I - \mathbf{I}_I) \otimes ((\mathbf{1}_J - \mathbf{I}_J) \otimes \mathbf{I}_K)}_{K\text{-cluster effects}} \Bigg). && (3.76)
\end{aligned}
$$

For both $\mathrm{E}\left[\mathbf{x}\mathbf{x}^\top\right]$ and $\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$, the number of nonzero elements is the function of the size of the clusters I, J and K, i.e. $f(I,J,K)$ is equal to $IJK(I(J-1)+J(K-1)+K(I-1)+1)$.

Hence f increases quadratically with I, J and K and f subject to $g(I,J,K) = IJK$ is therefore minimized at $I^* = J^* = K^* = \sqrt[3]{g}$. I simulate 5,000 panels using four different $I \times J \times K (= 1,728)$ combinations: (3,24,24), (6,12,24), (8,8,12) and (12,12,12). With $g = 1,728$, f is equal to 1,157,760 for the first combination, whereas it is equal to 686,016 for the last combination. The former is about 70% bigger than the latter hence its standard error estimator may perform worse than its counterpart.

I generate $\mathbf{x}$ and $\mathbf{u}$ as follows. First, I compute the second moments (3.75) and (3.76) using all parameters: I, J, K, σ_x^2, ρ_x, σ_u^2, and ρ_u. Though this setup employs only one parameter ρ that governs all clustering effects, it can be easily generalized. I exploit $\rho_x = \rho_u = 0.03$ because there exist a lot of off-diagonal elements. Second, I compute two lower triangular matrices $\mathbf{L}_x$ and $\mathbf{L}_u$ that respectively correspond to $\mathrm{E}\left[\mathbf{x}\mathbf{x}^\top\right]$ and $\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$ using Cholesky decomposition. Third, I generate two independent IJK-dimensional standard normal random vectors $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$. Fourth, I generate $\mathbf{x} = \mathbf{L}_x \boldsymbol{\varepsilon}_1$ and $\mathbf{u} = \mathbf{L}_u \boldsymbol{\varepsilon}_2$ using them hence both x and u have zero mean and unit variance, respectively.

The estimator $\hat{\beta}$ here is asymptotically normal as well with the following distribution.

$$\hat{\beta} \overset{a}{\sim} \mathcal{N}\left(\beta, \frac{\sigma_u^2}{IJK\sigma_x^2}\left(1 + (I(J-1) + J(K-1) + K(I-1))\rho_x \rho_\varepsilon\right)\right), \tag{3.77}$$

and I compute three-way clustered variance estimates using the sandwich estimator (Cameron and Miller, 2015) with no finite sample adjustment as follows.

$$\widehat{\mathrm{Var}}^{I\cap J\cap K}\left[\hat{\beta}\right] = \left(\mathbf{x}^\top\mathbf{x}\right)^{-1}\mathbf{x}^\top\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^\top \circ \mathbf{1}^{I\cup J\cup K}\right)\mathbf{x}\left(\mathbf{x}^\top\mathbf{x}\right)^{-1}. \tag{3.78}$$

This estimator is not always positive because there exist too many nonzero elements to be considered in the fillings of the sandwich estimator. I simulate 5,000 panels but exploit each panel only when its variance estimate is positive. About 5% (234 panels from the last combination) to 10% (479 panels from the first combination) of 5,000 panels are deleted as a result.

3.3 Main Result

In this section, I report the findings from the Monte Carlo simulations I introduce above. I discuss the result from the two-way clustered model first and the result from the three-way clustered model second. Section 3.3 contains further details of the models and Table 3.1 contains the estimators I employ.

3.3.1 Two-Way Clustered Model

As aforementioned, I simulate the two-way clustered model with 31 different $N \times T$ combinations first. 5,000 panels are generated for each combination.

Table 3.2 demonstrates the findings from the $\begin{pmatrix} 1,000 & 5 \end{pmatrix}$ combination, which is the most extreme one. The very first row contains the βs and the very first column contains the estimators I compute.

This table consists of five partitions. The first partition has two rows and focuses on the estimates of the βs. The second partition has 21 rows and focuses on the estimates of the standard errors. I compute t-statistics using the estimates above but there is no summary for these statistics in this table. Each of the last three partitions has 20 rows and focuses on the size. For each simulation, I use the null hypothesis that β is equal to 1 and reject it when the P-value of the t-statistic is less than 10%, 5%, and 1%, respectively.

The first row of the first partition of Table 3.2 contains $\boldsymbol{\beta}$ I exploit. The second row of this partition contains the average of 5,000 simulated $\boldsymbol{\beta}$ estimates. There exists neither multicollinearity nor endogeneity hence the OLS estimator of $\boldsymbol{\beta}$ is unbiased and consistent. As a result, the average is close to the true parameter.

3.3.1.1 Standard Error

The first row of the second partition of Table 3.2 contains the asymptotic standard error of $\hat{\boldsymbol{\beta}}$ I exploit. The second row of this partition contains the standard deviation of the simulated estimates. Here the OLS estimator of $\boldsymbol{\beta}$ is "asymptotically" normal hence its asymptotic standard error and the standard deviation from finite samples may not match closely. The standard deviation of $\hat{\beta}_0$s is 0.2627 hence about 1% bigger than its asymptotic reference. In contrast, the standard deviations of $\hat{\beta}_4$s and $\hat{\beta}_5$s are 0.1636 and 0.1279 hence about 11% and 15% smaller than their counterparts, respectively. When there exist both firm and time effects, asymptotic properties require both N and T to go to infinity as I analyze above. This affects $\hat{\beta}_3$s and $\hat{\beta}_4$s in particular because there exist time effects in x_3 and x_4 but $T = 5$ is too small here.

The third row of this partition contains the average of 5,000 simulated OLS standard error estimates. Here the OLS estimator underestimates the true standard deviation of $\hat{\boldsymbol{\beta}}$ except $\hat{\beta}_1$. x_1 has neither firm effects nor time effects hence, according to (3.67), u does not affect $\hat{\beta}_1$ even when it is doubly clustered. In addition, the average of the OLS estimates of $\hat{\beta}_2$s is about 27% smaller than its asymptotic reference while that of $\hat{\beta}_3$s is about 92% smaller than the counterpart. According to (3.68) and (3.69), these asymptotic standard errors are determined by N and T. Because $T = 5$ is smaller than $N = 1,000$, the OLS estimator less underestimates the former.

The next four rows display the performance of four HC standard error estimators. Since the regression errors are homoskedastic, the performance of these estimators is nearly identical to that of the OLS estimator. In addition, none of the finite sample adjustments improves the performance.

The next four rows display the performance of four NCR standard error estimators. These NCR estimators outperform the HC estimators with an intercept, x_2 and x_4. According to (3.66), (3.68) and (3.70), the estimators $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_4$ are affected by the firm effects of u. Because these NCR estimators capture the firm effects, they do not underestimate the asymptotic standard error of $\hat{\beta}_2$ and less underestimate the asymptotic standard errors of $\hat{\beta}_0$ and $\hat{\beta}_4$. Yet the NCR estimators still show three shortcomings. Though $N = 1,000$ is too large here, these estimators ignore the time effects hence (i) still underestimate their asymptotic references ($\hat{\beta}_0$, $\hat{\beta}_4$), (ii) even underperform when there exists not firm effects but time effects ($\hat{\beta}_3$) and (iii) the adjustments still are ineffective.

The next four rows display the performance of four TCR standard error estimators. TCR0 estimator outperform in estimating the asymptotic standard errors of $\hat{\beta}_0$, $\hat{\beta}_3$ and $\hat{\beta}_4$. In particular, the average of the TCR0 estimates of $\hat{\beta}_3$ is 0.0962 hence it is only 47% smaller than its asymptotic reference. Though this improvement from 91% to 47% is noticeable, TCR0 estimator still underestimates its counterpart. In addition, this estimator underperforms with x_1 and x_2. The average of the TCR0 estimates of $\hat{\beta}_2$ is 0.0092 hence it is 50% smaller than the asymptotic standard error.

Unlike HC and NCR estimators, TCR estimators are effectively improved by the adjustments. Firstly, TCR1 estimator reduces this underestimation using degrees of freedom. On average, it underestimates the asymptotic standard error of $\hat{\beta}_3$ by about 41%. Secondly, TCR2 estimator uses leverages and it underestimates its counterpart by about 19%. Lastly, TCR3 estimator uses the jackknife and it overestimates the counterpart by about 12%. According to this result, TCR2 and TCR3 are more effective than TCR1.

The last four rows of this partition display the performance of four NTCR standard error estimators. NTCR captures both firm and time effects exactly since the errors are doubly clustered. First, NTCR0 estimator underestimate its asymptotic reference on average by about 27% ($\hat{\beta}_0$), 24% ($\hat{\beta}_1$), 22% ($\hat{\beta}_2$), 48% ($\hat{\beta}_3$) and 51% ($\hat{\beta}_4$), respectively. That is, NTCR0 estimator tends to underestimate the asymptotic standard error in particular when the regressors have time effects. Second, NTCR1 estimator underestimates its counterpart on average by about 18%, 16%, 18%, 41% and 45%, respectively. NTCR1 estimator outperforms NTCR0 estimator everywhere but still underestimates the asymptotic standard errors of $\hat{\beta}_3$ and $\hat{\beta}_4$. Third, NTCR2 estimator underestimates the counterpart on average by about 11%, 13%, 16%, 28% and 33%, respectively. NTCR2 estimator even outperforms NTCR1 estimator everywhere. Fourth, NTCR3 estimator underestimates the asymptotic standard error of $\hat{\beta}_2$ on average by 3% and overestimates the rest by 23% ($\hat{\beta}_0$), 14% ($\hat{\beta}_1$), 12% ($\hat{\beta}_3$)

and 2% ($\hat{\beta}_4$), respectively. This result is ambiguous because NTCR3 estimator outperforms NTCR2 estimator with x_2, x_3 and x_4 but underperforms it with an intercept and x_1. Fifth, the performance of NTCR2 and that of NTCR4 are matching, and the performance of NTCR3 and that of NTCR5 are matching as well.

3.3.1.2 P-Value

As aforementioned, the third, fourth, and fifth partitions of Table 3.2 displays the performance of 19 P-values computed using 19 standard error estimators. Though standard error estimates are important, they are often employed as a means to test hypotheses. I examine whether these standard errors a posteriori replicate the size I intend a priori.

The third partition displays the performance of these P-values at a 10% significance level. The first row of this partition contains the P-value I intend ex ante. If one standard error is accurate, then the simulation must replicate this P-value ex post as well. In other words, I must reject the true null hypothesis about 500 times since 5,000 panels are simulated.

The second row of this partition displays the performance of OLS P-value. Though only about 10% of $\hat{\beta}_1$s are significant here, OLS P-value incorrectly rejects the null hypothesis too ofte. In particular, it rejects the true null hypothesis $\beta_0 = 1$ almost always with 92.6% probability.

The next four rows display the performance of HC P-values but none of them outperforms OLS P-value since there exists no heteroskedasticity. All Three finite sample adjustments are ineffective.

The next four rows display the performance of NCR P-values and they outperform with x_2, which has firm effects only. However, they still reject the true null too often when the regressors have time effects. All three finite sample adjustments are ineffective as well.

The next four rows display the performance of TCR P-values. TCR0 P-value outperforms NCR P-values when the regressors have time effects but underperform otherwise. However, the adjustments are so effective that TCR3 P-value closely replicate the true P-value.

The last four rows of this partition displays the performance of NTCR P-values. First, NTCR0 P-value underperforms TCR3 counterpart everywhere except the third column. In particular, NTCR0 P-value incorrectly rejects the null hypotheses $\beta_0 = 1$, $\beta_3 = 1$ and $\beta_4 = 1$ about 4 to 6 times more than the true P-value. Second, NTCR1 P-value outperforms NTCR0 counterpart everywhere but still underperforms TCR3 counterpart everywhere except the third column. Third, NTCR2 P-value even outperforms NTCR1 counterpart but still underperforms TCR3 counterpart everywhere except the third column. Fourth, NTCR3 P-value outperforms TCR3 counterpart everywhere. Fifth, consistent with the previous findings, NTCR2 and NTCR4 perform similarly in terms of the size. Likewise, NTCR3 and NTCR5 perform similarly with marginal difference.

The fourth and fifth partitions adopts 5% and 1% in rejecting the null hypothesis, respectively, rather than 10%. Though some estimators such as NCRs at β_2 and TCR3 at β_1, β_3, and β_4 are somewhat impressive, they exhibit unreasonable numbers at the rest. On the other hand, both NTCR3 and NTCR5 behave similarly and display the best performance among others. With a 5% significance level, both NTCR3 and NTCR5 reject the null hypothesis 7.34% at most and 3.42% at least. With a 1% significance level, both NTCR3 and NTCR5 reject the null hypothesis 3.56% at most and 0.98% at least. In addition, both NTCR2 and NTCR4 present the second best performance. Both reject the null hypothesis 15.60% at most and 9.78% at least with a 5% significance level, and 8.08% at most and 3.22% at least with a 1% significance level, respectively.

These findings imply that even when the regression error u is doubly clustered, double clustered standard errors are not always the best. When there is no finite sample adjustment but T (N) is too small, double clustered standard errors are not noticeably different from time (firm) clustered standard errors. According to (3.70), the asymptotic standard error of $\hat{\beta}_4$ consists of the firm effects $(T-1)\rho_u^N \rho_4^N$ and the time effects $(N-1)\rho_u^T \rho_4^T$. When T (N) is smaller than N (T), the time (firm) effects determines the asymptotic standard error more. Because time (firm) clustered standard errors capture these time (firm) effects, they less underperform double clustered standard errors.

Moreover, the findings present the monotonic relation among the finite sample adjustments. For instance, double clustered standard errors monotonically increase from NTCR0 to NTCR3 on average. NTCR1 outperforms NTCR0, NTCR2 outperforms NTCR1, and NTCR3 outperforms NTCR2 on average as well in terms of size. NTCR3, however, overestimates the asymptotic standard error while NTCR2 rather than NTCR3 displays more accurate estimates on average. In addition, these finite sample adjustments are more effective when there exist only few clusters. For example, the adjustments improve time clustered standard errors dramatically while they are ineffective for firm clustered standard errors.

Furthermore, this result, all things considered, indicates that NTCR3 and NTCR5 mostly outperform the others on average. Two potential challengers are NTCR2 and NTCR4 but they still have the over-rejection problem, despite their more accurate standard error estimates on average. For example, hey do not overestimate the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_3$ on average hence is less biased. However, they reject the true null hypothesis about 2 to 3 times more than the true P-value everywhere.

3.3.1.3 *t*-Statistic

Since I compute t-statistics already, I further investigate the distribution of each sample and test its normality using Kolmogorov–Smirnov D, Cramér–von Mises W^2, and Anderson–Darling A^2. They test the null hypothesis that the observations are standard normally distributed. Previous studies primarily suggest either NTCR0 (Thompson, 2011) or NTCR1 (Cameron et al., 2011; Cameron and Miller, 2015) when there exist both firm and time effects. Petersen (2009) even suggests NCR1 when there exist these effects but $N \gg T$. Figure 3.3 above discusses the use of one-way estimators under two-way clustering. Focusing in particular on β_3 and β_4, I compare two t-statistics computed using two double clustered standard errors NTCR1 and NTCR3.

Figure 3.4 displays the histograms of the t-statistics of $\hat{\beta}_4$s generated using 4 different

$N \times T (= 5,000)$ combinations and 3 different two-way clustered standard error estimators. This figure consists of 4 left histograms from NTCR1, 4 middle histograms from NTCR3, and 4 right histograms from NTCR5, respectively. The 4 $N \times T$ combinations are $\begin{pmatrix} 1,000 & 5 \end{pmatrix}$, $\begin{pmatrix} 500 & 10 \end{pmatrix}$, $\begin{pmatrix} 250 & 20 \end{pmatrix}$, and $\begin{pmatrix} 100 & 50 \end{pmatrix}$, respectively. Each histogram provides the standard normal density as well as two test statistics.

The upper-left histogram, i.e. the (1,1) histogram hereafter, indicates that the t-statistics computed using NTCR1 estimates are not standard normally distributed. About 22% of the t-statistics incorrectly reject the null hypothesis $\beta_4 = 1$ at a 5% significance level. In addition, Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling tests reject the null hypothesis at 1% significance level.

Though the (1,2) histogram also indicates that the t-statistics computed using NTCR3 estimates are not standard normally distributed, the behavior of these t-statistics is different in detail. About 3.4% of the t-statistics reject the true null at 5% significance level. Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling tests reject the null at 1% significance level as well but the tests statistics are noticeably smaller than those above. D decreases from 0.13 to 0.04, W^2 decreases from 37.80 to 2.13, and A^2 decreases from 468.12 to 19.12. In addition, both variance and excess kurtosis of the observations decrease from 2.91 to 0.82 and from 1.58 to 0.49, respectively. According to these findings, the latter histogram is closer to the standard normal density than the former one. Consistent with the previous findings, the outcomes from NTCR5, the (1,3) histogram, are similar to those from NTCR3 and marginally different.

The (2,1) histogram still exhibits that the t-statistics computed using NTCR1 estimates are not standard normally distributed. D, W^2, and A^2 reject the null hypothesis at 1% significance level. On the other hand, the (2,2) and (2,3) histograms exhibit that the t-statistics computed using NTCR3 and NTCR5 estimators are standard normally distributed. As T increases from 5 to 10, D, W^2, and A^2 decrease from 0.07 to 0.01, 11.84 to 0.11, and 126.31 to 0.86 (0.87), respectively. NTCR3 and NTCR5, in a nutshell, outperform NTCR1.

The other six histograms are consistent with the first six histograms as well. Both the (3,1) and the (4,1) histograms reject the null hypothesis that the t-statistics computed using NTCR1 are standard normally distributed at 1% significance level, though D, W^2, and A^2 decrease from 0.07 to 0.03 to 0.02, from 11.84 to 2.12 to 0.83, and from 126.31 to 20.95 to 6.94 as T increases from 10 to 20 to 50. In contrast, the (4,2) histogram does not reject the null hypothesis. The histogram only rejects the null hypothesis of of Anderson–Darling test at 10% significance level but does not reject at 5% significance level. The (4,3) histogram exhibits the numbers similar to the (4,2) histogram, but the Cramér–von Mises test does not reject the null hypothesis as well.

Because I use $N \geq T$, the time effects dominates the firm effects in determining the asymptotic standard error. Since there exist time effects in x_3, I further focus on β_3. Figure 3.6 displays the histograms of the t-statistics of $\hat{\beta}_3$s generated using 12 different combinations. This figure is consistent with Figure 3.4. First, all the histograms in the first column reject the null hypothesis of Kolmogorov–Smirnov, Cramér–von Mises, and Anderson–Darling tests at least at a 5% significance level while the last two histograms in the second and third columns do not reject the null hypothesis even at 10% level. Second, D, W^2, and A^2 decrease as T increases in Panel A. Third, the finite sample adjustment becomes less effective as T increases. In short, NTCR3 and NTCR5 outperform NTCR0, though they overestimate the asymptotic standard error.

3.3.1.4 Asymptotic Property

Figures 3.4 and 3.6 indicate that both NTCR1 and NTCR3 standard errors are improved as T increase. Since there exists both firm and time effects, both N and T determine the asymptotic behavior. First, I examine this asymptotic behavior using different $N \times T (= 5,000)$ combinations. Second, focusing in particular on β_4, I track the performance of the double clustered standard errors using several $N \geq T$ combinations from $\{10, 20, 30, 40, 50, 75, 100\}$.

Tables 3.3–3.5 replicate Table 3.2 but increase T from 5 to 10 to 20 to 50, respectively hence N decreases from 1,000 to 500 to 200 to 100 as a result. First, both OLS and HC standard errors become better in the first, fourth and fifth columns but become worse in the third column as T increases. According to (3.68), the asymptotic standard error of $\hat{\beta}_2$ increases as T increases. Because OLS and HC standard errors capture neither firm effects nor time effects, they therefore more underestimate the asymptotic standard error as T increases. According to (3.66), (3.69), and (3.70), in contrast, each of the asymptotic standard errors decreases as N decreases. Because the decrease of N is faster than the increase of T, the asymptotic standard errors of $\hat{\beta}_0$ and $\hat{\beta}_4$ decrease as N decreases Thus, OLS and HC standard errors less underestimate their references as N decreases.

Second, NCR standard errors become better everywhere except the third column as T increases. As I analyze above, the influence of the firm (time) effects increase (decrease) as T increases. Because the influence of the time effects becomes smaller, NCR standard errors less underestimate this influence as T increases. On the other hand, NCR standard errors capture the influence of firm effects hence they track this influence as T increases but lose asymptotic properties as N decreases.

Third, TCR standard errors demonstrate mixed findings. According to (3.26), the filling inside the TCR sandwich estimator exploits 5,000,000 elements inside the outer product of the residual when $N = 1,000$ and $T = 5$ while, according to (3.31), the filling inside the NTCR one exploits 5,020,000 elements. TCR standard errors therefore are not noticeably different from NTCR ones when $N \gg T$. There is a trade-off as T increases. They get asymptotic properties as T increases but lose the coverage as N decreases. When $N = 100$ and $T = 50$, the filling inside the TCR estimator exploits 500,000 elements while that inside the NTCR estimator exploits 745,000 elements. Due to this trade-off, TCR standard error estimators become better first and then become worse in the first and the last columns. They become better in the fourth column since they get asymptotic properties as T increases while become worse in the third column since the influence of the firm effects increases as T increases.

Fourth, NTCR standard errors by and large converge to the true P-value as T increases and NTCR0 standard error in particular converges dramatically. Unlike the other standard errors, NTCR3 and NTCR5 are close to the true P-value even when $N \gg T$. In general, NTCR3 and NTCR5 by and large performs best in all combinations.

As aforementioned Table 3.6 in particular focuses on β_4 and summarizes these findings. This table begins from the 10×10 combination and increases N from 10 to 100. Then it begins from the 20×20 combination again and increases N from 20 to 100 and so forth. Like Tables 3.2–3.5, Table 3.6 displays the average and standard deviation of $\boldsymbol{\beta}$ estimates, the average of 6 NTCR standard error estimates and the percentage of incorrectly significant t-statistics computed using these estimates.

According to the panel with $T = 10$ of this table, NTCR0 standard error estimator on average underestimates the asymptotic standard error by about 29% when $N = 10$ and about 25% when $N = 100$. The percentage of incorrectly significant t-statistics computed using NTCR0 estimates is 0.1944 when $N = 10$ and 0.1414 when $N = 100$. In contrast, NTCR3 estimator on average overestimates the asymptotic standard error by about 14% when $N = 10$ and about 4% when $N = 100$. The percentage of incorrectly significant t-statistics computed using NTCR3 estimates is 0.0528 when $N = 10$ and 0.0544 when $N = 100$. Consistent with the previous findings, NTCR5 exhibits reasonable properties similar to NTCR3.

These findings indicate that increasing N does not solve this problem. For instance, NTCR0 underestimates the asymptotic standard error by about 25% and rejects the true null hypothesis about three to four times more than the true P-value as long as $T = 10$. On the other hand, NTCR3 does not distort the size a lot, though it overestimates the asymptotic standard error a bit. Instead, the standard error estimators are noticeably improved as one goes from one to the next panel. According to the panel with $T = 100$, both NTCR0 and NTCR3 estimate the asymptotic standard error exactly and do not distort the size a lot, though NTCR0 rejects the true null about 20% more than its reference. This result implies that one must think of the panel as "finite" (i) if the regression error is doubly clustered and (ii) the panel contains only few clusters in one dimension.

3.3.2 Three-Way Clustered Model

I further simulate the triple clustered model with four different $I \times J \times K (= 1,728)$ combinations: (3,24,24), (6,12,24), (8,8,27) and (12,12,12). 5,000 panels are generated for each combination. Here I do not use any finite sample adjustment but focus on the problem that occurs when the panel has only "finite" clusters. Though the triple clustered standard error is consistent, it may be vulnerable to this problem.

Figure 3.8 displays the histograms of the t-statistics of $\hat{\beta}$s generated using 4 different $I \times J \times K (= 1,728)$ combinations. The first, second, third, and fourth histograms are from the triples (3,24,24), (6,12,24), (8,8,27) and (12,12,12), respectively. As I analyze above, the number of the nonzero elements of $\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$ for each triple is 1,157,760, 800,064, 784,512 and 686,016, respectively. Since the first (last) triple is worst (best) balanced, the t-statistics from this triple may underperform (outperform) those from the other three.

The first histogram consists of 4,521 $\hat{\beta}$s while the fourth histogram consists of 4,766 $\hat{\beta}$s. There exists no problem in estimating βs, but estimating the asymptotic standard error is less stable. The average of the triple clustered standard errors for the first triple on average underestimates the asymptotic standard error by 24% while that for the last triple underestimates the reference by about 14%. Likewise the t-statistics for the first triple reject the true null hypothesis about 4 times more than the true P-value while those for the last triple reject the null about 3 times more than the reference.

Collectively, this result implies that (i) the performance of one standard error is determined by the number of nonzero elements inside the expected value of the outer product of the error, (ii) even a small amount amount of correlation such as $\rho_x = \rho_u = 0.03$ largely affects the performance when the clusters are 3-dimensional and (iii) though the sandwich form without any finite sample adjustment is easy to use, it is vulnerable not to the "finite sample" problem but to the "finite cluster" problem. As aforementioned, the finite sample adjustment here may not be straightforward because $\mathbf{I} - \mathbf{P} \circ \mathbf{1}^{I \cup J \cup K}$ is not always invertible. Instead, the natural extension of the CR3 recipe (3.30) for practical purposes can be written as follows.

$$\begin{aligned}\widehat{\mathrm{Var}}_{\mathrm{CR3}}^{I \cup J \cup K}\left[\hat{\boldsymbol{\beta}}\right] =& \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{I}\left[\hat{\boldsymbol{\beta}}\right] + \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{J}\left[\hat{\boldsymbol{\beta}}\right] + \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{K}\left[\hat{\boldsymbol{\beta}}\right] \\ &- \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{I \cap J}\left[\hat{\boldsymbol{\beta}}\right] - \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{I \cap K}\left[\hat{\boldsymbol{\beta}}\right] - \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{J \cap K}\left[\hat{\boldsymbol{\beta}}\right] + \widehat{\mathrm{Var}}_{\mathrm{CR3}}^{I \cap J \cap K}\left[\hat{\boldsymbol{\beta}}\right]. \end{aligned} \tag{3.79}$$

3.4 Conclusion

In this paper, I discuss the bias of the well-known clustered standard error estimators, or non-standard standard error estimators. As introduced in many econometrics textbooks so far (Angrist and Pischke, 2008; Cameron and Trivedi, 2005; Campbell et al., 1997; Greene, 2003; Wooldridge, 2010), these standard error estimators exhibit reasonable asymptotic properties such as asymptotic normality. However, when there exists multiway clustering, asymptotic properties require harsh conditions to researchers, so the finite sample properties of the estimators get more important. There have been many attempts to investigate the bias and the size of the estimators in finite sample so far. Because of the extensive use of the clustered standard error estimators, their finite sample properties have been studied when there exists one-way clustering only. This paper extends this literature by comparing many candidate estimators that can address the bias issue and comparing their performance in finite sample using Monte Carlo simulations. All things considered, the NTCR3 estimator, the cluster-by-cluster jackknife estimator, and the NTCR5 estimators, the multiple-cluster jackknife estimator, exhibit the best performance in various finite sample environments. In particular, NTCR5 is attractive as more general, while NTCR3 is more practical as saves expensive numerical computations. Further research deserves as the Monte Carlo setup I employ only considers standard linear regressions.

3.A Proofs

3.A.1 Proof of Proposition 3.1

Proof. From (3.45),

$$\hat{\beta} = \beta + \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2 \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} u_{it},$$

so

$$\begin{aligned}
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}^2 =& \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i^2 + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_t^2 + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_{it}^2 \\
& + 2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \nu_t + 2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \nu_{it} + 2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_t \nu_{it} \\
\xrightarrow{p}& \mathrm{E}\left[\nu_{i=1}^2\right] + \mathrm{E}\left[\nu_{t=1}^2\right] + \mathrm{E}\left[\nu_{it=11}^2\right],
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} u_{it} =& \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \varepsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \varepsilon_t + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_i \varepsilon_{it} \\
& + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_t \varepsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_t \varepsilon_t + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_t \varepsilon_{it} \\
& + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_{it} \varepsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_{it} \varepsilon_t + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \nu_{it} \varepsilon_{it} \\
\xrightarrow{p}& 0,
\end{aligned}$$

and hence $\hat{\beta}$ converges in probability to β as (N, T) diverges to (∞, ∞). $\square$

3.A.2 Proof of Proposition 3.2

Proof. From Proposition 3.1,

$$\sqrt{NT}\left(\hat{\beta}-\beta\right)=\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}u_{it},$$

and

$$\begin{aligned}
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}u_{it}=&\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_i\varepsilon_i+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_i\varepsilon_t+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_i\varepsilon_{it}\\
&+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_t\varepsilon_i+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_t\varepsilon_t+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_t\varepsilon_{it}\\
&+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_{it}\varepsilon_i+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_{it}\varepsilon_t+\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\nu_{it}\varepsilon_{it}\\
\overset{a}{\sim}&\mathcal{N}\left(0,C_1\right)+\mathcal{N}\left(0,\mathrm{E}\left[\nu_{i=1}^2\right]\right)\mathcal{N}\left(0,\mathrm{E}\left[\varepsilon_{t=1}^2\right]\right)\\
&+\mathcal{N}\left(0,\mathrm{E}\left[\nu_{t=1}^2\right]\right)\mathcal{N}\left(0,\mathrm{E}\left[\varepsilon_{i=1}^2\right]\right),\\
\text{where } C_1=&T\mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right]+\mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right]\\
&+N\mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right]+\mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right]\\
&+\mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right]+\mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right]+\mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right].
\end{aligned}$$

Also,

$$\begin{aligned}
\widehat{\text{Avar}}\left[\hat{\beta}\right]=&\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2\right)^{-1}\left(C_2+(T-1)C_3+(N-1)C_4\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2\right)^{-1},\\
\text{where } C_2=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}^2\hat{u}_{it}^2\\
C_3=&\frac{1}{NT(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t\neq t'}^{T}x_{it}x_{it'}\hat{u}_{it}\hat{u}_{it'}\\
C_4=&\frac{1}{N(N-1)T}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{i\neq i'}^{N}x_{it}x_{i't}\hat{u}_{it}\hat{u}_{i't}.
\end{aligned}$$

and

$$\begin{aligned}
x_{it}^2\hat{u}_{it}^2 =& x_{it}^2u_{it}^2 - 2\left(\hat{\beta}-\beta\right)x_{it}^3u_{it} + \left(\hat{\beta}-\beta\right)^2 x_{it}^4 \\
=& \left(\nu_i+\nu_t+\nu_{it}\right)^2\left(\varepsilon_i+\varepsilon_t+\varepsilon_{it}\right)^2 \\
& -2\left(\hat{\beta}-\beta\right)\left(\nu_i+\nu_t+\nu_{it}\right)^3\left(\varepsilon_i+\varepsilon_t+\varepsilon_{it}\right) \\
& +\left(\hat{\beta}-\beta\right)^2\left(\nu_i+\nu_t+\nu_{it}\right)^4 \\
x_{it}x_{it'}\hat{u}_{it}\hat{u}_{it'} =& x_{it}x_{it'}u_{it}u_{it'} - \left(\hat{\beta}-\beta\right)x_{it}x_{it'}^2u_{it} - \left(\hat{\beta}-\beta\right)x_{it}^2x_{it'}u_{it'} + \left(\hat{\beta}-\beta\right)^2 x_{it}^2x_{it'}^2 \\
=& \left(\nu_i+\nu_t+\nu_{it}\right)\left(\nu_i+\nu_{t'}+\nu_{it'}\right)\left(\varepsilon_i+\varepsilon_t+\varepsilon_{it}\right)\left(\varepsilon_i+\varepsilon_{t'}+\varepsilon_{it'}\right) \\
& -\left(\hat{\beta}-\beta\right)\left(\nu_i+\nu_t+\nu_{it}\right)\left(\nu_i+\nu_{t'}+\nu_{it'}\right)^2\left(\varepsilon_i+\varepsilon_t+\varepsilon_{it}\right) \\
& -\left(\hat{\beta}-\beta\right)\left(\nu_i+\nu_t+\nu_{it}\right)^2\left(\nu_i+\nu_{t'}+\nu_{it'}\right)\left(\varepsilon_i+\varepsilon_{t'}+\varepsilon_{it'}\right) \\
& +\left(\hat{\beta}-\beta\right)^2\left(\nu_i+\nu_t+\nu_{it}\right)^2\left(\nu_i+\nu_{t'}+\nu_{it'}\right)^2 \\
x_{it}x_{i't}\hat{u}_{it}\hat{u}_{i't} =& x_{it}x_{i't}u_{it}u_{i't} - \left(\hat{\beta}-\beta\right)x_{it}x_{i't}^2u_{it} - \left(\hat{\beta}-\beta\right)x_{it}^2x_{i't}u_{i't} + \left(\hat{\beta}-\beta\right)^2 x_{it}^2x_{i't}^2 \\
=& \left(\nu_i+\nu_t+\nu_{it}\right)\left(\nu_{i'}+\nu_t+\nu_{i't}\right)\left(\varepsilon_i+\varepsilon_t+\varepsilon_{it}\right)\left(\varepsilon_{i'}+\varepsilon_t+\varepsilon_{i't}\right) \\
& -\left(\hat{\beta}-\beta\right)\left(\nu_i+\nu_t+\nu_{it}\right)\left(\nu_{i'}+\nu_t+\nu_{i't}\right)^2\left(\varepsilon_i+\varepsilon_t+\varepsilon_{it}\right) \\
& -\left(\hat{\beta}-\beta\right)\left(\nu_i+\nu_t+\nu_{it}\right)^2\left(\nu_{i'}+\nu_t+\nu_{i't}\right)\left(\varepsilon_{i'}+\varepsilon_t+\varepsilon_{i't}\right) \\
& +\left(\hat{\beta}-\beta\right)^2\left(\nu_i+\nu_t+\nu_{it}\right)^2\left(\nu_{i'}+\nu_t+\nu_{i't}\right)^2,
\end{aligned}$$

so

$$\begin{aligned}
C_2 \xrightarrow{p}& \mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] + \mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right] \\
& + \mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] + \mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right] \\
& + \mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] + \mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right]. \\
C_3 \xrightarrow{p}& \mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] \\
C_4 \xrightarrow{p}& \mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right].
\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\sqrt{NT}\left(\hat{\beta}-\beta\right)}{\sqrt{\widehat{\text{Avar}}\left[\hat{\beta}\right]}} \overset{a}{\sim} & \mathcal{N}\left(0, \frac{C_1}{C_5}\right) + \mathcal{N}\left(0, \frac{\mathrm{E}\left[\nu_{i=1}^2\right]}{\sqrt{C_5}}\right)\mathcal{N}\left(0, \frac{\mathrm{E}\left[\varepsilon_{t=1}^2\right]}{\sqrt{C_5}}\right) \\ & + \mathcal{N}\left(0, \frac{\mathrm{E}\left[\nu_{t=1}^2\right]}{\sqrt{C_5}}\right)\mathcal{N}\left(0, \frac{\mathrm{E}\left[\varepsilon_{i=1}^2\right]}{\sqrt{C_5}}\right), \\ \text{where } C_5 = & T\mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] + \mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\nu_{i=1}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right] \\ & + \mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] + N\mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\nu_{t=1}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right] \\ & + \mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{i=1}^2\right] + \mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{t=1}^2\right] + \mathrm{E}\left[\nu_{it=11}^2\right]\mathrm{E}\left[\varepsilon_{it=11}^2\right],\end{aligned}$$

and hence

$$\frac{\sqrt{NT}\left(\hat{\beta}-\beta\right)}{\sqrt{\widehat{\text{Avar}}\left[\hat{\beta}\right]}} \xrightarrow{d} \mathcal{N}(0,1),$$

as C_1/C_5 converges to unity and C_5 diverges to infinity as (N,T) diverges to (∞,∞). $\square$

Table 3.1. Assumptions on Regression Errors and Corresponding Variance Estimators

This table introduces several variance estimators for regression coefficients and their underlying assumptions. The first column classifies different estimators. OLS, HC and NCR are ordinary least squares, heteroskedasticity-consistent and cluster-robust (by firm), respectively. Here I exclude TCR (cluster-robust by time) since it is similar to NCR. The second column displays their background assumptions on the expected value of the outer product of $\mathbf{u}$. $\mathbf{I}_N$ and $\mathbf{1}_N$ are an N-dimensional identity matrix and an N-dimensional matrix of ones, respectively. $\mathbf{\Sigma}^N$ and $\mathbf{\Sigma}^T$ are the covariance matrices for firm and time effects, respectively. $\otimes$ is Kronecker product. The third column exhibits the corresponding estimators. $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is the regression residual and $\tilde{\mathbf{u}}$ is its adjusted counterpart, respectively. $\mathbf{1}^N$, $\mathbf{1}^T$ and $\mathbf{1}^{N\cup T} = \mathbf{1}^N + \mathbf{1}^T + \mathbf{1}^{N\cap T}$ are the indicator matrices for firm effects, time effects and both firm and time effects, respectively. $\circ$ is the Hadamard product. The fourth column shows the way to adjust the regression residual in detail. $\mathbf{M} = \mathbf{I}_{NT} - \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$ is the annihilator matrix of $\mathbf{X}$. $\mathbf{V}$ and $\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \cdots, \lambda_K\right)$ are the matrix of eigenvectors and eigenvalues, respectively. $\mathbf{\Lambda}^+$ is $\mathrm{diag}\left(\max\left(\lambda_1, 0\right), \cdots, \max\left(\lambda_K, 0\right)\right)$.

Estimator	$\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$	$\widehat{\mathrm{Var}}\left[\hat{\boldsymbol{\beta}}\right]$	$\tilde{\mathbf{u}}$
OLS	$\sigma^2\mathbf{I}_{NT}$	$s^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$	$s^2 = \hat{\mathbf{u}}^\top\hat{\mathbf{u}}/(NT-K)$
HC0 HC1 HC2 HC3	$\mathrm{diag}\left(\sigma_{11}^2, \cdots, \sigma_{NT}^2\right)$	$\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathrm{diag}\left(\hat{u}_{11}^2, \cdots, \hat{u}_{NT}^2\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$ $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathrm{diag}\left(\tilde{u}_{11}^2, \cdots, \tilde{u}_{NT}^2\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$	$\sqrt{NT/(NT-K)}\hat{\mathbf{u}}$ $\left(\mathbf{M}\circ\mathbf{I}_{NT}\right)^{-1/2}\hat{\mathbf{u}}$ $\left(\mathbf{M}\circ\mathbf{I}_{NT}\right)^{-1}\hat{\mathbf{u}}$
NCR0 NCR1 NCR2 NCR3	$\mathbf{I}_N\otimes\mathbf{\Sigma}^N$	$\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^\top\circ\mathbf{1}^N\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$ $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\tilde{\mathbf{u}}\tilde{\mathbf{u}}^\top\circ\mathbf{1}^N\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$	$\sqrt{N(NT-1)/((N-1)(NT-K))}\hat{\mathbf{u}}$ $\left(\mathbf{M}\circ\mathbf{1}^N\right)^{-1/2}\hat{\mathbf{u}}$ $\sqrt{N/(N-1)}\left(\mathbf{M}\circ\mathbf{1}^N\right)^{-1}\hat{\mathbf{u}}$
NTCR0 NTCR1 NTCR2 NTCR3 NTCR4 NTCR5	$\mathbf{I}_N\otimes\mathbf{\Sigma}^N + \left(\mathbf{1}_N - \mathbf{I}_N\right)\otimes\underbrace{\mathbf{\Sigma}^T}_{\text{diagonal}}$	$\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\hat{\mathbf{u}}\hat{\mathbf{u}}^\top\circ\mathbf{1}^{N\cup T}\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$ $\widehat{\mathrm{Var}}_{\mathrm{NCR}}\left[\hat{\boldsymbol{\beta}}\right] + \widehat{\mathrm{Var}}_{\mathrm{TCR}}\left[\hat{\boldsymbol{\beta}}\right] - \widehat{\mathrm{Var}}_{\mathrm{HC}}\left[\hat{\boldsymbol{\beta}}\right]$ $\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\left(\tilde{\mathbf{u}}\tilde{\mathbf{u}}^\top\circ\mathbf{1}^{N\cup T}\right)\mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$	$\widehat{\mathrm{Var}}^+ = \mathbf{V}\mathbf{\Lambda}^+\mathbf{V}^{-1}$ unless positive semi-definite $\left(\mathbf{M}\circ\mathbf{1}^{N\cup T}\right)^{-1/2}\hat{\mathbf{u}}$ $\sqrt{NT/(N-1)(T-1)}\left(\mathbf{M}\circ\mathbf{1}^{N\cup T}\right)^{-1}\hat{\mathbf{u}}$

Table 3.2. Result of Monte Carlo Simulation with $S = 5,000$, $N = 1,000$ and $T = 5$

This table displays the performance of different standard error estimators computed using simulated panel data sets. I simulate a panel of $N = 1,000$ and $T = 5$ with a regression model $y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it}$ 5,000 times. $\mathbf{x}_{it}$ includes 4 different regressors with a constant. $x_{it1} = \nu_{it1}$ has neither firm effects nor time effects. $x_{it2} = \sqrt{1/2}\nu_{i2} + \sqrt{1/2}\nu_{it3}$ only has firm effects. $x_{it3} = \sqrt{1/2}\nu_{t4} + \sqrt{1/2}\nu_{it5}$ only has time effects. $x_{it4} = \sqrt{1/3}\nu_{i6} + \sqrt{1/3}\nu_{t7} + \sqrt{1/3}\nu_{it8}$ has both firm and time effects. The error $u_{it} = \sqrt{1/3}\varepsilon_{i1} + \sqrt{1/3}\varepsilon_{t2} + \sqrt{1/3}\varepsilon_{it3}$ also has both firm and time effects. All 8 νs and 3 εs are i.i.d. standard normal and hence both $\mathbf{x}$ and u have zero mean and unit variance. I report (i) the averages and standard deviations of the simulated coefficient estimates, (ii) the averages of the simulated standard error estimates and (iii) the percentages of the simulated t-statistics computed with these standard error estimates that reject the null hypothesis $\beta = 1$ at 10%, 5% and 1% significance levels, respectively. I enclose all true counterparts as well.

Panel A.	β_0	β_1	β_2	β_3	β_4
True regression coefficient	1.0000	1.0000	1.0000	1.0000	1.0000
Estimate	0.9958	0.9999	1.0001	0.9968	1.0014
Asymptotic standard error	0.2590	0.0141	0.0183	0.1830	0.1500
Estimate	0.2627	0.0135	0.0172	0.1636	0.1279
OLS	0.0146	0.0133	0.0133	0.0146	0.0142
HC0	0.0145	0.0133	0.0133	0.0144	0.0140
HC1	0.0145	0.0133	0.0133	0.0144	0.0140
HC2	0.0145	0.0133	0.0133	0.0144	0.0140
HC3	0.0145	0.0133	0.0133	0.0144	0.0140
NCR0	0.0211	0.0133	0.0171	0.0130	0.0166
NCR1	0.0211	0.0133	0.0171	0.0130	0.0166
NCR2	0.0211	0.0133	0.0171	0.0131	0.0166
NCR3	0.0212	0.0133	0.0171	0.0131	0.0167
TCR0	0.1895	0.0106	0.0092	0.0962	0.0732
TCR1	0.2120	0.0119	0.0102	0.1076	0.0819
TCR2	0.2572	0.0137	0.0117	0.1477	0.1115
TCR3	0.3182	0.0160	0.0137	0.2055	0.1531
NTCR0	0.1903	0.0107	0.0143	0.0960	0.0741
NTCR1	0.2127	0.0119	0.0150	0.1074	0.0827
NTCR2	0.2307	0.0123	0.0153	0.1319	0.1004
NTCR3	0.3186	0.0161	0.0177	0.2054	0.1536
NTCR4	0.2305	0.0123	0.0149	0.1318	0.1002
NTCR5	0.3186	0.0161	0.0177	0.2052	0.1535
True P-value	0.1000	0.1000	0.1000	0.1000	0.1000
OLS	0.9260	0.1042	0.2030	0.8790	0.8420
HC0	0.9258	0.1032	0.2056	0.8808	0.8438
HC1	0.9258	0.1032	0.2056	0.8808	0.8434
HC2	0.9258	0.1032	0.2056	0.8808	0.8434
HC3	0.9258	0.1032	0.2048	0.8806	0.8434
NCR0	0.8886	0.1036	0.1030	0.8902	0.8122
NCR1	0.8886	0.1034	0.1030	0.8902	0.8120
NCR2	0.8886	0.1034	0.1028	0.8900	0.8118
NCR3	0.8886	0.1032	0.1024	0.8900	0.8116
TCR0	0.2880	0.2312	0.3982	0.3678	0.3794
TCR1	0.2460	0.1880	0.3522	0.3206	0.3196
TCR2	0.1754	0.1376	0.2940	0.1810	0.1694
TCR3	0.1142	0.0944	0.2320	0.0770	0.0640
NTCR0	0.2854	0.2302	0.1790	0.3700	0.3676
NTCR1	0.2434	0.1894	0.1616	0.3214	0.3106
NTCR2	0.2102	0.1798	0.1548	0.2302	0.2144
NTCR3	0.1132	0.0944	0.1128	0.0772	0.0620
NTCR4	0.2106	0.1794	0.1646	0.2308	0.2148
NTCR5	0.1130	0.0932	0.1130	0.0772	0.0622

Panel B.	β_0	β_1	β_2	β_3	β_4
True P-value	0.0500	0.0500	0.0500	0.0500	0.0500
OLS	0.9106	0.0572	0.1308	0.8608	0.8136
HC0	0.9110	0.0572	0.1320	0.8622	0.8152
HC1	0.9110	0.0570	0.1316	0.8622	0.8152
HC2	0.9110	0.0570	0.1314	0.8620	0.8152
HC3	0.9110	0.0570	0.1310	0.8620	0.8150
NCR0	0.8708	0.0560	0.0516	0.8722	0.7780
NCR1	0.8706	0.0556	0.0516	0.8718	0.7778
NCR2	0.8706	0.0556	0.0512	0.8716	0.7776
NCR3	0.8702	0.0556	0.0512	0.8714	0.7774
TCR0	0.2200	0.1678	0.3266	0.2932	0.2878
TCR1	0.1800	0.1302	0.2808	0.2406	0.2344
TCR2	0.1250	0.0916	0.2190	0.1168	0.1064
TCR3	0.0746	0.0606	0.1712	0.0434	0.0360
NTCR0	0.2166	0.1666	0.1166	0.2942	0.2778
NTCR1	0.1778	0.1276	0.1026	0.2420	0.2238
NTCR2	0.1536	0.1174	0.0978	0.1558	0.1356
NTCR3	0.0734	0.0590	0.0672	0.0440	0.0342
NTCR4	0.1540	0.1178	0.1066	0.1560	0.1362
NTCR5	0.0734	0.0594	0.0680	0.0440	0.0342
True P-value	0.0100	0.0100	0.0100	0.0100	0.0100
OLS	0.8832	0.0116	0.0482	0.8168	0.7558
HC0	0.8836	0.0116	0.0480	0.8190	0.7564
HC1	0.8836	0.0116	0.0476	0.8190	0.7564
HC2	0.8836	0.0116	0.0474	0.8190	0.7564
HC3	0.8836	0.0114	0.0474	0.8190	0.7564
NCR0	0.8378	0.0116	0.0110	0.8328	0.7126
NCR1	0.8378	0.0114	0.0110	0.8328	0.7122
NCR2	0.8378	0.0114	0.0110	0.8326	0.7118
NCR3	0.8378	0.0114	0.0110	0.8326	0.7110
TCR0	0.1340	0.0908	0.2124	0.1740	0.1602
TCR1	0.1024	0.0692	0.1768	0.1344	0.1216
TCR2	0.0618	0.0452	0.1354	0.0526	0.0446
TCR3	0.0368	0.0282	0.0994	0.0182	0.0114
NTCR0	0.1314	0.0874	0.0424	0.1770	0.1484
NTCR1	0.1008	0.0678	0.0348	0.1354	0.1134
NTCR2	0.0808	0.0618	0.0322	0.0776	0.0576
NTCR3	0.0356	0.0284	0.0182	0.0184	0.0098
NTCR4	0.0808	0.0612	0.0378	0.0776	0.0578
NTCR5	0.0356	0.0282	0.0208	0.0188	0.0098

Table 3.3. Result of Monte Carlo Simulation with $S = 5,000$, $N = 500$ and $T = 10$

This table displays the performance of different standard error estimators computed using simulated panel data sets. I simulate a panel of $N = 500$ and $T = 10$ with a regression model $y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it}$ 5,000 times. $\mathbf{x}_{it}$ includes 4 different regressors with a constant. $x_{it1} = \nu_{it1}$ has neither firm effects nor time effects. $x_{it2} = \sqrt{1/2}\nu_{i2} + \sqrt{1/2}\nu_{it3}$ only has firm effects. $x_{it3} = \sqrt{1/2}\nu_{t4} + \sqrt{1/2}\nu_{it5}$ only has time effects. $x_{it4} = \sqrt{1/3}\nu_{i6} + \sqrt{1/3}\nu_{t7} + \sqrt{1/3}\nu_{it8}$ has both firm and time effects. The error $u_{it} = \sqrt{1/3}\varepsilon_{i1} + \sqrt{1/3}\varepsilon_{t2} + \sqrt{1/3}\varepsilon_{it3}$ also has both firm and time effects. All 8 νs and 3 εs are i.i.d. standard normal and hence both $\mathbf{x}$ and u have zero mean and unit variance. I report (i) the averages and standard deviations of the simulated coefficient estimates, (ii) the averages of the simulated standard error estimates and (iii) the percentages of the simulated t-statistics computed with these standard error estimates that reject the null hypothesis $\beta = 1$ at 10%, 5% and 1% significance levels, respectively. I enclose all true counterparts as well.

Panel A.	β_0	β_1	β_2	β_3	β_4
True regression coefficient	1.0000	1.0000	1.0000	1.0000	1.0000
Estimate	0.9987	1.0003	1.0005	0.9985	0.9982
Asymptotic standard error	0.1846	0.0141	0.0224	0.1297	0.1072
Estimate	0.1863	0.0138	0.0212	0.1232	0.1020
OLS	0.0143	0.0137	0.0137	0.0144	0.0141
HC0	0.0143	0.0137	0.0137	0.0142	0.0140
HC1	0.0143	0.0137	0.0137	0.0142	0.0140
HC2	0.0143	0.0137	0.0137	0.0142	0.0140
HC3	0.0143	0.0137	0.0137	0.0142	0.0140
NCR0	0.0275	0.0136	0.0215	0.0124	0.0194
NCR1	0.0276	0.0137	0.0215	0.0124	0.0194
NCR2	0.0276	0.0137	0.0216	0.0124	0.0195
NCR3	0.0277	0.0137	0.0216	0.0124	0.0195
TCR0	0.1583	0.0123	0.0104	0.0942	0.0755
TCR1	0.1669	0.0129	0.0110	0.0994	0.0797
TCR2	0.1834	0.0139	0.0118	0.1163	0.0927
TCR3	0.2024	0.0151	0.0127	0.1367	0.1081
NTCR0	0.1602	0.0122	0.0196	0.0940	0.0770
NTCR1	0.1687	0.0129	0.0200	0.0991	0.0811
NTCR2	0.1757	0.0132	0.0202	0.1101	0.0892
NTCR3	0.2039	0.0151	0.0212	0.1366	0.1092
NTCR4	0.1756	0.0132	0.0200	0.1100	0.0891
NTCR5	0.2041	0.0151	0.0216	0.1365	0.1092
True P-value	0.1000	0.1000	0.1000	0.1000	0.1000
OLS	0.9052	0.1028	0.2910	0.8456	0.8110
HC0	0.9054	0.1054	0.2922	0.8470	0.8126
HC1	0.9052	0.1052	0.2918	0.8470	0.8124
HC2	0.9052	0.1050	0.2918	0.8470	0.8122
HC3	0.9052	0.1048	0.2914	0.8468	0.8122
NCR0	0.8172	0.1028	0.0938	0.8678	0.7538
NCR1	0.8166	0.1026	0.0936	0.8678	0.7536
NCR2	0.8158	0.1022	0.0924	0.8674	0.7532
NCR3	0.8156	0.1016	0.0914	0.8672	0.7528
TCR0	0.1838	0.1698	0.4312	0.2422	0.2454
TCR1	0.1654	0.1518	0.4048	0.2188	0.2214
TCR2	0.1310	0.1230	0.3710	0.1558	0.1564
TCR3	0.0978	0.0966	0.3380	0.1038	0.1016
NTCR0	0.1790	0.1708	0.1314	0.2446	0.2346
NTCR1	0.1612	0.1530	0.1232	0.2196	0.2094
NTCR2	0.1444	0.1428	0.1212	0.1802	0.1672
NTCR3	0.0946	0.0968	0.1076	0.1044	0.0952
NTCR4	0.1444	0.1426	0.1238	0.1806	0.1676
NTCR5	0.0946	0.0966	0.1012	0.1044	0.0952

Panel B.	β_0	β_1	β_2	β_3	β_4
True P-value	0.0500	0.0500	0.0500	0.0500	0.0500
OLS	0.8846	0.0492	0.2078	0.8166	0.7822
HC0	0.8852	0.0494	0.2064	0.8188	0.7836
HC1	0.8850	0.0488	0.2064	0.8188	0.7836
HC2	0.8850	0.0486	0.2062	0.8188	0.7836
HC3	0.8850	0.0486	0.2058	0.8188	0.7836
NCR0	0.7822	0.0506	0.0454	0.8418	0.7080
NCR1	0.7822	0.0504	0.0452	0.8414	0.7080
NCR2	0.7820	0.0502	0.0446	0.8412	0.7072
NCR3	0.7812	0.0494	0.0444	0.8412	0.7064
TCR0	0.1186	0.1068	0.3460	0.1680	0.1652
TCR1	0.1016	0.0892	0.3248	0.1472	0.1460
TCR2	0.0752	0.0730	0.2924	0.0966	0.0944
TCR3	0.0536	0.0548	0.2588	0.0584	0.0556
NTCR0	0.1140	0.1070	0.0734	0.1694	0.1548
NTCR1	0.0970	0.0900	0.0684	0.1490	0.1346
NTCR2	0.0850	0.0842	0.0664	0.1136	0.1026
NTCR3	0.0522	0.0548	0.0522	0.0586	0.0508
NTCR4	0.0850	0.0848	0.0684	0.1136	0.1030
NTCR5	0.0518	0.0554	0.0492	0.0586	0.0504
True P-value	0.0100	0.0100	0.0100	0.0100	0.0100
OLS	0.8556	0.0086	0.0960	0.7600	0.7206
HC0	0.8562	0.0086	0.0978	0.7628	0.7232
HC1	0.8560	0.0086	0.0978	0.7626	0.7228
HC2	0.8560	0.0086	0.0978	0.7624	0.7226
HC3	0.8556	0.0086	0.0974	0.7622	0.7224
NCR0	0.7108	0.0088	0.0098	0.7890	0.6198
NCR1	0.7096	0.0088	0.0098	0.7884	0.6192
NCR2	0.7092	0.0088	0.0098	0.7882	0.6170
NCR3	0.7076	0.0086	0.0094	0.7882	0.6160
TCR0	0.0486	0.0422	0.2246	0.0758	0.0746
TCR1	0.0400	0.0344	0.1996	0.0634	0.0602
TCR2	0.0278	0.0246	0.1726	0.0388	0.0310
TCR3	0.0180	0.0182	0.1484	0.0198	0.0152
NTCR0	0.0444	0.0426	0.0190	0.0770	0.0660
NTCR1	0.0370	0.0348	0.0172	0.0648	0.0528
NTCR2	0.0324	0.0314	0.0166	0.0476	0.0334
NTCR3	0.0166	0.0186	0.0132	0.0200	0.0140
NTCR4	0.0324	0.0316	0.0172	0.0480	0.0334
NTCR5	0.0162	0.0192	0.0114	0.0200	0.0140

Table 3.4. Result of Monte Carlo Simulation with $S = 5,000$, $N = 250$ and $T = 20$

This table displays the performance of different standard error estimators computed using simulated panel data sets. I simulate a panel of $N = 250$ and $T = 20$ with a regression model $y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it}$ 5,000 times. $\mathbf{x}_{it}$ includes 4 different regressors with a constant. $x_{it1} = \nu_{it1}$ has neither firm effects nor time effects. $x_{it2} = \sqrt{1/2}\nu_{i2} + \sqrt{1/2}\nu_{it3}$ only has firm effects. $x_{it3} = \sqrt{1/2}\nu_{t4} + \sqrt{1/2}\nu_{it5}$ only has time effects. $x_{it4} = \sqrt{1/3}\nu_{i6}+\sqrt{1/3}\nu_{t7}+\sqrt{1/3}\nu_{it8}$ has both firm and time effects. The error $u_{it} = \sqrt{1/3}\varepsilon_{i1}+\sqrt{1/3}\varepsilon_{t2}+\sqrt{1/3}\varepsilon_{it3}$ also has both firm and time effects. All 8 νs and 3 εs are i.i.d. standard normal and hence both $\mathbf{x}$ and u have zero mean and unit variance. I report (i) the averages and standard deviations of the simulated coefficient estimates, (ii) the averages of the simulated standard error estimates and (iii) the percentages of the simulated t-statistics computed with these standard error estimates that reject the null hypothesis $\beta = 1$ at 10%, 5% and 1% significance levels, respectively. I enclose all true counterparts as well.

Panel A.	β_0	β_1	β_2	β_3	β_4
True regression coefficient	1.0000	1.0000	1.0000	1.0000	1.0000
Estimate	0.9986	1.0000	1.0002	0.9994	0.9996
Asymptotic standard error	0.1344	0.0141	0.0289	0.0922	0.0785
Estimate	0.1370	0.0140	0.0279	0.0904	0.0740
OLS	0.0142	0.0139	0.0139	0.0142	0.0141
HC0	0.0142	0.0139	0.0139	0.0141	0.0140
HC1	0.0142	0.0139	0.0139	0.0141	0.0141
HC2	0.0142	0.0139	0.0139	0.0141	0.0141
HC3	0.0142	0.0139	0.0139	0.0141	0.0141
NCR0	0.0375	0.0138	0.0280	0.0120	0.0242
NCR1	0.0376	0.0139	0.0281	0.0120	0.0242
NCR2	0.0378	0.0139	0.0282	0.0120	0.0244
NCR3	0.0379	0.0140	0.0284	0.0121	0.0245
TCR0	0.1210	0.0131	0.0111	0.0793	0.0635
TCR1	0.1242	0.0135	0.0113	0.0814	0.0652
TCR2	0.1301	0.0140	0.0118	0.0881	0.0704
TCR3	0.1366	0.0146	0.0122	0.0956	0.0760
NTCR0	0.1261	0.0131	0.0267	0.0789	0.0668
NTCR1	0.1291	0.0134	0.0269	0.0810	0.0684
NTCR2	0.1317	0.0136	0.0271	0.0856	0.0717
NTCR3	0.1411	0.0146	0.0277	0.0953	0.0789
NTCR4	0.1316	0.0136	0.0270	0.0855	0.0716
NTCR5	0.1415	0.0146	0.0281	0.0954	0.0791
True P-value	0.1000	0.1000	0.1000	0.1000	0.1000
OLS	0.8648	0.1010	0.4126	0.8010	0.7562
HC0	0.8646	0.1024	0.4130	0.8012	0.7588
HC1	0.8646	0.1024	0.4128	0.8012	0.7588
HC2	0.8646	0.1022	0.4128	0.8012	0.7588
HC3	0.8646	0.1022	0.4124	0.8010	0.7586
NCR0	0.6672	0.1022	0.1022	0.8290	0.5992
NCR1	0.6656	0.1018	0.1014	0.8288	0.5980
NCR2	0.6650	0.1008	0.1004	0.8284	0.5960
NCR3	0.6644	0.0992	0.0980	0.8280	0.5946
TCR0	0.1544	0.1274	0.5158	0.1644	0.1832
TCR1	0.1458	0.1178	0.5048	0.1548	0.1732
TCR2	0.1262	0.1068	0.4898	0.1256	0.1440
TCR3	0.1124	0.0946	0.4720	0.0994	0.1124
NTCR0	0.1366	0.1300	0.1214	0.1668	0.1598
NTCR1	0.1276	0.1210	0.1176	0.1564	0.1478
NTCR2	0.1214	0.1158	0.1154	0.1378	0.1314
NTCR3	0.0994	0.0940	0.1068	0.1000	0.0958
NTCR4	0.1216	0.1156	0.1166	0.1384	0.1320
NTCR5	0.0990	0.0952	0.1012	0.1000	0.0952

Panel B.	β_0	β_1	β_2	β_3	β_4
True P-value	0.0500	0.0500	0.0500	0.0500	0.0500
OLS	0.8418	0.0508	0.3304	0.7686	0.7110
HC0	0.8428	0.0506	0.3300	0.7698	0.7128
HC1	0.8426	0.0506	0.3298	0.7698	0.7126
HC2	0.8426	0.0506	0.3298	0.7698	0.7126
HC3	0.8426	0.0506	0.3298	0.7698	0.7124
NCR0	0.6082	0.0534	0.0492	0.8000	0.5260
NCR1	0.6076	0.0534	0.0490	0.7996	0.5254
NCR2	0.6066	0.0520	0.0474	0.7992	0.5234
NCR3	0.6050	0.0512	0.0462	0.7986	0.5208
TCR0	0.0920	0.0748	0.4406	0.1014	0.1138
TCR1	0.0840	0.0696	0.4276	0.0930	0.1036
TCR2	0.0734	0.0588	0.4102	0.0728	0.0802
TCR3	0.0624	0.0482	0.3922	0.0548	0.0568
NTCR0	0.0784	0.0774	0.0644	0.1034	0.0908
NTCR1	0.0726	0.0706	0.0622	0.0946	0.0816
NTCR2	0.0668	0.0662	0.0604	0.0794	0.0690
NTCR3	0.0536	0.0482	0.0552	0.0554	0.0456
NTCR4	0.0670	0.0666	0.0608	0.0796	0.0692
NTCR5	0.0530	0.0498	0.0508	0.0554	0.0448
True P-value	0.0100	0.0100	0.0100	0.0100	0.0100
OLS	0.7972	0.0112	0.2054	0.6884	0.6344
HC0	0.7968	0.0112	0.2056	0.6908	0.6348
HC1	0.7966	0.0110	0.2054	0.6908	0.6348
HC2	0.7966	0.0110	0.2054	0.6908	0.6348
HC3	0.7966	0.0106	0.2048	0.6906	0.6346
NCR0	0.4932	0.0126	0.0106	0.7402	0.4142
NCR1	0.4926	0.0126	0.0106	0.7392	0.4124
NCR2	0.4900	0.0126	0.0102	0.7380	0.4090
NCR3	0.4876	0.0122	0.0094	0.7370	0.4078
TCR0	0.0320	0.0224	0.3190	0.0378	0.0366
TCR1	0.0292	0.0204	0.3078	0.0332	0.0324
TCR2	0.0228	0.0168	0.2900	0.0240	0.0214
TCR3	0.0188	0.0138	0.2714	0.0166	0.0152
NTCR0	0.0262	0.0254	0.0140	0.0388	0.0244
NTCR1	0.0230	0.0218	0.0134	0.0344	0.0214
NTCR2	0.0206	0.0198	0.0128	0.0276	0.0168
NTCR3	0.0144	0.0140	0.0120	0.0174	0.0106
NTCR4	0.0206	0.0198	0.0130	0.0276	0.0168
NTCR5	0.0140	0.0146	0.0098	0.0170	0.0106

Table 3.5. Result of Monte Carlo Simulation with $S = 5,000$, $N = 100$ and $T = 50$

This table displays the performance of different standard error estimators computed using simulated panel data sets. I simulate a panel of $N = 100$ and $T = 50$ with a regression model $y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + u_{it}$ 5,000 times. $\mathbf{x}_{it}$ includes 4 different regressors with a constant. $x_{it1} = \nu_{it1}$ has neither firm effects nor time effects. $x_{it2} = \sqrt{1/2}\nu_{i2} + \sqrt{1/2}\nu_{it3}$ only has firm effects. $x_{it3} = \sqrt{1/2}\nu_{t4} + \sqrt{1/2}\nu_{it5}$ only has time effects. $x_{it4} = \sqrt{1/3}\nu_{i6} + \sqrt{1/3}\nu_{t7} + \sqrt{1/3}\nu_{it8}$ has both firm and time effects. The error $u_{it} = \sqrt{1/3}\varepsilon_{i1} + \sqrt{1/3}\varepsilon_{t2} + \sqrt{1/3}\varepsilon_{it3}$ also has both firm and time effects. All 8 νs and 3 εs are i.i.d. standard normal and hence both $\mathbf{x}$ and u have zero mean and unit variance. I report (i) the averages and standard deviations of the simulated coefficient estimates, (ii) the averages of the simulated standard error estimates and (iii) the percentages of the simulated t-statistics computed with these standard error estimates that reject the null hypothesis $\beta = 1$ at 10%, 5% and 1% significance levels, respectively. I enclose all true counterparts as well.

Panel A.	β_0	β_1	β_2	β_3	β_4
True regression coefficient	1.0000	1.0000	1.0000	1.0000	1.0000
Estimate	1.0006	1.0000	1.0003	1.0011	1.0017
Asymptotic standard error	0.1003	0.0141	0.0428	0.0592	0.0591
Estimate	0.1025	0.0142	0.0424	0.0588	0.0590
OLS	0.0142	0.0140	0.0141	0.0142	0.0141
HC0	0.0142	0.0140	0.0140	0.0141	0.0141
HC1	0.0142	0.0140	0.0141	0.0141	0.0141
HC2	0.0142	0.0140	0.0141	0.0141	0.0141
HC3	0.0142	0.0140	0.0141	0.0141	0.0141
NCR0	0.0577	0.0138	0.0411	0.0117	0.0343
NCR1	0.0580	0.0139	0.0414	0.0118	0.0345
NCR2	0.0585	0.0140	0.0420	0.0119	0.0350
NCR3	0.0591	0.0141	0.0427	0.0120	0.0356
TCR0	0.0802	0.0137	0.0115	0.0555	0.0458
TCR1	0.0811	0.0138	0.0116	0.0561	0.0462
TCR2	0.0826	0.0141	0.0118	0.0580	0.0477
TCR3	0.0842	0.0143	0.0120	0.0599	0.0492
NTCR0	0.0980	0.0135	0.0404	0.0550	0.0558
NTCR1	0.0988	0.0137	0.0406	0.0556	0.0563
NTCR2	0.0995	0.0138	0.0411	0.0569	0.0573
NTCR3	0.1020	0.0144	0.0421	0.0594	0.0594
NTCR4	0.0995	0.0138	0.0411	0.0568	0.0572
NTCR5	0.1026	0.0144	0.0424	0.0596	0.0597
True P-value	0.1000	0.1000	0.1000	0.1000	0.1000
OLS	0.8218	0.1054	0.5856	0.6962	0.6944
HC0	0.8216	0.1072	0.5876	0.6968	0.6968
HC1	0.8216	0.1072	0.5876	0.6966	0.6968
HC2	0.8216	0.1072	0.5876	0.6966	0.6966
HC3	0.8212	0.1062	0.5876	0.6966	0.6962
NCR0	0.3564	0.1146	0.1142	0.7470	0.3462
NCR1	0.3536	0.1130	0.1126	0.7460	0.3430
NCR2	0.3492	0.1092	0.1084	0.7424	0.3352
NCR3	0.3434	0.1054	0.1038	0.7402	0.3292
TCR0	0.2006	0.1164	0.6658	0.1326	0.2104
TCR1	0.1962	0.1130	0.6628	0.1302	0.2044
TCR2	0.1882	0.1084	0.6576	0.1174	0.1918
TCR3	0.1808	0.1032	0.6534	0.1068	0.1772
NTCR0	0.1214	0.1288	0.1212	0.1376	0.1232
NTCR1	0.1202	0.1218	0.1188	0.1330	0.1200
NTCR2	0.1182	0.1180	0.1158	0.1246	0.1160
NTCR3	0.1096	0.1052	0.1078	0.1092	0.1040
NTCR4	0.1182	0.1180	0.1162	0.1252	0.1160
NTCR5	0.1074	0.1060	0.1060	0.1088	0.1026

Panel B.	β_0	β_1	β_2	β_3	β_4
True P-value	0.0500	0.0500	0.0500	0.0500	0.0500
OLS	0.7884	0.0568	0.5156	0.6398	0.6380
HC0	0.7888	0.0570	0.5172	0.6424	0.6426
HC1	0.7888	0.0568	0.5170	0.6422	0.6426
HC2	0.7888	0.0568	0.5166	0.6420	0.6426
HC3	0.7886	0.0566	0.5162	0.6418	0.6424
NCR0	0.2706	0.0628	0.0634	0.6996	0.2624
NCR1	0.2684	0.0610	0.0618	0.6980	0.2602
NCR2	0.2636	0.0588	0.0588	0.6960	0.2528
NCR3	0.2588	0.0572	0.0546	0.6928	0.2478
TCR0	0.1328	0.0658	0.6022	0.0756	0.1376
TCR1	0.1284	0.0638	0.5986	0.0740	0.1336
TCR2	0.1232	0.0596	0.5928	0.0652	0.1220
TCR3	0.1166	0.0554	0.5844	0.0554	0.1120
NTCR0	0.0678	0.0716	0.0684	0.0782	0.0720
NTCR1	0.0658	0.0664	0.0662	0.0758	0.0684
NTCR2	0.0630	0.0636	0.0634	0.0704	0.0644
NTCR3	0.0544	0.0540	0.0596	0.0566	0.0536
NTCR4	0.0638	0.0638	0.0640	0.0708	0.0644
NTCR5	0.0536	0.0546	0.0576	0.0562	0.0522
True P-value	0.0100	0.0100	0.0100	0.0100	0.0100
OLS	0.7160	0.0114	0.3918	0.5436	0.5376
HC0	0.7168	0.0118	0.3938	0.5440	0.5416
HC1	0.7164	0.0118	0.3938	0.5440	0.5414
HC2	0.7164	0.0118	0.3938	0.5440	0.5412
HC3	0.7162	0.0118	0.3932	0.5438	0.5410
NCR0	0.1510	0.0136	0.0164	0.6130	0.1414
NCR1	0.1490	0.0130	0.0158	0.6110	0.1382
NCR2	0.1464	0.0118	0.0140	0.6080	0.1334
NCR3	0.1406	0.0108	0.0128	0.6044	0.1266
TCR0	0.0486	0.0136	0.4884	0.0174	0.0560
TCR1	0.0464	0.0128	0.4848	0.0166	0.0528
TCR2	0.0434	0.0122	0.4776	0.0126	0.0458
TCR3	0.0386	0.0100	0.4718	0.0094	0.0392
NTCR0	0.0142	0.0186	0.0192	0.0186	0.0172
NTCR1	0.0132	0.0168	0.0182	0.0176	0.0170
NTCR2	0.0132	0.0162	0.0170	0.0152	0.0160
NTCR3	0.0104	0.0124	0.0134	0.0100	0.0120
NTCR4	0.0132	0.0164	0.0172	0.0154	0.0160
NTCR5	0.0100	0.0124	0.0132	0.0098	0.0114

Table 3.6. Result of Monte Carlo Simulations for β_4 with $S = 5,000$ by Combination

This table compares the performance of 2-way clustered standard error estimators using simulated panel data sets. I simulate 5,000 panels for each of all different combinations of $N \geq T$ from $\{10, 20, 30, 40, 50, 75, 100\}$. The remainder setting here is identical to one introduced in Tables 3.2–3.5. I focus on the result about β_4 and compare 6 different 2-way clustered standard error estimators NTCR0–NTCR5. I report (i) the averages and the standard deviations of the simulated β_4 estimates, (ii) the averages of the simulated standard error estimates and (iii) the percentages of t-statistics significant at 10%, 5% and 1% levels, respectively. I enclose all true counterparts as well.

Panel A. N	10	20	30	40	50	75	100	20	30	40	50	75	100	30
T				10							20			30
True regression coefficient	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Estimate	0.9957	0.9974	1.0006	0.9979	0.9999	1.0006	1.0020	1.0015	1.0006	1.0016	1.0023	0.9997	1.0007	0.9993
Asymptotic standard error	0.1732	0.1434	0.1319	0.1258	0.1220	0.1167	0.1140	0.1143	0.1027	0.0965	0.0925	0.0869	0.0840	0.0909
Estimate	0.1754	0.1404	0.1300	0.1202	0.1157	0.1106	0.1081	0.1152	0.1025	0.0948	0.0891	0.0853	0.0822	0.0894
NTCR0	0.1233	0.1084	0.1010	0.0957	0.0929	0.0875	0.0852	0.0959	0.0883	0.0834	0.0807	0.0756	0.0726	0.0812
NTCR1	0.1362	0.1157	0.1069	0.1008	0.0978	0.0919	0.0895	0.0995	0.0910	0.0857	0.0827	0.0774	0.0743	0.0830
NTCR2	0.1510	0.1257	0.1158	0.1092	0.1060	0.0999	0.0976	0.1051	0.0953	0.0894	0.0863	0.0808	0.0776	0.0860
NTCR3	0.1976	0.1537	0.1398	0.1315	0.1275	0.1204	0.1180	0.1187	0.1054	0.0982	0.0945	0.0885	0.0849	0.0929
NTCR4	0.1451	0.1230	0.1140	0.1080	0.1050	0.0993	0.0971	0.1039	0.0946	0.0889	0.0859	0.0805	0.0774	0.0856
NTCR5	0.1909	0.1516	0.1388	0.1309	0.1272	0.1203	0.1179	0.1186	0.1057	0.0986	0.0949	0.0888	0.0852	0.0934
True P-value	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
NTCR0	0.2692	0.2284	0.2206	0.2052	0.2122	0.2098	0.2074	0.1856	0.1652	0.1560	0.1480	0.1538	0.1606	0.1442
NTCR1	0.2244	0.1970	0.1954	0.1816	0.1906	0.1860	0.1866	0.1690	0.1544	0.1458	0.1368	0.1436	0.1504	0.1362
NTCR2	0.1800	0.1664	0.1624	0.1536	0.1596	0.1560	0.1566	0.1524	0.1378	0.1310	0.1208	0.1276	0.1322	0.1260
NTCR3	0.0896	0.0990	0.0974	0.0992	0.0946	0.1000	0.0978	0.1050	0.1050	0.0998	0.0914	0.0988	0.1020	0.0962
NTCR4	0.1992	0.1736	0.1682	0.1582	0.1634	0.1584	0.1590	0.1562	0.1410	0.1316	0.1230	0.1288	0.1328	0.1272
NTCR5	0.1002	0.1006	0.0992	0.1010	0.0956	0.0998	0.0976	0.1062	0.1040	0.0992	0.0896	0.0976	0.1006	0.0948
True P-value	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
NTCR0	0.1944	0.1560	0.1448	0.1368	0.1366	0.1386	0.1414	0.1184	0.1038	0.0944	0.0832	0.0906	0.0946	0.0866
NTCR1	0.1492	0.1312	0.1240	0.1196	0.1168	0.1218	0.1256	0.1056	0.0948	0.0878	0.0766	0.0810	0.0872	0.0792
NTCR2	0.1182	0.1042	0.1000	0.1006	0.0932	0.1006	0.0994	0.0872	0.0788	0.0770	0.0648	0.0708	0.0754	0.0694
NTCR3	0.0528	0.0562	0.0494	0.0520	0.0530	0.0554	0.0544	0.0588	0.0566	0.0542	0.0450	0.0524	0.0498	0.0512
NTCR4	0.1322	0.1102	0.1044	0.1040	0.0962	0.1020	0.1008	0.0914	0.0810	0.0786	0.0668	0.0714	0.0756	0.0710
NTCR5	0.0598	0.0598	0.0514	0.0524	0.0528	0.0550	0.0544	0.0592	0.0556	0.0536	0.0442	0.0518	0.0482	0.0512
True P-value	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100
NTCR0	0.0978	0.0706	0.0598	0.0600	0.0534	0.0612	0.0612	0.0456	0.0366	0.0338	0.0288	0.0312	0.0254	0.0252
NTCR1	0.0720	0.0550	0.0460	0.0476	0.0442	0.0498	0.0506	0.0378	0.0312	0.0294	0.0248	0.0286	0.0226	0.0228
NTCR2	0.0504	0.0390	0.0332	0.0344	0.0300	0.0368	0.0376	0.0276	0.0246	0.0252	0.0200	0.0236	0.0198	0.0190
NTCR3	0.0168	0.0152	0.0170	0.0148	0.0112	0.0180	0.0176	0.0162	0.0140	0.0174	0.0108	0.0146	0.0118	0.0130
NTCR4	0.0574	0.0446	0.0348	0.0352	0.0324	0.0380	0.0382	0.0302	0.0262	0.0254	0.0208	0.0242	0.0200	0.0194
NTCR5	0.0224	0.0176	0.0168	0.0150	0.0114	0.0174	0.0176	0.0162	0.0138	0.0172	0.0104	0.0142	0.0114	0.0126

Panel B. N	40	50	75	100	40	50	75	100	50	75	100	75	100	100
T		30				40				50		75		100
True regression coefficient	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Estimate	0.9988	1.0001	0.9995	0.9984	0.9993	1.0014	1.0004	0.9988	1.0010	1.0003	1.0017	0.9994	0.9992	0.9997
Asymptotic standard error	0.0844	0.0803	0.0744	0.0712	0.0777	0.0734	0.0672	0.0639	0.0690	0.0625	0.0591	0.0557	0.0519	0.0480
Estimate	0.0835	0.0780	0.0737	0.0709	0.0774	0.0726	0.0667	0.0630	0.0694	0.0611	0.0590	0.0544	0.0514	0.0478
NTCR0	0.0758	0.0726	0.0676	0.0651	0.0711	0.0675	0.0624	0.0595	0.0641	0.0589	0.0558	0.0531	0.0498	0.0463
NTCR1	0.0772	0.0738	0.0687	0.0661	0.0722	0.0684	0.0632	0.0602	0.0649	0.0595	0.0563	0.0535	0.0502	0.0465
NTCR2	0.0797	0.0761	0.0707	0.0680	0.0742	0.0701	0.0646	0.0615	0.0663	0.0606	0.0573	0.0543	0.0508	0.0470
NTCR3	0.0853	0.0811	0.0750	0.0722	0.0785	0.0738	0.0676	0.0643	0.0694	0.0630	0.0594	0.0559	0.0522	0.0481
NTCR4	0.0794	0.0758	0.0705	0.0679	0.0740	0.0699	0.0645	0.0614	0.0662	0.0605	0.0572	0.0543	0.0508	0.0470
NTCR5	0.0858	0.0815	0.0753	0.0725	0.0790	0.0742	0.0680	0.0646	0.0698	0.0633	0.0597	0.0562	0.0524	0.0483
True P-value	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000	0.1000
NTCR0	0.1422	0.1368	0.1438	0.1448	0.1444	0.1386	0.1272	0.1238	0.1352	0.1126	0.1232	0.1112	0.1200	0.1142
NTCR1	0.1348	0.1302	0.1372	0.1384	0.1396	0.1332	0.1236	0.1184	0.1308	0.1096	0.1200	0.1082	0.1170	0.1118
NTCR2	0.1258	0.1194	0.1262	0.1290	0.1288	0.1230	0.1140	0.1102	0.1218	0.1040	0.1160	0.1038	0.1120	0.1098
NTCR3	0.1034	0.0968	0.1080	0.1056	0.1060	0.1032	0.1002	0.0960	0.1084	0.0950	0.1040	0.0928	0.1034	0.1014
NTCR4	0.1270	0.1200	0.1264	0.1294	0.1302	0.1238	0.1146	0.1106	0.1226	0.1046	0.1160	0.1038	0.1124	0.1098
NTCR5	0.1006	0.0950	0.1048	0.1046	0.1044	0.1002	0.0984	0.0944	0.1068	0.0930	0.1026	0.0920	0.1010	0.1004
True P-value	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
NTCR0	0.0852	0.0788	0.0820	0.0814	0.0830	0.0786	0.0706	0.0666	0.0780	0.0594	0.0720	0.0592	0.0612	0.0594
NTCR1	0.0816	0.0760	0.0778	0.0758	0.0776	0.0722	0.0684	0.0644	0.0746	0.0572	0.0684	0.0574	0.0600	0.0576
NTCR2	0.0728	0.0680	0.0682	0.0676	0.0694	0.0676	0.0646	0.0580	0.0674	0.0542	0.0644	0.0546	0.0560	0.0558
NTCR3	0.0578	0.0522	0.0554	0.0532	0.0542	0.0528	0.0560	0.0486	0.0568	0.0460	0.0536	0.0488	0.0496	0.0502
NTCR4	0.0738	0.0684	0.0684	0.0680	0.0708	0.0684	0.0646	0.0582	0.0684	0.0546	0.0644	0.0550	0.0562	0.0558
NTCR5	0.0568	0.0512	0.0536	0.0526	0.0534	0.0518	0.0554	0.0478	0.0544	0.0454	0.0522	0.0472	0.0494	0.0500
True P-value	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100	0.0100
NTCR0	0.0276	0.0230	0.0246	0.0202	0.0246	0.0192	0.0228	0.0160	0.0208	0.0154	0.0172	0.0130	0.0156	0.0144
NTCR1	0.0256	0.0206	0.0232	0.0182	0.0230	0.0180	0.0220	0.0154	0.0188	0.0146	0.0170	0.0122	0.0152	0.0140
NTCR2	0.0214	0.0168	0.0210	0.0140	0.0198	0.0150	0.0188	0.0132	0.0168	0.0128	0.0160	0.0114	0.0138	0.0132
NTCR3	0.0148	0.0112	0.0148	0.0094	0.0144	0.0106	0.0138	0.0102	0.0134	0.0118	0.0120	0.0088	0.0110	0.0114
NTCR4	0.0222	0.0170	0.0212	0.0140	0.0200	0.0156	0.0190	0.0136	0.0168	0.0128	0.0160	0.0116	0.0138	0.0132
NTCR5	0.0140	0.0108	0.0144	0.0092	0.0142	0.0100	0.0132	0.0102	0.0126	0.0106	0.0114	0.0082	0.0106	0.0112

Figure 3.1. Expected Value of the Error Outer Product $\mathbf{u}\mathbf{u}^\top$ with Different Assumptions

This figure shows the matrices of $\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$, the expected value of the outer product of the regression error, with four assumptions. This visualization uses $N = 2$ and $T = 5$, respectively. First, the upper-left case assumes neither firm nor time effects, that is, both cross and serial correlations are zero and hence the matrix is diagonal with 10 nonzero elements. Second, the upper-right case assumes only firm effects, that is, cross correlations are zero and hence the matrix is block diagonal with 50 nonzero elements. Third, the lower-left case assumes only time effects, that is, serial correlations are zero and hence the matrix is diagonal block by block with 20 nonzero elements. Fourth, the lower-right case assumes both firm and time effects, that is, neither cross nor serial correlations are zero and hence the matrix has 60 nonzero elements.

$$\begin{matrix}
u_{11}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & u_{12}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & u_{13}^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & u_{14}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & u_{15}^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & u_{21}^2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & u_{22}^2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & u_{23}^2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & u_{24}^2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & u_{25}^2
\end{matrix}$$

Neither firm nor time effects ($NT = 10$)

$$\begin{matrix}
u_{11}^2 & u_{11}u_{12} & u_{11}u_{13} & u_{11}u_{14} & u_{11}u_{15} & 0 & 0 & 0 & 0 & 0 \\
u_{12}u_{11} & u_{12}^2 & u_{12}u_{13} & u_{12}u_{14} & u_{12}u_{15} & 0 & 0 & 0 & 0 & 0 \\
u_{13}u_{11} & u_{13}u_{12} & u_{13}^2 & u_{13}u_{14} & u_{13}u_{15} & 0 & 0 & 0 & 0 & 0 \\
u_{14}u_{11} & u_{14}u_{12} & u_{14}u_{13} & u_{14}^2 & u_{14}u_{15} & 0 & 0 & 0 & 0 & 0 \\
u_{15}u_{11} & u_{15}u_{12} & u_{15}u_{13} & u_{15}u_{14} & u_{15}^2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & u_{21}^2 & u_{21}u_{22} & u_{21}u_{23} & u_{21}u_{24} & u_{21}u_{25} \\
0 & 0 & 0 & 0 & 0 & u_{22}u_{21} & u_{22}^2 & u_{22}u_{23} & u_{22}u_{24} & u_{22}u_{25} \\
0 & 0 & 0 & 0 & 0 & u_{23}u_{21} & u_{23}u_{22} & u_{23}^2 & u_{23}u_{24} & u_{23}u_{25} \\
0 & 0 & 0 & 0 & 0 & u_{24}u_{21} & u_{24}u_{22} & u_{24}u_{23} & u_{24}^2 & u_{24}u_{25} \\
0 & 0 & 0 & 0 & 0 & u_{25}u_{21} & u_{25}u_{22} & u_{25}u_{23} & u_{25}u_{24} & u_{25}^2
\end{matrix}$$

Only firm effects ($NT^2 = 50$)

$$\begin{matrix}
u_{11}^2 & 0 & 0 & 0 & 0 & u_{11}u_{21} & 0 & 0 & 0 & 0 \\
0 & u_{12}^2 & 0 & 0 & 0 & 0 & u_{12}u_{22} & 0 & 0 & 0 \\
0 & 0 & u_{13}^2 & 0 & 0 & 0 & 0 & u_{13}u_{23} & 0 & 0 \\
0 & 0 & 0 & u_{14}^2 & 0 & 0 & 0 & 0 & u_{14}u_{24} & 0 \\
0 & 0 & 0 & 0 & u_{15}^2 & 0 & 0 & 0 & 0 & u_{15}u_{25} \\
u_{21}u_{11} & 0 & 0 & 0 & 0 & u_{21}^2 & 0 & 0 & 0 & 0 \\
0 & u_{22}u_{12} & 0 & 0 & 0 & 0 & u_{22}^2 & 0 & 0 & 0 \\
0 & 0 & u_{23}u_{13} & 0 & 0 & 0 & 0 & u_{23}^2 & 0 & 0 \\
0 & 0 & 0 & u_{24}u_{14} & 0 & 0 & 0 & 0 & u_{24}^2 & 0 \\
0 & 0 & 0 & 0 & u_{25}u_{15} & 0 & 0 & 0 & 0 & u_{25}^2
\end{matrix}$$

Only time effects ($N^2T = 20$)

$$\begin{matrix}
u_{11}^2 & u_{11}u_{12} & u_{11}u_{13} & u_{11}u_{14} & u_{11}u_{15} & u_{11}u_{21} & 0 & 0 & 0 & 0 \\
u_{12}u_{11} & u_{12}^2 & u_{12}u_{13} & u_{12}u_{14} & u_{12}u_{15} & 0 & u_{12}u_{22} & 0 & 0 & 0 \\
u_{13}u_{11} & u_{13}u_{12} & u_{13}^2 & u_{13}u_{14} & u_{13}u_{15} & 0 & 0 & u_{13}u_{23} & 0 & 0 \\
u_{14}u_{11} & u_{14}u_{12} & u_{14}u_{13} & u_{14}^2 & u_{14}u_{15} & 0 & 0 & 0 & u_{14}u_{24} & 0 \\
u_{15}u_{11} & u_{15}u_{12} & u_{15}u_{13} & u_{15}u_{14} & u_{15}^2 & 0 & 0 & 0 & 0 & u_{15}u_{25} \\
u_{21}u_{11} & 0 & 0 & 0 & 0 & u_{21}^2 & u_{21}u_{22} & u_{21}u_{23} & u_{21}u_{24} & u_{21}u_{25} \\
0 & u_{22}u_{12} & 0 & 0 & 0 & u_{22}u_{21} & u_{22}^2 & u_{22}u_{23} & u_{22}u_{24} & u_{22}u_{25} \\
0 & 0 & u_{23}u_{13} & 0 & 0 & u_{23}u_{21} & u_{23}u_{22} & u_{23}^2 & u_{23}u_{24} & u_{23}u_{25} \\
0 & 0 & 0 & u_{24}u_{14} & 0 & u_{24}u_{21} & u_{24}u_{22} & u_{24}u_{23} & u_{24}^2 & u_{24}u_{25} \\
0 & 0 & 0 & 0 & u_{25}u_{15} & u_{25}u_{21} & u_{25}u_{22} & u_{25}u_{23} & u_{25}u_{24} & u_{25}^2
\end{matrix}$$

Both firm and time effects ($NT(N + T - 1) = 60$)

Figure 3.2. Structure of Residual Outer Product with Different Standard Error Estimators

This figure visualizes four standard error estimators: heteroskedasticity-consistent, cluster-robust by firm, cluster-robust by time and cluster-robust by both firm and time estimators, respectively. I uses $N = 2$ and $T = 5$ and assume both firm and time effects and hence the matrix $\mathrm{E}\left[\mathbf{uu}^\top\right]$ has 60 nonzero elements. Each estimator estimates the standard error using highlighted elements. First, the upper-left one visualizes a heteroskedasticity-consistent standard error estimator, which uses 10 diagonal elements out of 60 nonzero elements. Second, the upper-right one visualizes a cluster-robust standard error estimator by firm, which uses 50 block diagonal elements. Third, the lower-left one visualizes a cluster-robust standard error estimator by time, which block by block uses 5 diagonal elements, that is, 20 elements. Fourth, the lower-right one visualizes a two-way cluster-robust standard error estimator, which uses 25 elements from diagonal blocks and 5 elements from off-diagonal blocks, that is, 60 elements.

$$\begin{array}{cccccccccc}
u_{11}^2 & u_{11}u_{12} & u_{11}u_{13} & u_{11}u_{14} & u_{11}u_{15} & u_{11}u_{21} & 0 & 0 & 0 & 0 \\
u_{12}u_{11} & u_{12}^2 & u_{12}u_{13} & u_{12}u_{14} & u_{12}u_{15} & 0 & u_{12}u_{22} & 0 & 0 & 0 \\
u_{13}u_{11} & u_{13}u_{12} & u_{13}^2 & u_{13}u_{14} & u_{13}u_{15} & 0 & 0 & u_{13}u_{23} & 0 & 0 \\
u_{14}u_{11} & u_{14}u_{12} & u_{14}u_{13} & u_{14}^2 & u_{14}u_{15} & 0 & 0 & 0 & u_{14}u_{24} & 0 \\
u_{15}u_{11} & u_{15}u_{12} & u_{15}u_{13} & u_{15}u_{14} & u_{15}^2 & 0 & 0 & 0 & 0 & u_{15}u_{25} \\
u_{21}u_{11} & 0 & 0 & 0 & 0 & u_{21}^2 & u_{21}u_{22} & u_{21}u_{23} & u_{21}u_{24} & u_{21}u_{25} \\
0 & u_{22}u_{12} & 0 & 0 & 0 & u_{22}u_{21} & u_{22}^2 & u_{22}u_{23} & u_{22}u_{24} & u_{22}u_{25} \\
0 & 0 & u_{23}u_{13} & 0 & 0 & u_{23}u_{21} & u_{23}u_{22} & u_{23}^2 & u_{23}u_{24} & u_{23}u_{25} \\
0 & 0 & 0 & u_{24}u_{14} & 0 & u_{24}u_{21} & u_{24}u_{22} & u_{24}u_{23} & u_{24}^2 & u_{24}u_{25} \\
0 & 0 & 0 & 0 & u_{25}u_{15} & u_{25}u_{21} & u_{25}u_{22} & u_{25}u_{23} & u_{25}u_{24} & u_{25}^2
\end{array}$$

Heteroskedasticity-consistent (10/60=16.67%)

$$\begin{array}{cccccccccc}
u_{11}^2 & u_{11}u_{12} & u_{11}u_{13} & u_{11}u_{14} & u_{11}u_{15} & u_{11}u_{21} & 0 & 0 & 0 & 0 \\
u_{12}u_{11} & u_{12}^2 & u_{12}u_{13} & u_{12}u_{14} & u_{12}u_{15} & 0 & u_{12}u_{22} & 0 & 0 & 0 \\
u_{13}u_{11} & u_{13}u_{12} & u_{13}^2 & u_{13}u_{14} & u_{13}u_{15} & 0 & 0 & u_{13}u_{23} & 0 & 0 \\
u_{14}u_{11} & u_{14}u_{12} & u_{14}u_{13} & u_{14}^2 & u_{14}u_{15} & 0 & 0 & 0 & u_{14}u_{24} & 0 \\
u_{15}u_{11} & u_{15}u_{12} & u_{15}u_{13} & u_{15}u_{14} & u_{15}^2 & 0 & 0 & 0 & 0 & u_{15}u_{25} \\
u_{21}u_{11} & 0 & 0 & 0 & 0 & u_{21}^2 & u_{21}u_{22} & u_{21}u_{23} & u_{21}u_{24} & u_{21}u_{25} \\
0 & u_{22}u_{12} & 0 & 0 & 0 & u_{22}u_{21} & u_{22}^2 & u_{22}u_{23} & u_{22}u_{24} & u_{22}u_{25} \\
0 & 0 & u_{23}u_{13} & 0 & 0 & u_{23}u_{21} & u_{23}u_{22} & u_{23}^2 & u_{23}u_{24} & u_{23}u_{25} \\
0 & 0 & 0 & u_{24}u_{14} & 0 & u_{24}u_{21} & u_{24}u_{22} & u_{24}u_{23} & u_{24}^2 & u_{24}u_{25} \\
0 & 0 & 0 & 0 & u_{25}u_{15} & u_{25}u_{21} & u_{25}u_{22} & u_{25}u_{23} & u_{25}u_{24} & u_{25}^2
\end{array}$$

Cluster-robust by firm (50/60=83.33%)

$$\begin{array}{cccccccccc}
u_{11}^2 & u_{11}u_{12} & u_{11}u_{13} & u_{11}u_{14} & u_{11}u_{15} & u_{11}u_{21} & 0 & 0 & 0 & 0 \\
u_{12}u_{11} & u_{12}^2 & u_{12}u_{13} & u_{12}u_{14} & u_{12}u_{15} & 0 & u_{12}u_{22} & 0 & 0 & 0 \\
u_{13}u_{11} & u_{13}u_{12} & u_{13}^2 & u_{13}u_{14} & u_{13}u_{15} & 0 & 0 & u_{13}u_{23} & 0 & 0 \\
u_{14}u_{11} & u_{14}u_{12} & u_{14}u_{13} & u_{14}^2 & u_{14}u_{15} & 0 & 0 & 0 & u_{14}u_{24} & 0 \\
u_{15}u_{11} & u_{15}u_{12} & u_{15}u_{13} & u_{15}u_{14} & u_{15}^2 & 0 & 0 & 0 & 0 & u_{15}u_{25} \\
u_{21}u_{11} & 0 & 0 & 0 & 0 & u_{21}^2 & u_{21}u_{22} & u_{21}u_{23} & u_{21}u_{24} & u_{21}u_{25} \\
0 & u_{22}u_{12} & 0 & 0 & 0 & u_{22}u_{21} & u_{22}^2 & u_{22}u_{23} & u_{22}u_{24} & u_{22}u_{25} \\
0 & 0 & u_{23}u_{13} & 0 & 0 & u_{23}u_{21} & u_{23}u_{22} & u_{23}^2 & u_{23}u_{24} & u_{23}u_{25} \\
0 & 0 & 0 & u_{24}u_{14} & 0 & u_{24}u_{21} & u_{24}u_{22} & u_{24}u_{23} & u_{24}^2 & u_{24}u_{25} \\
0 & 0 & 0 & 0 & u_{25}u_{15} & u_{25}u_{21} & u_{25}u_{22} & u_{25}u_{23} & u_{25}u_{24} & u_{25}^2
\end{array}$$

Cluster-robust by time (20/60=33.33%)

$$\begin{array}{cccccccccc}
u_{11}^2 & u_{11}u_{12} & u_{11}u_{13} & u_{11}u_{14} & u_{11}u_{15} & u_{11}u_{21} & 0 & 0 & 0 & 0 \\
u_{12}u_{11} & u_{12}^2 & u_{12}u_{13} & u_{12}u_{14} & u_{12}u_{15} & 0 & u_{12}u_{22} & 0 & 0 & 0 \\
u_{13}u_{11} & u_{13}u_{12} & u_{13}^2 & u_{13}u_{14} & u_{13}u_{15} & 0 & 0 & u_{13}u_{23} & 0 & 0 \\
u_{14}u_{11} & u_{14}u_{12} & u_{14}u_{13} & u_{14}^2 & u_{14}u_{15} & 0 & 0 & 0 & u_{14}u_{24} & 0 \\
u_{15}u_{11} & u_{15}u_{12} & u_{15}u_{13} & u_{15}u_{14} & u_{15}^2 & 0 & 0 & 0 & 0 & u_{15}u_{25} \\
u_{21}u_{11} & 0 & 0 & 0 & 0 & u_{21}^2 & u_{21}u_{22} & u_{21}u_{23} & u_{21}u_{24} & u_{21}u_{25} \\
0 & u_{22}u_{12} & 0 & 0 & 0 & u_{22}u_{21} & u_{22}^2 & u_{22}u_{23} & u_{22}u_{24} & u_{22}u_{25} \\
0 & 0 & u_{23}u_{13} & 0 & 0 & u_{23}u_{21} & u_{23}u_{22} & u_{23}^2 & u_{23}u_{24} & u_{23}u_{25} \\
0 & 0 & 0 & u_{24}u_{14} & 0 & u_{24}u_{21} & u_{24}u_{22} & u_{24}u_{23} & u_{24}^2 & u_{24}u_{25} \\
0 & 0 & 0 & 0 & u_{25}u_{15} & u_{25}u_{21} & u_{25}u_{22} & u_{25}u_{23} & u_{25}u_{24} & u_{25}^2
\end{array}$$

Cluster-robust by both firm and time (60/60=100.00%)

Figure 3.3. Comparison of Clustered Standard Error Estimators by Firm and by Time

This figure displays the percentage of the nonzero elements of $\mathrm{E}\left[\mathbf{u}\mathbf{u}^\top\right]$ that are correctly captured by four standard error estimators when there exist both firm and time effects. For example, when $N = 50$ and $T = 100$, the expected value of the outer product $\mathbf{u}\mathbf{u}^\top$ has 745,000 nonzero elements. First, a heteroskedasticity-consistent standard error estimator exploits the diagonal elements from the outer product and hence the percentage is equal to 5,000/745,000=0.67%. Second, a clustered standard error estimator by firm exploits the block diagonal elements from the outer product and hence the percentage is equal to 500,000/745,000=67.11%. Third, a clustered standard error estimator by time exploits the diagonal elements from the outer product block by block and hence the percentage is equal to 250,000/745,000=33.56%. Fourth, a two-way clustered standard error estimator exploits all 60 elements from the outer product and hence the percentage correctly captured by the estimator is equal to 745,000/745,000=100%. Here the horizontal axis represents N, the number of firms per year and the number of observations $g(N, T) = NT$ is equal to 5,000 with $N, T \in \mathbb{N}$.

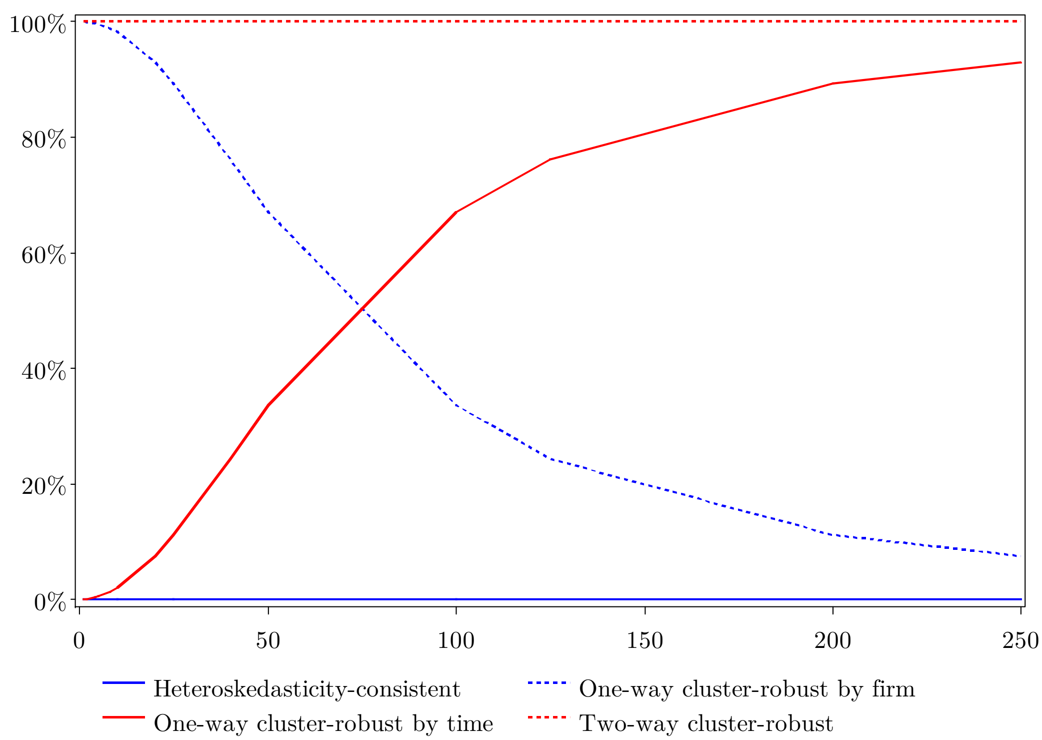


Figure 3.4. Distribution of *t*-Statistics for $\hat{\beta}_4$s by Monte Carlo Simulation with $S = 5,000$

This figure displays the histograms of the t-statistics of $\hat{\beta}_4$s simulated with four different $N \times T = 5,000$ combinations. I specify each combination using the first column. The second, third and fourth columns use NTCR1, NTCR3 and NTCR5 estimates for t-statistics, respectively. For each distribution, I compute Kolmogorov–Smirnov D, Cramér–von Mises W^2 and Anderson–Darling A^2 statistics to test its normality. I attach * (10%), ** (5%) and *** (1%) to indicate significance levels.

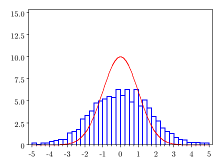
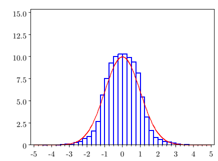
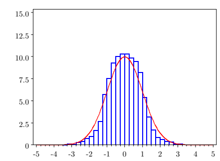
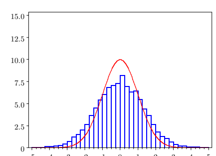
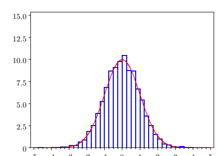
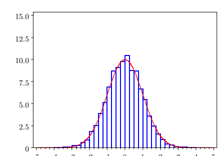
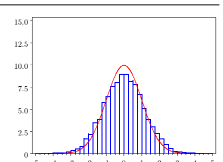
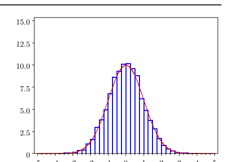
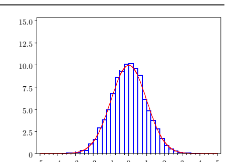
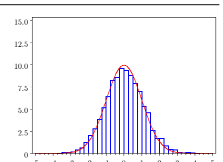
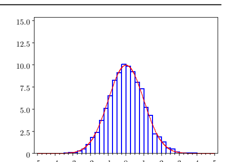
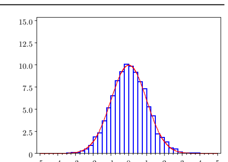
$N \times T$	NTCR1	NTCR3	NTCR5
$1,000 \times 5$			
D	0.1276^{***}	0.0392^{***}	0.0392^{***}
W^2	37.7970^{***}	2.1320^{***}	2.1293^{***}
A^2	468.1177^{***}	19.1213^{***}	19.1399^{***}
500×10			
D	0.0739^{***}	0.0111	0.0111
W^2	11.8414^{***}	0.1113	0.1131
A^2	126.3144^{***}	0.8627	0.8742
250×20			
D	0.0332^{***}	0.0133	0.0138
W^2	2.1242^{***}	0.1149	0.1289
A^2	20.9489^{***}	0.7689	0.8782
100×50			
D	0.0248^{***}	0.0137	0.0145
W^2	0.8347^{***}	0.3478^{*}	0.3466
A^2	6.9412^{***}	2.1430^{*}	2.1008^{*}

Figure 3.5. Distribution of t-Statistics for $\hat{\beta}_4$s by Monte Carlo Simulation with $S = 5,000$

This figure displays the histograms of the t-statistics of $\hat{\beta}_4$s simulated with four different $N \times T = 5,000$ combinations. I specify each combination using the first column. The second, third and fourth columns use HC3, NCR3 and TCR3 estimates for t-statistics, respectively. For each distribution, I compute Kolmogorov–Smirnov D, Cramér–von Mises W^2 and Anderson–Darling A^2 statistics to test its normality. I attach * (10%), ** (5%) and *** (1%) to indicate significance levels.

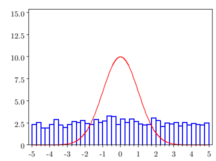
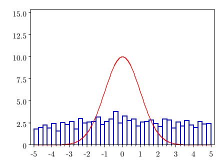
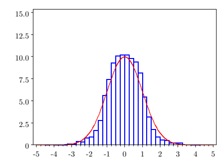
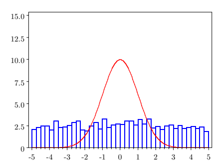
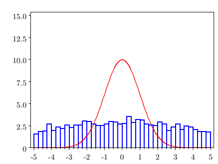
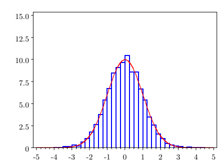
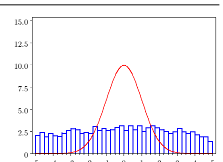
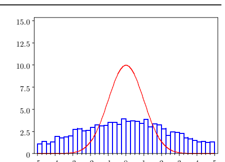
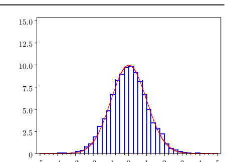
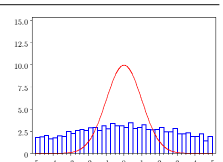
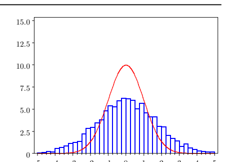
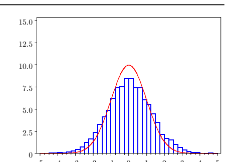
$N \times T$	HC3	NCR3	TCR3
$1,000 \times 5$			
D	0.3895***	0.3716***	0.0369***
W^2	291.4041***	269.6508***	1.8916***
A^2	36,399.7058***	26,407.4282***	16.9206***
500×10			
D	0.3763***	0.3360***	0.0098
W^2	275.2947***	229.5035***	0.0923
A^2	24,194.3751***	12,378.6334***	0.8486
250×20			
D	0.3357***	0.2536***	0.0106
W^2	236.1808***	141.6175***	0.1092
A^2	12,845.9145***	3,666.4240***	1.4025
100×50			
D	0.3096***	0.1324***	0.0587***
W^2	197.2110***	37.8214***	6.2788***
A^2	7,605.4118***	518.6871***	66.9676***

Figure 3.6. Distribution of t-Statistics for $\hat{\beta}_3$s by Monte Carlo Simulation with $S = 5,000$

This figure displays the histograms of the t-statistics of $\hat{\beta}_3$s simulated with four different $N \times T = 5,000$ combinations. I specify each combination using the first column. The second, third and fourth columns use NTCR1, NTCR3 and NTCR5 estimates for t-statistics, respectively. For each distribution, I compute Kolmogorov–Smirnov D, Cramér–von Mises W^2 and Anderson–Darling A^2 statistics to test its normality. I attach * (10%), ** (5%) and *** (1%) to indicate significance levels.

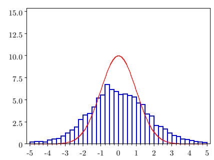
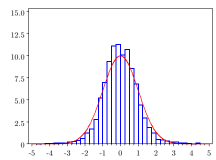
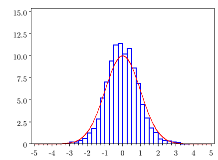
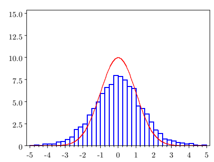
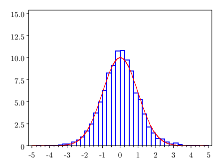
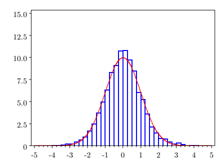
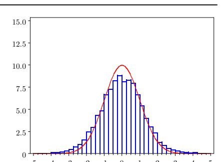
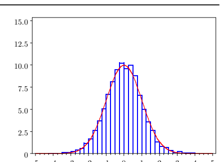
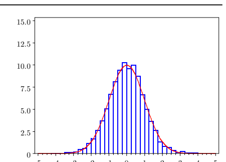
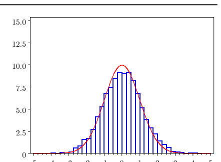
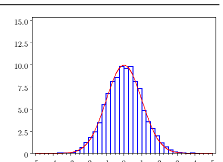
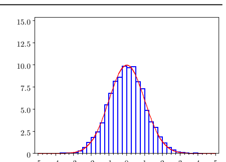
$N \times T$	NTCR1	NTCR3	NTCR5
$1,000 \times 5$			
D	0.1218^{***}	0.0379^{***}	0.0377^{***}
W^2	37.5423^{***}	2.7240^{***}	2.6906^{***}
A^2	539.0116^{***}	19.1646^{***}	18.9372^{***}
500×10			
D	0.0722^{***}	0.0175^{*}	0.0173^{*}
W^2	11.0561^{***}	0.3451	0.3398
A^2	140.8576^{***}	2.8273^{**}	2.8085^{**}
250×20			
D	0.0384^{***}	0.0107	0.0108
W^2	3.0601^{***}	0.0970	0.0992
A^2	33.6466^{***}	1.0222	1.0313
100×50			
D	0.0228^{**}	0.0131	0.0130
W^2	1.1003^{***}	0.2262	0.2176
A^2	10.3956^{***}	1.5206	1.4095

Figure 3.7. Power of t-Tests with Some Two-Way Clustered Standard Error Estimators

This figure displays the power of t-tests for β_4 with six different two-way clustered standard error estimators. Table 3.1 introduces these estimators in detail. I simulate a regression model with $N = 100$ and $T = 10$ $S = 5,000$ times using 21 β_4s from 0 to 2 by 0.1 and then test the null hypothesis $\beta_4 = 1$ at a 5% significance level using t-statistics for $\hat{\beta}_4$s with six standard error estimates. For example, NTCR0 and NTCR1 reject the hypothesis $\beta_4 = 1$ 707 (14.14%) and 628 (12.56%) out of 5,000 times, respectively when $\beta_4 = 1$, while NTCR5 and NTCR3 reject it 330 (6.60%) and 272 (5.44%) out of 5,000 times, respectively. Here the horizontal axis represents $\beta_4 \in [0, 2]$.

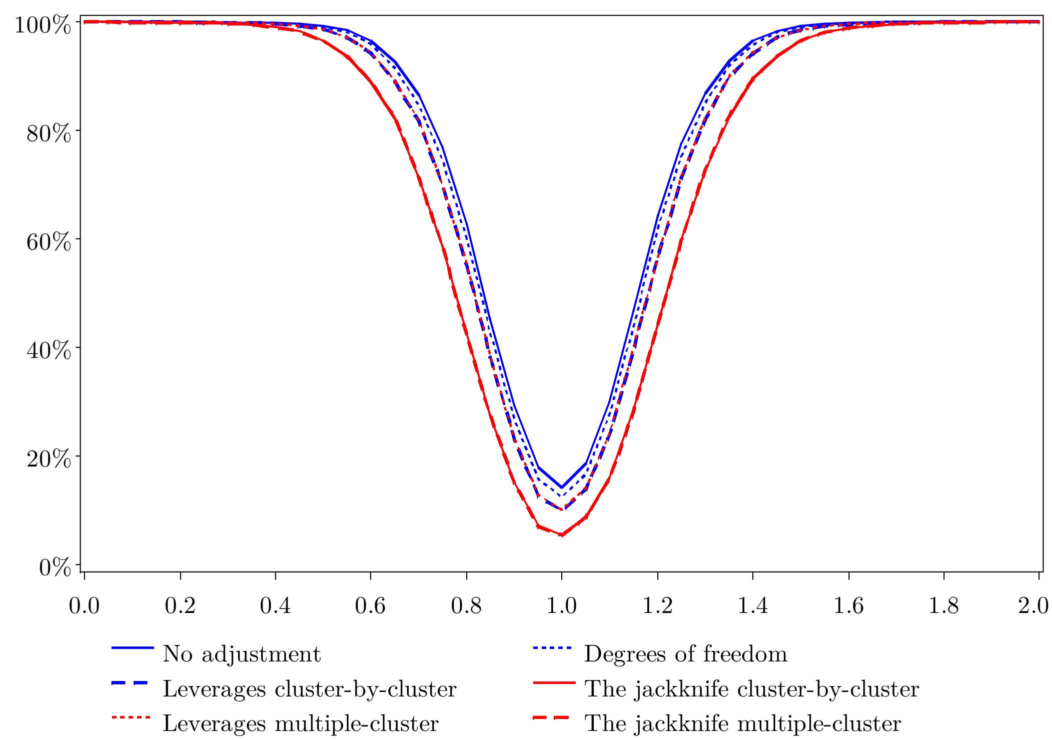
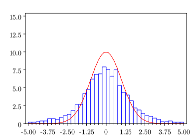
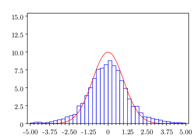
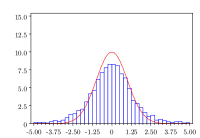
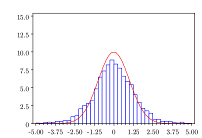
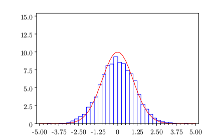


Figure 3.8. Distribution of t-Statistics for $\hat{\beta}$s under Three-Way Effects by Combination

This figure displays the histograms of the t-statistics of $\hat{\beta}$s simulated under three-way effects with four different $I \times J \times K = 1,728$ combinations: (3,24,24), (6,12,24), (8,8,12) and (12,12,12). S, the number of simulations, is equal to 5,000 for all triples. I specify each combination using the first row. For the first four combinations, I use neither degrees of freedom nor leverages to adjust three-way clustered standard error estimates. For the last combination, I use leverages to adjust the estimates. I report the number of positive standard error estimates as well as other major findings because the standard error estimates are not always positive. For each distribution, I compute Kolmogorov–Smirnov D, Cramér–von Mises W^2 and Anderson–Darling A^2 statistics to test its normality. I attach * (10%), ** (5%) and *** (1%) to indicate significance levels.

$I \times J \times K$	$3 \times 24 \times 24$	$6 \times 12 \times 24$	$8 \times 8 \times 27$	$12 \times 12 \times 12$	
Finite sample adjustment	No				Yes
					
# of positive SEs	4,521	4,695	4,709	4,766	5,000
Average of $\hat{\beta}$s	0.9995	0.9997	0.9994	0.9997	0.9997
Standard deviation of $\hat{\beta}$s	0.0301	0.0283	0.0289	0.0278	0.0279
Average of 3-way SEs	0.0229	0.0241	0.0239	0.0240	0.0253
Kurtosis of t-statistics	280.4199	17.4842	150.3746	16.4502	0.1093
% of rejecting H_0 at 10%	0.2617	0.2034	0.2158	0.2042	0.1340
% at 5%	0.2000	0.1410	0.1571	0.1399	0.0738
% at 1%	0.1232	0.0818	0.0862	0.0638	0.0202
D	0.0860***	0.0563***	0.0680***	0.0572***	0.0269***
W^2	15.6234***	6.0083***	7.9954***	5.9559***	1.4547***
A^2	313.6319***	122.6492***	155.0202***	99.5726***	13.2014***

Bibliography

Ali, Usman, Kent Daniel, and David Hirshleifer, 2017, One brief shining moment(um): Past momentum performance and momentum reversals, *Working Paper* .

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006, The cross-section of volatility and expected returns, *Journal of Finance* 61, 259–299.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2009, High idiosyncratic volatility and low returns: International and further us evidence, *Journal of Financial Economics* 91, 1–23.

Angrist, Joshua D, and Jörn-Steffen Pischke, 2008, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press).

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, 2013, Value and momentum everywhere, *Journal of Finance* 68, 929–985.

Baele, Lieven, Geert Bekaert, Koen Inghelbrecht, and Min Wei, 2019, Flights to safety, *Review of Financial Studies* .

Bali, Turan G, Nusret Cakici, and Robert F Whitelaw, 2011, Maxing out: Stocks as lotteries and the cross-section of expected returns, *Journal of Financial Economics* 99, 427–446.

Barberis, Nicholas, and Ming Huang, 2008, Stocks as lotteries: The implications of probability weighting for security prices, *American Economic Review* 98, 2066–2100.

Barroso, Pedro, and Pedro Santa-Clara, 2015, Momentum has its moments, *Journal of Financial Economics* 116, 111 – 120.

Barsky, Robert, 1989, Why don't the prices of stocks and bonds move together?, *American Economic Review* 79, 1132–45.

Bekaert, Geert, Robert J Hodrick, and Xiaoyan Zhang, 2009, International stock return comovements, *Journal of Finance* 64, 2591–2626.

Bell, Robert M, and Daniel F McCaffrey, 2002, Bias reduction in standard errors for linear regression with multi-stage samples, *Survey Methodology* 28, 169–182.

Blitz, David, Joop Huij, and Martin Martens, 2011, Residual momentum, *Journal of Empirical Finance* 18, 506–521.

Boudry, Walter I., Robert A. Connolly, and Eva Steiner, 2019, What happens during flight to safety: Evidence from public and private real estate markets, *Real Estate Economics* 1–26.

Boyer, Brian, Todd Mitton, and Keith Vorkink, 2010, Expected idiosyncratic skewness, *Review of Financial Studies* 23, 169–202.

Caballero, Ricardo J., and Arvind Krishnamurthy, 2008, Collective risk management in a flight to quality episode, *Journal of Finance* 63, 2195–2230.

Callen, Jeffrey L, and Dan Segal, 2004, Do accruals drive firm-level stock returns? A variance decomposition analysis, *Journal of Accounting Research* 42, 527–560.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, 2008, Bootstrap-based improvements for inference with clustered errors, *Review of Economics and Statistics* 90, 414–427.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, 2011, Robust inference with multiway clustering, *Journal of Business & Economic Statistics* 29, 238–249.

Cameron, A Colin, and Douglas L Miller, 2015, A practitioner's guide to cluster-robust inference, *Journal of Human Resources* 50, 317–372.

Cameron, A Colin, and Pravin K Trivedi, 2005, *Microeconometrics: Methods and Applications* (Cambridge University Press).

Campbell, John, Christopher Polk, and Tuomo Vuolteenaho, 2010, Growth or glamour? Fundamentals and systematic risk in stock returns, *Review of Financial Studies* 23, 305–344.

Campbell, John Y, 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.

Campbell, John Y, and John Ammer, 1993, What moves the stock and bond markets? A variance decomposition for long-term asset returns, *Journal of Finance* 48, 3–37.

Campbell, John Y, Stefano Giglio, and Christopher Polk, 2013, Hard times, *Review of Asset Pricing Studies* 3, 95–132.

Campbell, John Y, Martin Lettau, Burton G Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of Finance* 56, 1–43.

Campbell, John Y, Andrew Wen-Chuan Lo, Archie Craig MacKinlay, et al., 1997, *The Econometrics of Financial Markets* (Princeton University Press).

Campbell, John Y, and Robert J Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y, and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.

Cao, Jie, and Bing Han, 2016, Idiosyncratic risk, costly arbitrage, and the cross-section of stock returns, *Journal of Banking & Finance* 73, 1–15.

Carhart, Mark M, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Cenedese, Gino, and Enrico Mallucci, 2016, What moves international stock and bond markets?, *Journal of International Money and Finance* 60, 94–113.

Chen, Linda, George Jiang, Danielle Xu, and Tong Yao, 2012, Dissecting the idiosyncratic volatility anomaly, *Working Paper* .

Chen, Long, Zhi Da, and Xinlei Zhao, 2013, What drives stock price movements?, *Review of Financial Studies* 26, 841–876.

Chen, Long, and Xinlei Zhao, 2009, Return decomposition, *Review of Financial Studies* 22, 5213–5249.

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

Cooper, Michael J., Roberto C. Gutierrez Jr., and Allaudeen Hameed, 2004, Market states and momentum, *Journal of Finance* 59, 1345–1365.

Cribari-Neto, Francisco, 2004, Asymptotic inference under heteroskedasticity of unknown form, *Computational Statistics & Data Analysis* 45, 215–233.

Cribari-Neto, Francisco, and Maria da Glória A Lima, 2009, Heteroskedasticity-consistent interval estimators, *Journal of Statistical Computation and Simulation* 79, 787–803.

Daniel, Kent, and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221 – 247.

Davidson, Russell, and James G MacKinnon, 2004, *Econometric theory and methods* (Oxford University Press).

Efron, Bradley, 1982, *The Jackknife, the Bootstrap, and Other Resampling Plans* (SIAM).

Egginton, Jared, and Jungshik Hur, 2018, The robust "maximum daily return effect as demand for lottery" and "idiosyncratic volatility puzzle", *Journal of Empirical Finance* .

Ehsani, Sina, and Juhani Linnainmaa, 2019, Factor momentum and the momentum factor, *Working Paper* .

Engsted, Tom, Thomas Q Pedersen, and Carsten Tanggaard, 2012, Pitfalls in var based return decompositions: A clarification, *Journal of Banking & Finance* 36, 1255–1265.

Fama, Eugene F, and Kenneth R French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F, and Kenneth R French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

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

Fama, Eugene F, and Kenneth R French, 2016, Dissecting anomalies with a five-factor model, *Review of Financial Studies* 29, 69–103.

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

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

Fink, Jason D, Kristin E Fink, and Hui He, 2012, Expected idiosyncratic volatility measures and expected returns, *Financial Management* 41, 519–553.

Froot, Kenneth A, 1989, Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data, *Journal of Financial and Quantitative Analysis* 24, 333–355.

Fu, Fangjian, 2009, Idiosyncratic risk and the cross-section of expected stock returns, *Journal of Financial Economics* 91, 24–37.

Goetzmann, William N., and Simon Huang, 2018, Momentum in Imperial Russia, *Journal of Financial Economics* 130, 579 – 591.

Goyal, Amit, and Pedro Santa-Clara, 2003, Idiosyncratic risk matters!, *Journal of Finance* 58, 975–1007.

Goyal, Amit, and Sunil Wahal, 2015, Is momentum an echo?, *Journal of Financial and Quantitative Analysis* 50, 1237–1267.

Greene, William H, 2003, *Econometric Analysis* (Pearson Education).

Griffin, John M, Xiuqing Ji, and J Spencer Martin, 2003, Momentum investing and business cycle risk: Evidence from pole to pole, *Journal of Finance* 58, 2515–2547.

Grundy, Bruce D., and J. Spencer Martin Martin, 2001, Understanding the nature of the risks and the source of the rewards to momentum investing, *Review of Financial Studies* 14, 29–78.

Guo, Hui, Haimanot Kassa, and Michael F Ferguson, 2014, On the relation between egarch idiosyncratic volatility and expected stock returns, *Journal of Financial and Quantitative Analysis* 49, 271–296.

Han, Yufeng, and David Lesmond, 2011, Liquidity biases and the pricing of cross-sectional idiosyncratic volatility, *Review of Financial Studies* 24, 1590–1629.

Harvey, Campbell R, Edward Hoyle, Russell Korgaonkar, Sandy Rattray, Matthew Sargaison, and Otto Van Hemert, 2018, The impact of volatility targeting, *Journal of Portfolio Management* 45, 14–33.

Hausman, Jerry, and Christopher Palmer, 2012, Heteroskedasticity-robust inference in finite samples, *Economics Letters* 116, 232–235.

He, Zhiguo, and Arvind Krishnamurthy, 2012, A model of capital and crises, *Review of Economic Studies* 2, 735–777.

Hinkley, David V, 1977, Jackknifing in unbalanced situations, *Technometrics* 19, 285–292.

Horn, Susan D, Roger A Horn, and David B Duncan, 1975, Estimating heteroscedastic variances in linear models, *Journal of the American Statistical Association* 70, 380–385.

Hou, Kewei, and Roger K Loh, 2016, Have we solved the idiosyncratic volatility puzzle?, *Journal of Financial Economics* 121, 167–194.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Huang, Wei, Qianqiu Liu, S Ghon Rhee, and Liang Zhang, 2010, Return reversals, idiosyncratic risk, and expected returns, *Review of Financial Studies* 23, 147–168.

Huang, Wei, Qianqiu Liu, S Ghon Rhee, and Liang Zhang, 2011, Another look at idiosyncratic volatility and expected returns, *Journal of Investment Management* 9, 26–51.

Hurst, Brian, Yao Hua Ooi, and Lasse Heje Pedersen, 2017, A century of evidence on trend-following investing, *Journal of Portfolio Management* 44, 15–29.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993a, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993b, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jiang, George J, Danielle Xu, and Tong Yao, 2009, The information content of idiosyncratic volatility, *Journal of Financial and Quantitative Analysis* 44, 1–28.

Kothari, Shanker P, and Jay Shanken, 1992, Stock return variation and expected dividends: A time-series and cross-sectional analysis, *Journal of Financial Economics* 31, 177–210.

Krishnamurthy, Arvind, 2010, Amplification mechanisms in liquidity crises, *American Economic Journal: Macroeconomics* 2, 1–30.

Lehmann, Bruce N, 1990, Residual risk revisited, *Journal of Econometrics* 45, 71–97.

Lochstoer, Lars A, and Paul C Tetlock, 2019, What drives anomaly returns?, *Working Paper* .

Long, J Scott, and Laurie H Ervin, 2000, Using heteroscedasticity consistent standard errors in the linear regression model, *American Statistician* 54, 217–224.

Longstaff, Francis A, 1989, Temporal aggregation and the continuous-time capital asset pricing model, *Journal of Finance* 44, 871–887.

MacKinnon, James G, 2013, Thirty years of heteroskedasticity-robust inference, in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, 437–461 (Springer).

MacKinnon, James G, and Halbert White, 1985, Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties, *Journal of Econometrics* 29, 305–325.

Merton, Robert C, 1987, A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42, 483–510.

Moreira, Alan, and Tyler Muir, 2017, Volatility-managed portfolios, *Journal of Finance* 72, 1611–1644.

Moskowitz, Tobias J, Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.

Novy-Marx, Robert, 2012, Is momentum really momentum?, *Journal of Financial Economics* 103, 429 – 453.

Novy-Marx, Robert, 2015, Fundamentally, momentum is fundamental momentum, *Working Paper* .

Petersen, Mitchell A, 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Poirier, Dale J, 2011, Bayesian interpretations of heteroskedastic consistent covariance estimators using the informed bayesian bootstrap, *Econometric Reviews* 30, 457–468.

Politis, Dimitris N, 2011, Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices, *Econometric Theory* 27, 703–744.

Rogers, William H, 1993, sg17: Regression standard errors in clustered samples, *Stata Technical Bulletin* 3, 19–23.

Shao, Jun, and JNK Rao, 1993, Jackknife inference for heteroscedastic linear regression models, *Canadian Journal of Statistics* 21, 377–395.

Skoulakis, Georgios, 2008, Panel data inference in finance: Least-squares vs Fama-Macbeth, Working Paper.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan, 2015, Arbitrage asymmetry and the idiosyncratic volatility puzzle, *Journal of Finance* 70, 1903–1948.

Stivers, Chris, and Licheng Sun, 2010, Cross-sectional return dispersion and time variation in value and momentum premiums, *Journal of Financial and Quantitative Analysis* 45, 987–1014.

Thompson, Samuel B, 2011, Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* 99, 1–10.

Vayanos, Dimitri, 2004, Flight to quality, flight to liquidity, and the pricing of risk, *Working Paper* .

Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns?, *Journal of Finance* 57, 233–264.

Webb, Matthew D, 2008, Reworking wild bootstrap based inference for clustered errors, Working Paper.

White, Halbert, 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

Wooldridge, Jeffrey M, 2010, *Econometric Analysis of Cross Section and Panel Data* (MIT Press).

Xu, Yexiao, and Burton G Malkiel, 2004, Idiosyncratic risk and security returns, *Working Paper* .

Yeh, Chung-Ying, Junming Hsu, Kai-Li Wang, and Che-Hui Lin, 2015, Explaining the default risk anomaly by the two-beta model, *Journal of Empirical Finance* 30, 16–33.

Curriculum Vitae

Junyong Kim

Place of birth: Seoul, Korea

Education

B.B.A., Kyung Hee University, February 2013

Major: Business Administration

M.S., Seoul National University, February 2015

Major: Finance

Ph.D., University of Wisconsin–Milwaukee, August 2020

Major: Management Science, Finance

Minor: Economics, Econometrics

Dissertation Title: Three Essays on Market Anomalies and Financial Econometrics