The London School of Economics and Political Science

Essays in Factor-Based Investing

Tengyu (James) Guo

A thesis submitted to the Department of Finance of the London School of Economics and Political Science for the degree of Doctor of Philosophy, London, September 2019

Declaration

I certify that the thesis I have presented for examination for the Ph.D. degree of the London School of Economics and Political Science is solely my own work.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 52,098 words.

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Dong Lou for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. I would also like to thank Christopher Polk and Thummim Cho for their supervision, guidance, and help. I could not have imagined having better advisors and mentors for my Ph.D. study and research.

Besides my advisors, I would like to thank the faculty members that I have met at LSE: Ashwini Agrawal, Ron Anderson, Ulf Axelson, Elisabetta Bertero, Mike Burkart, Georgy Chabakauri, Vicente Cuñat, Jon Danielsson, Amil Dasgupta, Daniel Ferreira, Juanita Gonzalez-Uribe, Charles Goodhart, Moqi Groen-Xu, Dirk Jenter, Christian Julliard, Peter Kondor, Paula Lopes-Cocco, Igor Makarov, Ian Martin, Martin Oehmke, Daniel Paravisini, Cameron Peng, Rohit Rahi, Domingos Romualdo, Dimitri Vayanos, Michela Verardo, David Webb, Kathy Yuan, Hongda Zhong, and Jean-Pierre Zigrand. I benefit from their classes in various courses, from insightful comments in seminars, and encouragement, but also from the hard questions which incentivized me to broaden my research and skill set.

My sincere thanks also goes to researchers that I have met outside LSE and along my studies: Dante Amengual, Manuel Arellano, Pedro Barroso, Samuel Bentolila, Simon Gervais, Jeewon Jang, Liguori Jego, Marcin Kacperczyk, Ralph Koijen, Daisuke Miyakawa, Omar Licandro, Rafael Repullo, Enrique Sentana, Javier Suarez, Sunil Wahal, Nancy Xu, and Yu Yuan among the others.

My time at LSE Finance would have been different without many great friends that I have met during my years as a Ph.D. student. I would like to thank them for their friendship and support during all these years: Huaizhi Chen, Jingxuan Chen, Juan Chen, Hoyong Choi, Fabrizio Core, Bernardo De Oliveira Guerra Ricca, Andreea Englezu, Jesus Gorrin, Sergei Glebkin, David Haller, Brandon Yueyang Han, Miao Han, Zhongchen Hu, Jiantao Huang, Shiyang Huang, Bruce Iwadate, Lukas Kremens, Yiqing Lu, Francesco Nicolai, Olga Obishaeva, Dimitris Papadimitriou, Jiho Park, Alberto Pellicioli, Marco Pelosi, Poramapa Poonpakdee, Michael Punz, Amirabas Salarkia, Ran Shi, Simona Risteska, Irina Stanciu, Claudia Robles-Garcia, Gosia Malgorzata Ryduchowska, Petar Sabtchevsky, Una Savic, Seyed Seyedan, Ji Shen, Arthur Taburet, Bo Tang, Jiaxing Tian, Karamfil Todorov, Su Wang, Yue

Wu, Yun Xue, Xiang Yin, Cheng Zhang and most importantly, Yue Yuan. I also appreciate the company from friends that I made at different places and the help from the administration team at LSE Finance department, and in particular, Mary Comben.

Last but not least, I would like to thank my parents, Jiali and Chenggang, for giving me life and supporting me as much as they can.

To this beautiful world. *Rerum cognoscere causas.*

Abstract

My thesis explores three questions in factor-based investing.

In the first chapter, I study the correlation risk in trading stock market anomalies. I propose a simple time-series risk measure in trading stock market anomalies, CoAnomaly, the time-varying average pairwise correlation among 34 anomalies, which helps to explain both the time-series and the cross-sectional anomaly return patterns. Since correlations among underlying assets determine the portfolio variance, CoAnomaly is an important state variable for arbitrageurs who hold diversified portfolios of anomalies to boost their performance. Empirically, I show that, first, CoAnomaly is persistent and forecasts long-run aggregate volatility of the diversified anomaly portfolio. Second, CoAnomaly positively predicts future average anomaly returns in the time series. Third, in the cross-section of these 34 anomaly portfolios, CoAnomaly carries a negative price of risk.

In the second chapter, instead of studying multiple anomalies in a portfolio, I focus on one specific anomaly, momentum. I find that the momentum spread negatively predicts momentum returns in the long-term, but not in the following month. I further decompose the momentum spread into the spread of young or old momentum stocks based on how long the stock has been identified as a momentum stock. I show that the negative predictability is mainly driven by the old momentum spread. As these old momentum stocks are more likely to be exploited by arbitrageurs, these findings suggest that momentum is amplified by arbitrage activity and excessive arbitrage destabilizes the asset prices and generates strong reversals.

In the third chapter, I revisit the robust diversification of factor investing and study the intertemporal consideration of an anomaly investor. Motivated by Campbell et al. (2017), I use vector autoregressions (VAR) and estimate an intertemporal CAPM with stochastic volatility for market-neutral investing with the focus on a portfolio of 34 anomalies. Interestingly, based on my estimation, only the correlation-induced volatility news carries a significant risk premium, which echos my findings in the first chapter.

Contents

List of Tables

List of Figures

Chapter 1

CoAnomaly: Correlation Risk in Stock Market Anomalies

I propose a simple time-series risk measure in trading stock market anomalies, *CoAnomaly*, the time-varying average correlation among 34 anomalies, which helps to explain both the time-series and the cross-sectional anomaly return patterns. Since correlations among underlying assets determine the portfolio variance, CoAnomaly is an important state variable for arbitrageurs who hold diversified portfolios of anomalies to boost their performance. Empirically, I show that, first, CoAnomaly is persistent and forecasts long-run aggregate volatility of the diversified anomaly portfolio. Second, CoAnomaly positively predicts future average anomaly returns in the time series. Third, in the cross-section of these 34 anomaly portfolios, CoAnomaly carries a negative price of risk. These return patterns suggest that arbitrageurs take the time-varying correlation into account and their intertemporal hedging demand plays an important role in setting asset prices.

JEL-Classification: G11, G23.

Keywords: Stock Market Anomalies, Time-Varying Correlation Risk, Hedging Demand of Sophisticated Investors.

1.1 Introduction

Stock market anomalies are long-short portfolios exploiting stock characteristics known to predict returns. Since these anomalies generate returns beyond what standard notions of risk suggest, they have been the subject of a large body of literature which tries to understand the origins of these anomalies. More recently, researchers have started to study multiple anomalies jointly; however, they mainly focus on either reducing the dimensionality of the anomaly space[1], evaluating a new factor given existing factors[2], or comparing the costs/risks of trading each of these anomalies[3].

I propose a novel way to understand anomaly return dynamics through the lens of sophisticated arbitrageurs who trade a diversified portfolio of anomalies. These arbitrageurs understand the return-volatility trade-off and recognize that the portfolio volatility is time-varying. If cross-anomaly correlation is the main source of news about current and future volatility of a diversified portfolio of the anomalies, then the time-varying cross-anomaly correlation arises as a *state variable* that predicts the future returns on a diversified portfolio of anomalies in the time series and as a *risk factor* that gets priced in the cross-section.

Based on this idea, I propose a simple measure of correlation risk faced by sophisticated arbitrageurs. This measure, which I term *CoAnomaly*, works both as a predictor of future average anomaly return and as a risk factor that prices the cross-section of anomaly portfolios.

To construct CoAnomaly, at each point of time, I calculate the average partial correlation among anomaly returns using daily data within a time window. Effectively, this measure evaluates how much these anomalies comove with each other across time. I find that CoAnomaly shows time-series persistence and it is not particularly correlated with major existing risk measures, nor strongly predicted by other state variables. More importantly, I find that CoAnomaly is the only variable that can strongly and persistently predict future aggregate variance of the diversified anomaly portfolio up to one year. On the other hand, aggregate variance or

[1]See Fama and French (1996), Hou et al. (2015), Stambaugh and Yuan (2016), Harvey et al. (2016) and Green et al. (2017) among the others.

[2]See Harvey et al. (2016), Feng et al. (2017), and Chinco et al. (2019) among the others.

[3]See Barroso and Santa-Clara (2015), Novy-Marx and Velikov (2016), Moreira and Muir (2017), and Barroso and Maio (2018) among the others.

average variance of anomalies is quickly mean-reverting and shows no such long-run predictability. I do not intend to explain the source of correlations among anomalies at the moment and simply take the time-varying correlation structure as given and study the asset pricing implications for anomaly investing.

I first explore the time-series predictability of CoAnomaly. As expected, I find that the naïve equal-weighted anomaly return (E.A.R.) is higher after high CoAnomaly periods. I focus on E.A.R. for simplicity and I also show that this predictability applies to the in-sample mean-variance portfolio, which consists of these 34 anomalies. The predictive regression shows that the quarterly E.A.R. will be 70 basis points higher following a one-standard-deviation increase in CoAnomaly. This predictability shows up robustly after controlling for other predictors, spans from one month to one year, and does not suffer from the Stambaugh (1999) bias. This predictability is robust to long and short legs of anomalies as well as different sets of anomalies. To ease up the interpretation, I provide time-series sorting evidence. For 30% of the sample period associated with high levels of CoAnomaly, the E.A.R. yields 1.20% higher returns over the next 6 months, relative to its performance following the 30% of the sample period associated with low CoAnomaly. The periods with higher future anomaly returns do not mean that they are good states for arbitrageurs, as I find that the future standard deviation of E.A.R. is higher and its skewness is more negative after high CoAnomaly periods. Moreover, I find that the predictive power is stronger in the recent half of my sample period, as the E.A.R. performance difference between high CoAnomaly periods and low CoAnomaly periods almost doubles (from 1.20% to 2.36%). This sample period coincides with the period of fast emergence and growth of professional asset managers. Furthermore, the predictability is even more pronounced when I focus on the periods when quantitative equity hedge funds suffer a low return (a proxy for high risk aversion), which is consistent with the risk-return trade-off story from the perspective of these arbitrageurs. I also conduct a battery of robustness checks to make sure my results would not be driven by several possible mechanisms.

In the cross-section, I find that the innovation of CoAnomaly carries a negative price of risk. The negative sign of risk premium suggests that higher CoAnomaly periods are hard times for these arbitrageurs as their portfolio variance goes up. I check two sets of testing assets: the anomaly set including the 34 anomaly portfolios, which are used to calculate the CoAnomaly measure, and a standard set which includes equity portfolios sorted by size, book-to-market, momentum, and industry, as well as the cross-section of Treasury bond portfolios sorted by maturity. The price of risk is significant, and its magnitude is consistent across test portfolios and specifications. On average, the quarterly risk premium associated with one unit of the CoAnomaly beta is around -5%. In other words, investors demand lower expected returns for portfolios that covary positively with CoAnomaly, since these portfolios effectively provide hedges against the CoAnomaly risk. This finding is consistent the with the recent event, 'quant meltdown' in August 2007, documented by Khandani and Lo (2007) and Khandani and Lo (2011), in which they argue that this hard period for quantitative investors is accompanied by high correlation among their strategies. The fact that the correlation risk also gets priced in a broader set of portfolios (industry portfolios and bonds) suggests the large presence of these asset managers in different financial markets. Furthermore, the loadings on CoAnomaly innovation help explain the cross-sectional return dispersion across these portfolios. That said, my single factor does not fully explain the cross-section of anomaly returns as large intercepts are left unexplained. I further construct CoAnomaly beta sorted portfolios from stock level based on their estimated real-time CoAnomaly betas. I find large dispersion in adjusted returns that line up well with the post-ranking CoAnomaly betas.

My paper contributes to the literature in two aspects. First, I show that both the time-series and cross-sectional return dynamics of stock market anomalies can be partially explained by taking a portfolio view of these anomalies. This view stems from the perspective of portfolio managers in quantitative equity hedge funds who are betting on these anomalies, goes back to the basic trade-off between risk and return, and studies the time-varying risk of trading these anomalies as diversified portfolios. By focusing on the correlations among these anomalies that vary across time, I provide evidence supporting the basic risk-return trade-off relationship in anomaly asset prices. Moreover, this also sheds light on the understanding of the volatility-managed portfolios by highlighting the importance of the correlation risk. Existing literature mainly focuses on the volatility of a specific anomaly alone. In contrast, I study these anomalies jointly in a portfolio and show that it is the comovement among these anomalies that plays a key role in evaluating the risk.

Second, CoAnomaly gives a better understanding of the investing behaviors of the fast-growing sophisticated institutional investors, as I find that the time-series predictability is stronger in the recent half of the sample, which coincides the period with rapid growth in asset management industry[4]. Moreover, the predictability is even more pronounced when these arbitrageurs are suffering low returns, presumably inducing a higher risk aversion and making their trade-off incentives stronger. On the other hand, since the CoAnomaly risk mechanically affects the investment opportunity of these arbitrageurs, the hedging demand on this risk feeds back into asset prices and a negative risk premium of CoAnomaly shows up in the cross-sectional dispersion of average asset returns. My findings suggest that, first, these arbitrageurs understand that the risk they face is time-varying and know how to pick certain assets to hedge it; and second, their impacts to the market are substantial so we can observe these return patterns from asset prices.

1.2 Related Literature

The exploration of *stock market anomalies* starts from the early testing on CAPM and the empirical failure of a single market factor witnesses the explosion of identifying stock market anomalies by academics as well as by practitioners. Finance researchers find it difficult to reconcile them with standard asset pricing models, and several streams of approaches have been proposed to understand them: ex-post factor models, principal component analysis, behavioral stories, intermediary asset pricing, etc. On the practitioners' side, it is not just the arbitrageurs strictly chasing market neutrality, but also investors whose portfolio deviates from the market portfolio are loading on certain strategies to some extent. Most recently, the Exchange-Traded Fund (ETF) industry also started issuing factor-based '*smart beta*' products, and both long-term investors and retail investors are investing in these assets, hoping to boost their Sharpe ratio (see Cao et al. (2018)). There has always been a debate about whether these anomalies represent true risk-adjusted excess

[4]Schwert (2003) and Chordia et al. (2011) find evidence that arbitrage activity on anomalies increased following early 1990s.

returns or whether they are compensation for omitted risk factors. I take no stand on this issue and simply adopt the consensus interpretation of the market-neutral quantitative equity investors who believe these anomalies do generate alphas with respect to the market portfolio. Though concerns have been raised on whether there are too many anomalies (see Hou et al. (2017)) or whether they can survive transaction costs (see Novy-Marx and Velikov (2016)), Martin Utrera et al. (2017) find that transaction costs increase the number of significant characteristics due to the canceling-out of transaction costs when combining and rebalancing characteristics.

It has been documented that these anomalies are indeed exploited by sophisticated investors. Hanson and Sunderam (2013) use the short interest to show that the amount of capital devoted to value and momentum, the two most prominent strategies, has grown significantly since the late 1980s; McLean and Pontiff (2016) find that anomaly returns are lower after publication. Because of the sophisticated nature of these *large and professional agents*, some researchers argue that they are aware of the (endogenous) systemic risk and will internalize the impact of their behavior. Koijen and Yogo (2015) find that most cross-sectional variation in stock returns is contributed to retail investors instead of large asset managers. Meanwhile, there are also concerns about their roles and impacts, as Stein (2009) points out that *crowding* and *leverage* can impair market efficiency and argue that capital regulation may be helpful in dealing with the latter problem. Both theoretical work and empirical evidence show this destabilizing effect of arbitrageurs, see Vayanos and Woolley (2013) and Lou and Polk (2013).

When it comes to trading these anomalies, a proper risk measure is necessary to access the cost and benefit. However, recent literature has documented the failure in proper pricing of *variance risk* in finance as well as in macroeconomics. Dew-Becker et al. (2017) find it was costless on average to hedge news about future variance at horizons ranging from 1 quarter to 14 years between 1996 and 2014, and only unexpected, transitory realized variance was significantly priced. Berger et al. (2017) find that shocks to uncertainty have no significant effect on the economy, even though shocks to realized stock market volatility are contractionary according to a wide range of VAR specifications. On the other hand, as assets comove together, the magnitude of the average correlation clearly affects how much investors can diversify and hence the future risk premium. *Correlation Risk* is studied pervasively. Pollet and Wilson (2010) show that the average correlation between daily stock returns predicts subsequent quarterly stock market excess returns since the market risk is determined by the *individual risks* and the *correlation* among them. They start from a measurement error issue of aggregate risk and show that changes in true aggregate risk may nevertheless reveal themselves through changes in the correlation between observable stock returns. However, I take a portfolio view of anomalies from the arbitrageurs and show that the average correlation is an important state variable, which acts as both a predictor in the time-series and a priced risk in the cross-section. Driessen et al. (2009) study the different exposures to correlation risk between index options and individual stock options and find that correlation risk exposure explains the cross-section of the index and individual option returns well. Buraschi et al. (2010) provides a theoretical model in which the degree of correlation across industries, countries, or asset classes is stochastic. Buraschi et al. (2013) find that the ability of hedge funds to create market-neutral returns is often associated with significant exposure to correlation risk, which helps explain the large abnormal returns found in previous models, and they also estimate a significant negative market price of correlation risk. Adrian and Brunnermeier (2016) propose a measure for systemic risk: CoVaR, the value at risk (VaR) of financial institutions conditional on other institutions being in distress.

However, most research on the correlation risk mainly focuses on the correlation risk in the aggregate stock market. This paper takes a novel perspective of looking at the anomaly space from the scope of a portfolio manager chasing market neutrality and studies the time-series predictability and cross-sectional pricing together. As a closely related research, Stambaugh et al. (2012) find that investor sentiment positively predicts anomaly returns and argue that short-sale impediments contribute to their finding as their effect concentrates on the short legs of anomalies. My result is different from theirs in the sense that the predictability of CoAnomaly shows up for both long legs and short legs, which is in line with the basic trade-off between risk and expected return. Sotes-Paladino (2017) explores the optimal dynamic investment problem when mispricing assets are correlated, in which he considers a constant correlation structure. On the other hand, this paper highlights the importance of time-variation in the correlation structure.

1.3 CoAnomaly

1.3.1 Data and Anomaly Construction

To construct the stock market anomalies, I use the stock return data from the Center for Research in Security Prices (CRSP). The accounting data is taken from Compustat - Capital IQ, and short interest data from Supplemental Short Interest File of Compustat - Capital IQ. Since I use quarterly accounting information with valid *Report Date of Quarterly Earnings (RDQ)* starting from late 1971 to construct some anomalies, my main sample period starts in 1973 and ends in 2017. As my later analysis does not depend on a balanced panel, I extend my main sample period back to 1963 as a robustness check in the appendix (with anomalies which do not require *RDQ* information), which generates consistent results.

The hedge fund index data is taken from the Hedge Fund Research website. The HFRI® Indices are broadly constructed indices designed to capture the breadth of hedge fund performance trends across all strategies and regions. I use the Equity Market Neutral Index (HFRIEMNI), which studies the quantitative equity funds and dates back to the beginning of 1990. Mispricing factors data is taken from Stambaugh's website[5]. TED rate data is downloaded from the website of Federal Reserve Bank of St. Louis.

I consider a combined set of 34 stock market anomalies[6] studied in the literature. For each anomaly, I compute the time-series of monthly value-weighted (VW) returns on a long-short self-financed portfolio over the period 1973m1-2017m12. I use the NYSE breakpoints for the anomaly characteristics to sort all stocks traded on the NYSE, AMEX, and NASDAQ. To make sure my results are not driven by micro-cap stocks and other microstructure issues, I exclude stocks with prices below \$5 per share or are in the bottom NYSE size decile. I closely follow Stambaugh

[5] I thank Robert Stambaugh for providing the daily mispricing factors data.

[6] 32 anomalies are following the *the data library (click here)* for Novy-Marx and Velikov (2016), though they only use 23 of them in their analysis of their paper. Another 2 anomalies are from Stambaugh et al. (2012), as 9 out of 11 anomalies they study overlap with Novy-Marx and Velikov (2016)'s. I thank Thummim Cho for generously sharing the detailed anomaly replication procedure and the portfolio data with me to make sure the coding errors are minimized.

et al. (2012), Novy-Marx and Velikov (2016) and Cho (2017) to construct anomalies, and the full set of anomalies is shown in appendix' Table 1.15. I also normalize all anomalies to make sure they have positive abnormal returns so that arbitrageurs would hold a positive position on them.

1.3.2 CoAnomaly Calculation

I first construct value-weighted anomaly portfolios by sorting stocks into deciles based on their anomaly characteristics available at the end of month t-1. Here, I follow the standard procedure in the literature by using the NYSE breakpoints for the sorting and the anomaly portfolio is longing the top decile and shorting the bottom decile[7]. I construct both daily and monthly anomaly portfolios, and I also use the cumulative monthly anomaly returns in a quarter as the quarterly returns for the anomaly. After obtaining all the anomaly portfolios, I then compute partial correlations using daily returns for each anomaly portfolio with respect to the equal-weight of other anomaly portfolios[8]. $CoAnomaly^{LS}$ is the *average* partial correlation for long-short anomaly portfolios. Short-leg CoAnomaly ($CoAnomaly^{S}$) is the *average* partial correlation for the bottom deciles of all anomalies, and long-leg CoAnomaly ($CoAnomaly^{L}$) is the *average* partial correlation for the top deciles of all anomalies. Lou and Polk (2013) use this procedure to proxy the crowdedness of momentum arbitrage activity; however, I use this to measure the correlation risk.

$$CoAnomaly_t^{LS} = \frac{1}{N}\sum_{n=1}^{N} \underbrace{partialCorr_t(ret_n^{LS}, ret_{-n}^{LS}|MktRf)}_{\substack{\text{Average partial correlation} \\ \text{for anomaly } n \\ \text{with respect to all other anomalies } -n}} = \frac{1}{N}\sum_{n=1}^{N} \rho_{n,-n}^{LS},$$

$$CoAnomaly_t^{S} = \frac{1}{N}\sum_{n=1}^{N} partialCorr_t(ret_n^{S}, ret_{-n}^{S}|MktRf) = \frac{1}{N}\sum_{n=1}^{N} \rho_{n,-n}^{S},$$

[7]I adjust the signs of anomaly characteristics so that the outperforming stocks are always on the top deciles (example: small stocks and value stocks).

[8]There is concern about nonsynchronous trading as in Frazzini and Pedersen (2014), but the concern is mostly over the stock level and not on portfolio level, as they will cancel out within diversified portfolios.

$$CoAnomaly_t^L = \frac{1}{N}\sum_{n=1}^{N} partialCorr_t(ret_n^L, ret_{-n}^L | MktRf) = \frac{1}{N}\sum_{n=1}^{N} \rho_{n,-n}^L,$$

where N is the number of total test anomalies, $ret_n^{LS(/S/L)}$ is the daily return of the long-short(/short-leg/long-leg) portfolio for anomaly n, and $ret_{-n}^{LS(/S/L)}$ is the equal-weight daily return of long-short(/short-leg/long-leg) portfolios for all test anomalies apart from anomaly n.[9]

The correlation is partial in the sense that I control for market exposure when computing this correlation to purge away any comovement in anomaly returns induced by the market. Two practical facts can justify this consideration. First, most arbitrageurs (like hedge funds) who are the main traders and exploiters of these anomalies are chasing market neutrality; and in general, the market betas on portfolio level are fairly stable and can be predicted and hedged well.

I use the look-back period for three months, which means the CoAnomaly measure at the end of June is constructed using daily returns in April, May, and June. However, my main results are robust to other specifications, including the one-month look-back window (with stronger effects). The CoAnomaly time series can also be calculated by calculating the summation of the non-diagonal elements in the correlation matrix for all anomalies, which generates the same results in the qualitative sense.

1.3.3 Time Variation and Determinants of CoAnomaly

(Insert Table 1.1)

Table 1.1 reports the summary statistics of the CoAnomaly measure. I find that CoAnomaly for long-short anomaly portfolios is mainly driven by the short leg. CoAnomaly behaves quite differently for the long legs and the short legs, with a negative correlation. In terms of magnitude, the short leg CoAnomaly is larger than long leg CoAnomaly, which could be justified by the following: 1) Apart from size

[9] I also calculate the average pairwise partial correlation of the anomalies by averaging the non-diagonal numbers in the estimated $N \times N$ correlation matrix. I find this average partial correlation, slightly defined differently from CoAnomaly, have a strong correlation with the CoAnomaly, more than 0.97.

anomaly, most anomalies tend to have large firms on the long leg and small firms on the short leg, so a larger price impact should be expected on the short leg than on the long leg, and 2) arbitrageurs have a relatively higher trading presence than other investors on the short legs and they tend to trade all these assets simultaneously. Note that CoAnomaly is not driven by the average correlation between the constituents in the aggregate market by Pollet and Wilson (2010) as they are mildly correlated. Later, I also include the average correlation in the market in my predictive regression and find no effect.

I also check the contemporaneous correlations among different CoAnomaly measures and other market indices in the second half of my sample since some of them are only available from the 1990s. I find that CoAnomaly is related to the market realized variance and the VIX index, which suggests that CoAnomaly is related to the risk in the market, but the comovement is not entirely matched. This pattern can be seen in Figure 1.1 as well. The figure also shows an increasing trend in CoAnomaly, which may be linked to the growth of sophisticated investors in the last few decades. The market excess returns, TED rate, market liquidity level, and equity neutral hedge fund index do not have a particularly strong correlation with the CoAnomaly measure.

I also find that CoAnomaly is highly correlated with the correlation between two mispricing factors as in Stambaugh and Yuan (2016), who argue that most stock market anomalies can be explained by these two mispricing factors. Finally, the short leg CoAnomaly exhibits positive correlation with sentiment index[10] from Baker and Wurgler (2006), which is consistent with Stambaugh et al. (2012) and suggests that the high sentiment from overoptimistic retail investors may cause excess correlation on overpriced assets.

Predicting CoAnomaly

$$\text{CoAnomaly}_t = a + b \times \text{CoAnomaly}_{t-1} + \sum_p m_p \times \text{Controls}_{p,t-1} + t \times \text{Trend} + e_t. \quad (1.1)$$

(Insert Table 1.2)

[10]Thank Jeffrey Wurgler for sharing the sentiment index data on his website.

In panel A of Table 1.2, I conduct a predictive analysis of CoAnomaly to find the *potential* determinants of the time-varying CoAnomaly. I find that all CoAnomaly measures are fairly persistent and have a positive trend (except for the long leg CoAnomaly). I run extra regressions of long-short CoAnomaly on various variables with a one-quarter lag. The investor sentiment appears to be the only consistently strong predictor of CoAnomaly[11]. As the standard deviations of other non-CoAnomaly regressors are normalized to 1, a one-standard-deviation increase in market sentiment predicts around a 3 percent increase in CoAnomaly in the economic magnitude. Apart from sentiment, none of the coefficients for other predictors shows persistent significance, which includes market excess return, average liquidity, TED rate, and Equity Market Neutral Index (HFRIEMNI)[12].

One final observation is that the trend of CoAnomaly does not go beyond the linear trend as the F-test for R-squared yields a value of 1.59, which indicates that the increase in R-squared is not significant on adding more polynomials, as shown by specification [4] in panel A of Table 1.2. From this point, unless explicitly addressed, when I talk about CoAnomaly, I refer to the long-short CoAnomaly[13].

1.3.4 Risk Measure: Variance or Correlation?

Predicting the Aggregate Anomaly Variance The variance of a diversified portfolio, which serves as the traditional measure of the risk, is determined by both the average variance of the constituent assets and the average correlation among them[14]. I focus on a naïve portfolio by investing in these anomalies with equal amounts, and I term the return as *Equal-weighted Anomaly Return (E.A.R.)*, which is the simple equal-weight mean return of 34 stock market anomalies. DeMiguel

[11]However, I do not include the investor sentiment in the VAR for later analysis as Jeffery Wurgler argues in his website: *Do not use these series for measuring changes in sentiment (e.g. sentiment(t)-sentiment(t-1)) due to lag structures, among other considerations, as these are low-frequency levels indicators.*

[12]From a crowdedness of arbitrage capital perspective, this is not surprising considering that while high arbitrage capital (lower TED rate, larger higher market, and hedge fund return, lower volatility or higher liquidity) would induce more capital allocated in these anomalies, it is also possible that arbitrageurs trade more frequently when they are facing capital constraints. Both of these mechanisms may increase the comovement among assets. Given these effects are mixed, I do not overinterpret the signs at the moment.

[13]Empirical results remain qualitatively the same if I use the short leg CoAnomaly.

[14]Consider a simple case: a portfolio with N symmetric assets, when N is large: $\sigma_p^2 = N \times (\frac{1}{N})^2\sigma^2 + 2 \times \frac{N(N-1)}{2}(\frac{1}{N})(\frac{1}{N})\rho\sigma\sigma \approx \rho\sigma^2$

et al. (2007) show that many in-sample optimized techniques fail to beat the naïve 1/N rule out of sample in terms of Sharpe ratio, certainty-equivalent return, or turnover. Later, I also extend my analysis to the in-sample mean-variance efficient portfolio for robustness check.

In the first part of panel B of Table 1.2, I report the results of regressing the realized aggregate variance of the equal-weighted anomaly returns (E.A.R.) on the contemporaneous CoAnomaly and the average realized variance of single anomalies. The aggregate variance of the Equal-weighted Anomaly Returns (E.A.R.) is measured as the variance of daily returns within a given quarter. Average realized variance is equally averaging the realized daily variances for the 34 stock market anomalies in the same quarter. The results show that both CoAnomaly and the average variance contribute to the variance of the E.A.R., and these two components capture almost all of the time-series variation in the E.A.R. aggregate variance (as the R-Squared reaches 80%).

The remaining parts of panel B of Table 1.2 conduct predictive regression for the aggregate variance of E.A.R. by using CoAnomaly, average variance, and aggregate variance of E.A.R with lags up to four quarters. However, for average variance and aggregate variance, the predictability quickly dies out after 2 quarters as the R-squared dropped from around 50% for one-quarter lag to below 10% for above two-quarter lags. On the contrary, CoAnomaly robustly and persistently explains 20% to 10% of the aggregate variance variation, for lags from one to four quarters. These results (both the coefficients and the R-squared) clearly show that CoAnomaly is the only component that maintains a strong predictive power across the one-year period, and they also support the choice of CoAnomaly as a risk measure.

Economic Intuition Pollet and Wilson (2010) presents a stylized model to show that the average correlation among assets, but not the average variance, is positively related to the risk premium. They show that the risk premium is given by:

$$\mathbb{E}_t[r_{s,t+1}] - r_{f,t+1} + \frac{\overline{\rho}_t \overline{\sigma}_t^2}{2} = \frac{\gamma}{\beta_t(1-\theta_t)} \overline{\rho}_t \overline{\sigma}_t^2 - \frac{\gamma}{\beta_t(1-\theta_t)} \theta_t \overline{\sigma}_t^2, \tag{1.2}$$

where $r_{s,t+1}$ is the return on the stock market, $r_{f,t+1}$ is the risk-free rate, $\overline{\rho}_t$ and $\overline{\sigma}_t^2$ are the average correlation and the average variance of single stocks, β_t is the beta of stock market on the aggregate wealth portfolio, θ_t is the proportion of the stock market risk component to the total risk for a single stock.

As shown in the equation, the relationship between risk premium and average variance is not clear; however, the relationship between risk premium and the average correlation is positive because the stock market is part of the aggregate wealth portfolio. The reason behind this is that if the changes in the stock market variance are orthogonal to the risk in aggregate wealth portfolio, then such changes in stock market variance should be offset by changes in the covariance of the stock market with the rest of the aggregate wealth portfolio, holding the risk of aggregate wealth portfolio constant.

From another point of view, if single assets share common components from the aggregate portfolio, the increase in volatility of this common component will, first, drive up the volatility of single assets, and second and more importantly, induce stronger comovement among these single assets. When the aggregate portfolio cannot be measured perfectly, the volatility of an alternative pseudo-aggregate portfolio can be a bad proxy for the aggregate risk. However, the correlation effect between single assets remains robust.

In the market-neutral investment setting, these stock market anomalies constitute only a subset of the whole investment universe of the sophisticated investors, which is a perfect scenario that fits Roll (1977)'s critique. As argued in Pollet and Wilson (2010), if the Roll (1977)'s critique is important, the variance may be weakly correlated with the aggregate risk and subsequent excess returns. Following the same logic, the variance in these anomalies may provide contaminated information about the risk of their aggregate portfolio.

Why not covariance? Standard portfolio theory states that the risk of a single asset evaluated with respect to a diversified portfolio is measured by the covariance between the asset and the portfolio. However, in the case of stock market anomalies, I argue that CoAnomaly is a better measure than the covariance to proxy the risk: to access the covariance, a benchmark portfolio is required, which is unrealistic in the case of sophisticated institutional investors. Unlike the standard macrofinance models normally assuming that long-term investors hold the aggregate market, the investment universe of institutional investors, like hedge funds, go way beyond the stock market, to fixed-income, derivatives, and even to real estates and antiques. On the other hand, even if the composition of the portfolio is identified, the exact weight on each asset (strategy) is still unknown. Meanwhile, leverage is widely used by these professional institutional investors and they manage their leverage ratio across time and strategies. This effect also obfuscates the estimation of weights in different strategies/anomalies. Therefore, in the case of sophisticated arbitrageurs, a benchmark portfolio like the market portfolio cannot be observed, hence the co-variance measure lacks a clear definition to measure the risk.

However, if single assets share common components from the aggregate portfolio, the increase in volatility of this common component will induce stronger comovement among these single assets. This effect on comovement also justifies my choice of using the average correlation to calculate the CoAnomaly measure.

1.4 Time-Series Predictability of Anomaly Returns

Suppose an investor invests in a bundle of risky assets. *Ceteris paribus*, the increase in the average return correlation among all these assets will make the optimal portfolio riskier by increasing the variance, and the rational investor will ask for a higher return on holding these assets as compensation. Given the sophisticated nature of arbitrageurs, the risk-return trade-off is expected to hold in the anomaly investment setting.

1.4.1 Predictive Regression

I first conduct the following predictive regression analysis to see whether CoAnomaly can predict future anomaly returns:

$$\begin{aligned} \text{E.A.R.}_{t+1} = \quad & a + b \times \text{CoAnomaly}_t + t \times \text{Trend} \\ & + \sum_p m_p \times \text{Other.Predictors}_{p,t} \quad (+\sum_j \times \beta_j Benchmark.Factors_{t+1}) + e_{t+1}. \end{aligned} \tag{1.3}$$

(Insert Table 1.3)

In panel A of Table 1.3, I regress the equal-weighted anomaly Returns (E.A.R.) in the *next* quarter on the observables in *current* quarters using non-overlapping data. I also include a trend variable in all regressions, which turns out to carry a significant negative coefficient, and this is consistent with recent findings of the return decay in these anomalies (see McLean and Pontiff (2016)). The standard deviations of all predictive regressors are normalized to 1. Column (1) shows that CoAnomaly is a strong predictor of the E.A.R. and it alone with the trend can explain 8.1 percent of the time-series variation of the E.A.R.. Columns (2) and (3) show that the predictive powers of the average variance and aggregate variance of anomalies are negligible, consistent with recent findings of the weak or negative predictability of variance measures (see Barroso and Maio (2018)). The economic magnitude of CoAnomaly predictability on the E.A.R. is more than 70 basis points per quarter given the one standard deviation change in CoAnomaly. Columns (4) and (5) include both CoAnomaly, the average variance, and the aggregate variance of E.A.R., and they show that only the average correlation predicts future returns. Next, I control for the anomaly value spread and the sentiment. Anomaly value spread is the average value spread for all anomalies, which is the difference in weighted average log book-to-market ratio between the long legs and short legs. As Cohen et al. (2003) argue the value spread must predict future returns, profitability and/or the persistence of valuation levels following the firm-level decomposition by Vuolteenaho (2002). Stambaugh et al. (2012) find that anomaly returns are higher following high investor sentiment constructed by Baker and Wurgler (2006) and I find that the sentiment has a strong predictive power. Nevertheless, the predictive power of my CoAnomaly survives with statistical significance, albeit with a smaller scale. In the last specification, the predictability remains strong after controlling other potential predictors, including TED spread (TED), market excess return, market average correlation, and E.A.R..

In panel B of Table 1.3, I report the benchmark-factor-adjusted regression results as a clean-up test. I want to make sure that the time-variation in average anomaly returns that CoAnomaly predicts is not driven by the benchmark factors as many papers[15] have argued that there are risk factors behind these factors. I can also tease

[15]See Fama and French (1992), Campbell and Vuolteenaho (2004), Zhang (2005), Lettau and Wachter (2007), Garleanu et al. (2012) and Campbell et al. (2017) among the others.

out the part of variation associated without any 'premium' like size factor, in which alpha-seeking arbitrageurs will have less interest. Here I use Carhart (1997) four-factor model as a benchmark, and my results are robust to different specifications, including a single market factor, Fama-French three-factor model, and Fama-French five-factor model. Once the returns are adjusted with contemporaneous benchmark factors, CoAnomaly shows much stronger predictability. Note that the E.A.R. loadings on benchmark factors are consistent with literature: in general, (overpriced) stocks in the short legs tend to be small stocks with higher market beta, which results in negative loadings on both market factor and size factor; and anomalies tend to load on value and momentum as Asness et al. (2013) point out.

In nontabulated results, I also show that my results are robust, albeit with smaller magnitudes, when including two contemporaneous mispricing factors in Stambaugh and Yuan (2016). Note that their study is focusing on explaining the cross-sectional dispersion of anomaly returns by measuring their loadings on the two mispricing factors; however, my result is focusing on the time-variation of anomaly returns. The smaller magnitudes after controlling two mispricing factors, together with the fact that E.A.R. is strongly loading on these two factors, suggest that the time-varying returns of the two mispricing factors can also be explained, at least partially, by CoAnomaly.

Anomaly Set Specification To make sure my results are robust to anomaly specifications, I conduct the same analysis for two different sets of anomalies: 23 anomalies studied in Novy-Marx and Velikov (2016) (NMV) and 11 anomalies studied in Stambaugh et al. (2012) (SYY) separately. CoAnomaly, E.A.R., and other anomaly-relevant measures are constructed with only these two sets, respectively. The results are reported in the appendix, and I find that the predictability of CoAnomaly is robust. However, I also find that the market sentiment has a relatively stronger predictive power in the original anomaly set studied in Stambaugh et al. (2012).

Sample Periods I check if my results are robust to 1) extending the main sample period back to 1965, or 2) excluding the most recent financial crisis in 2008. The market turmoil in 2008 has a large impact on broad financial markets and asset prices. As for the stock market anomalies, there has been some research on the different behaviors during these periods: Daniel and Moskowitz (2016) show that momentum strategy lost close to 50 percent following the 2008 turmoil.

Separate Long Legs and Short Legs Moreover, I show that CoAnomaly predicts future returns on anomalies for both their long legs and short legs, which is more aligned to the risk-return trade-off story. The results can be found in the appendix (see Table 1.17). My results are different from Stambaugh et al. (2012), as they find that their predictability is concentrated on the short legs and argue that this is consistent with a combination of short-sale impediments and market-wide sentiment.

Different Horizons The predictive power of CoAnomaly is also robust for shorter or longer horizons (see Table 1.18 in the appendix). However, the regression result for 6-month E.A.R. is the strongest in terms of both coefficients and adjusted R-squared, which is not surprising considering there is more noise in short run, and the predictability will die out in the long run owing to the time-varying nature of CoAnomaly. This also helps to partially alleviate the concern of spurious regression raised by Ferson et al. (2003), since they argue that if the expected returns are persistent, there is a risk of finding a spurious relation between the return and an independent, highly autocorrelated lagged variable. Here, I show that the predictability is strongest for 6 months, and the CoAnomaly measure has an autocorrelation coefficient around 0.5 for 3 months, so in the 6-month window, the persistence level of CoAnomaly is relatively low, which violates the 'highly autocorrelated' condition of the spurious regression concern.

Mean-Variance Efficient Portfolio I also check the predictability for a mean-variance efficient (MVE) portfolio consisting of these 34 anomalies, though I do not focus my analysis on this ex-post efficient portfolio, which neither researchers nor arbitrageurs know ex-ante. I compute the MVE portfolio weights by maximizing the in-sample Sharpe ratio with respect to a zero-beta rate equal to 0, based on the in-sample average returns and covariance matrix. Sophisticated arbitrageurs will hold this optimal portfolio if anomalies arise from mispricing. I expect the pattern of risk-return trade-off to show up for the MVE portfolio. Consistent with my conjecture, Table 1.19 in the appendix shows that the predictability for the MVE portfolio is similar to that for the E.A.R..

1.4.2 Sorting in Time-series

(Insert Table 1.4)

To ease up the interpretation, I also provide evidence by sorting all months in the time series. As shown in Panel A of Table 1.4, all months are sorted into three groups based on CoAnomaly in month t. All sorting variables are detrended[16] and the time-series sorting uses 30% and 70% as breakpoints.

First, I show that higher past CoAnomaly does predict higher future CoAnomaly. Note that the persistence in CoAnomaly is a necessary condition for the predictability under the risk-return trade-off mechanism. I then check the returns of equal weighting the long-short returns of all anomalies from the next month (t+1) to half of a year (t+6). There is a monotonic and persistent pattern across groups, as high CoAnomaly months are followed by high average returns on anomalies. On average, the difference in returns between following a high CoAnomaly and following a low CoAnomaly (Diff 3-1) is 120 basis points in the following 6 months, which is economically large and statistically significant.

I split my full sample periods, from 1973Q1 to 2017Q4, into two halves: 1) pre-1994 and from 1973Q1 to 1993Q4, and 2) post-1994 and from 1994Q1 to 2017Q4. As I turn to the second half of my sample (Panel B1 of Table 1.4), the predictability is much stronger in both economic and statistical senses, as illustrated graphically in Figure 1.2. This fact could be partially justified by the explosion of finding stock market anomalies and the emergence of sophisticated institutional investors since the early 90s. This split is also robust to different cutting points.

Double-sorting I further check whether this pattern will hold under different market conditions for anomaly investors. I expect the asset pricing pattern from the risk-return trade-off will be stronger when the investors are more risk-averse.

[16]Detrending means that I am not biasing my analysis to the front or tail of my sample periods. The results are qualitatively the same if I use undetrended time-series.

Given the fact that hedge funds are the main players in trading these anomalies, I use the quantitative equity hedge fund returns (later HF Ret.) to proxy the risk-aversion of these arbitrageurs. Among Hedge Fund Research Indices, I choose the HFRI Equity Market Neutral Index, HFRIEMNI[17], which is the average return of all market-neutral quantitative equity funds in their database, to proxy the shocks to these arbitrageurs. Therefore, I first sort all months based on the hedge fund returns in month t, and then within each group, I sort on the CoAnomaly level. As shown in Panel B1 of Table 1.4, the two sorting variables, column HF Ret. t, and column CoAnomaly t do not show any particular relationship, so if I conduct the double sorting independently, the results remain unchanged qualitatively.

Panel B2 of Table 1.4 shows the results of the double sorting. The predictability is much stronger and more significant in the distress periods of hedge funds especially in the short-run. Although the evidence is by no means definitive, it is consistent with the possible explanation that due to various reasons, including the withdrawal of capital by investors, hedge fund managers show higher risk aversion after poor returns so the risk-return trade-off pattern in asset prices is stronger.

The main advantage of using the HFRIEMNI is that it is a direct measure of the shocks to the arbitrageurs who are mainly trading stock market anomalies, which is better than using the average returns on all anomalies as a proxy to the shocks to the arbitrageurs because I cannot assume that arbitrageurs are betting these anomalies consistently across time. There is a large body of literature documenting the timing ability of different anomalies (e.g. Cohen et al. (2003) for timing value, Lou and Polk (2013) and Barroso and Santa-Clara (2015) for timing momentum, Moreira and Muir (2017) for timing an extensive set of factors based on their realized volatility). Barroso et al. (2017) directly test the behavior of institutional investors with 13F institutional holdings data, and find that these investors actually decrease their loading on momentum before momentum crash, which rejects the idea that

[17] On their website, they state that '*Equity Market Neutral strategies employ sophisticated quantitative techniques of analyzing price data to ascertain information about future price movement and relationships between securities, select securities for* ***purchase*** *and* ***sale****. These can include both* ***Factor-based*** *and* ***Statistical Arbitrage/Trading strategies****. Factor-based investment strategies include strategies in which the investment thesis is predicated on the systematic analysis of common relationships between securities. In many but not all cases, portfolios are constructed to be neutral to one or multiple variables, such as broader equity markets in dollar or beta terms, and* ***leverage*** *is frequently employed to enhance the return profile of the positions identified.*'

momentum crashes relate to institutional crowding. However, in results not shown here, I also did the same test using the equal-weighted return on all anomalies as a proxy of shocks to arbitrageurs and find similar patterns, but with less statistical significance.

Control for Benchmark Factors Consistent with predictive regression, the results are stronger if I control for contemporaneous benchmark factors (a single market factor and Carhart (1997) 4-factor model), as shown in Table 1.5. In non-tabulated results, I find that my results are robust if I use post-1990 as the second half of my sample, as the hedge fund index starts from 1990.

(Insert Table 1.5)

Evidence from Daily Returns and Higher Moments I report the summary statistics of the daily raw returns of E.A.R.. The average daily raw returns show similar patterns on the monthly level. The standard deviation of the E.A.R. is also high after high CoAnomaly periods. More importantly, I find that the E.A.R. returns tend to left-skewed more strongly following high CoAnomaly periods. In the second sample period, the daily raw returns of E.A.R. show higher standard deviations as well as stronger crash risks.

These empirical patterns clearly show that though the average (first moment) of E.A.R. returns are higher after high CoAnomaly period, the risks (higher moments) in trading anomalies are also higher. This pattern echoes the '*Unwind Hypothesis*' analyzed in Khandani and Lo (2011), in which they observed large correlation and crash in returns of long-short equity strategies during the 'quant meltdown' in August 2007.

(Insert Table 1.6)

Other Robustness Checks In the appendix, I explore other drivers that may affect my results: principal component analysis of these anomalies, small and high idiosyncratic stocks driving results, too many anomalies and lack of dimension, different CoAnomaly calculation window, and anomalies sharing same stocks. None of these have a significant impact on my results.

1.4.3 Out-of-Sample Tests and Econometric Issues

Here I conduct out-of-sample (OOS) tests and also explore two econometric issues that might undermine the validity of my results, and I find evidence that the predictability of CoAnomaly is robust.

Out-of-Sample Tests In the out-of-sample tests, I compare the predictive power between the historical sample mean and the predictive regressions. For both methods, the forcast for quarter t only uses information up to quarter $t-1$. The out-of-sample statistics are standard in the literature[18] following McCracken (2007), Welch and Goyal (2008) and Huang (2015).

One predictive regression of quarterly E.A.R. returns on CoAnomaly produces an out-of-sample R-square of 2.16%. This contrasts with the results of other predictors for E.A.R. as well as the usual negative OOS R-squares of similar predictive regressions for the market excess return.

(Insert Table 1.7)

Generated Regressors The standard errors in the above predictive regression require a caveat because the regressor CoAnomaly is generated from the daily anomaly returns. To make sure this extra layer of noise does not undermine my main result, I conduct a double-layer block bootstrap based on Djogbenou et al. (2015) and the detailed process is described in the appendix. Table 1.8 reports the t-stats with Newey and West (1987) correction and the t-stats with bootstrap standard errors. The differences between these two are marginal and my results remain significant robust.

(Insert Table 1.8)

This is not surprising since effectively, the CoAnomaly measure is calculated quite precisely with high-frequency data used in the process. In the appendix, I show that the standard errors in the CoAnomaly measure are relatively small compared to its own time-variation.

[18] $R^2 = 1 - \frac{MSE_A}{MSE_N}$, $\overline{R^2} = 1-(1-R^2)\frac{T-k-1}{T-k-p-1}$, $\Delta RMSE = \sqrt{MSE_N} - \sqrt{MSE_A}$ and $MSE-F = (T-k)\frac{MSE_N - MSE_A}{MSE_A}$, where MSE_A and MSE_N are the mean square errors of the predictive regression and the historical sample mean respectively, T is the sample periods, k is the trading periods, and p is the number of predictors.

Biased Estimators As Stambaugh (1999) points out: when a rate of return is regressed on a lagged stochastic regressor, the OLS estimator will be biased if the innovations of the dependent variables and the innovations of the regressors are correlated. A simple illustration of the Stambaugh (1999) bias (omitting means):

$$\begin{aligned} r_{t+1} &= b \times pred_t + e_{t+1} \\ pred_{t+1} &= \phi \times pred_t + u_{t+1}. \end{aligned} \tag{1.4}$$

In small-sample AR(1) regression, the estimate $\hat{\phi}$ tends to be downward biased. If $Cov(e_{t+1}, u_{t+1}) < 0$, then the coefficient in the predictive regression will be inflated, as

$$\mathbb{E}(\hat{b} - b) = \frac{Cov(e_{t+1}, u_{t+1})}{Var(u_{t+1})}\mathbb{E}(\hat{\phi} - \phi).$$

In my setting, there can be a potential problem since CoAnomaly is proxying the risk. If the latent risk gets higher, the innovation in the CoAnomaly measure will increase, and in the meantime, the assets will suffer a contemporaneous bad shock as the future discount rate increases due to higher risk. This mechanism may lead to a negative relationship between e_{t+1} and u_{t+1}.

To make sure my results do not suffer from this bias, I estimate a restricted vector autoregression (VAR(1)) for the E.A.R. and CoAnomaly. I am interested in the covariance between two shocks as well as the variance of the CoAnomaly shock.

$$\begin{aligned} E.A.R._{t+1} &= a + b \times CoAnomaly_t + e_{t+1} \\ CoAnomaly_{t+1} &= c + d \times CoAnomaly_t + u_{t+1}. \end{aligned} \tag{1.5}$$

By estimating a VAR with only E.A.R. and CoAnomaly (both are detrended first to avoid complication, and CoAnomaly is also normalized that standard deviation equals to 1), I find that the estimated E.A.R. shocks e_{t+1} and CoAnomaly shocks u_{t+1} are almost uncorrelated, with the correlation coefficient being equal 0.08. Moreover, I also follow Baker et al. (2006) and conduct a Monte Carlo simulation under the null, i.e., there is no predictability ($b = 0$). In the last column of panel B in Table 1.9, I show that across several specifications, the probability that the simulated coefficient, b, is higher than my estimated coefficient is always less than 0.01%. This suggests that the predictability of my finding is extremely unlikely to be driven by the Stambaugh (1999) bias.

(Insert Table 1.9)

1.5 Price of Risk in Cross-Sectional Asset Prices

CoAnomaly positively predicts both higher volatility and higher returns on these stock market anomalies; in other words, it forecasts changes in the distribution of future return (investment opportunity). If there are some assets that comove with CoAnomaly, I would expect sophisticated investors to use them as hedges.

Here I follow the procedure of Fama and MacBeth (1973) and Adrian et al. (2014) to conduct a standard asset pricing test of whether CoAnomaly is priced in the market. I use the simple detrended-AR(1) innovation e_t in the CoAnomaly measure (the error terms in the specification (1) in Panel A of Table 1.2) as the shock here for simplicity although my results are robust to other alternative specifications.

$$\text{CoAnomaly}_t = a + t \times \text{Trend} + b \times \text{CoAnomaly}_{t-1} + \epsilon_t \tag{1.6}$$

In the first step, based on different pricing models that I check, I regress the excess returns on the different factors (including the innovations of CoAnomaly ϵ_t) to get their risk exposures or betas:

$$\mathrm{R}^e_{i,t} = c_i + \sum_j \beta_{i,j} f_j + e_{i,t}, \qquad t = 1, 2, ..., T \text{ for all } i.$$

Then I run a cross-sectional regression of time-series average excess returns, $\mathbb{E}[R^e_{i,t}]$, on risk factor exposures $\widehat{\beta_i}$ estimated from the last step:

$$\mathbb{E}[R^e_i] = \lambda_0 + \sum_j \widehat{\beta_{i,j}} \lambda_j + \xi_i.$$

By doing this, the risk premia of different factors λ_j as well as the zero-beta rate λ_0 are calculated. Here, I assume that all assets have constant betas on different factors.

1.5.1 Price of Risk in Anomaly Portfolios

I first use the 34 long-short stock market anomalies as test portfolios. As they are traded by anomaly arbitrageurs, I expect CoAnomaly to be priced among them.

(Insert Table 1.10)

Table 1.10 shows that CoAnomaly carries a significant and negative price of risk. The negative sign of CoAnomaly risk means that investors are willing to accept a lower average return if the asset positively comoves with CoAnomaly. In other words, assets with a positive loading on CoAnomaly will give a lower average return because these assets tend to do well in high CoAnomaly periods, providing hedges against the CoAnomaly risk. This is consistent with the empirical fact that CoAnomaly is a strong and persistent predictor of future aggregate variance for these anomalies: so the increase in CoAnomaly implies higher aggregate volatility, hence worse investment opportunity. This result echoes the finding in Driessen et al. (2009). As both returns and CoAnomaly are measured in percentage, the economic magnitude is that, given other things equal, one unit increase in the CoAnomaly beta will lower the quarterly return of the asset by around 5%.

In the last column of Table 1.10, I also show the adjusted R-squared in the cross-sectional regression, which represents how much the cross-sectional dispersion of average returns among these anomaly portfolios can be explained by different loadings on the factors. The single market factor only explains 3%, and once it is augmented with CoAnomaly, they can explain 11.3% of the return dispersion. This is a good performance given two facts: first, the test portfolios are 34 stock market *anomalies* which are famous for being difficult to price, and second, contemporaneous return benchmarks only explain the dispersion with a similar proportion (7.8% for Fama and French (1996) three-factor model and 13.8% for Carhart (1997) four-factor model in nontabulated result).

I also control for two intermediary asset pricing factors: leverage of securities broker-dealers from Adrian et al. (2014) and equity capital ratio of primary dealers He et al. (2017). Both of these studies find a positive price of risk of the shocks to financial intermediaries. I find that CoAnomaly maintains significant pricing power and these two intermediary asset pricing factors do not show a large (though positive) risk premium.

Note that there are large intercepts left for these anomaly portfolios. Consequently, the standard Chi-square test of pricing errors gets rejected overwhelmingly. However, pricing all assets is *not* the main purpose of the exercise here. All evidence shown here is to support the fact that the CoAnomaly risk gets priced among anomalies themselves. Moreover, it also helps explain the cross-sectional return dispersions partially, as shown in the top two figures in Figure 1.3.

(Insert Figure 1.3)

1.5.2 Price of Risk in a Standard Set of Portfolios

I also use the standard set of test portfolios, which are the commonly studied equity and government bond portfolios: 25 Fama-French size-value portfolios, 10 momentum portfolios, 5 industry portfolios, and 6 treasury bond portfolios sorted by maturity[19]. Five industry portfolios are included to make sure that my results are not driven by the strong factor structure within size, value and momentum portfolios, as suggested by Lewellen et al. (2010).

(Insert Table 1.11)

The point estimates of risk premia are different across specifications, but they stay in a stable range and are slightly larger than the estimates in anomalies: On average, one unit of loading on CoAnomaly generates -8% percent risk premium per quarter. In nontabulated results, I find that CoAnomaly risk premium gets subsumed to zero if I include all size, value and momentum factors together, which is not surprising since most test portfolios are based on the characteristics behind these factors and hence have a strong factor structure that can be explained by '*themselves*'[20].

[19] Portfolio returns are downloaded from French's website. Thank him for providing the data.

[20] Cochrane (2009) states this point (page 126): '*Thus, it is probably not a good idea to evaluate economically interesting models with statistical horse races against models that use portfolio returns as factors. Economically interesting models, even if true and perfectly measured, will just equal the performance of their own factor-mimicking portfolios, even in large samples. Add any measurement error, and the economic model will underperform its own factor-mimicking portfolios. And both models will always lose in sample against ad hoc factor models that find nearly ex-post efficient portfolios.*'

In panel B of Table 1.11, I exclude the 6 bond portfolios, and the risk premium of CoAnomaly remains large and significant. The results also show that market factor explains little cross-sectional variation of portfolio returns (and also comes with a wrong sign of risk premium). However, on augmenting the CAPM with CoAnomaly, I find that close to one-fourth of the total cross-sectional return dispersion can be explained. I plot these results in the bottom figures in Figure 1.3.

In non-tabulated results, I find that the hedging portfolios for CoAnomaly risks, i.e. portfolios with high CoAnomaly betas, tend to be large stocks, loser stocks, and long-term bond portfolios.

These results provide strong evidence supporting that CoAnomaly is priced in the market, and the most well-known stock market anomalies (size, value, and momentum) can be explained, at least partially, by the different loadings on the CoAnomaly risk. On the other hand, these results also suggest that arbitrageurs play important roles in setting asset prices even for the standard assets, given the fact that the priced CoAnomaly risk is a matter of concern to these arbitrageurs.

1.5.3 CoAnomaly Beta Sorted Portfolios

I use the entire cross-section of CRSP stock returns to construct portfolios based on real-time CoAnomaly innovation betas. I find the post-ranking CoAnomaly betas is consistent with estimated and the return dispersion in average returns line up well with the betas after controlling other benchmark factors.

Constructing CoAnomaly Beta Sorted Portfolios I follow Fama and French (1993) and Adrian et al. (2014) and form portfolios based on pre-ranking CoAnomaly betas. The betas are computed by, at the end of every month, regressing past monthly returns on the monthly CoAnomaly innovation, which follows the same specification as in Equation 1.6 but on the monthly frequency. I use a 3-year rolling window and require stocks to have more than 24 monthly returns. Specifically, at each month, I sort the universe of Amex, NASDAQ, and NYSE stocks from CRSP into quintiles based on their estimated CoAnomaly betas over the last 3 years (36 months) with NYSE breakpoints. This CoAnomaly factor is constructed in real-time, so this procedure is tradable for investors. To avoid market microstructure issues, I drop stocks smaller than NYSE bottom 10% cutoff and stocks with a share price under \$5. I form value-weighted portfolios based on the CoAnomaly beta group and also construct a long-short portfolio by taking the return difference between the top quintile and the bottom quintile.

Betas and Adjusted Returns Panel A of Table 1.12 reports the post-ranking betas of the quintile portfolios. The betas on different factors are estimated by running a regression on all the factors. The post-ranking liquidity betas increase monotonically across quintiles, consistent with the objective of the sorting procedure. The '5-1' spread, which goes long stocks with high CoAnomaly beta and short stocks with low CoAnomaly beta, has a CoAnomaly beta of 0.030, with a t-statistic of 1.89. I also find this long-short portfolio tilts towards large stocks, growth stocks, and winner stocks. Average market capitalization is also reported and the low CoAnomaly beta stocks are generally smaller than others.

Panel B of Table 1.12 reports the return dispersion among these CoAnomaly beta sorted portfolios. In general, the spread between the top quintile and the bottom quintile is negative. This negative return spread shows significance after purging away the effects of Fama-French five factors and momentum factor, which yields 13 basis points per month. These results are consistent with the negative price of risk in CoAnomaly that I find earlier, but the lack of strong statistical significance also suggests that the portfolio construction is still fairly noisy[21].

(Insert Table 1.12)

1.6 Drivers of CoAnomaly

Up to this point, I am treating CoAnomaly as an exogenous risk in the sense that I do not take any stand on what drives the time-variation of CoAnomaly. It is worth taking a look at the potential driver of the correlations among anomalies. Why do assets comove with each other in the first place? From the perspective

[21] Adrian et al. (2014) argue that: *It is well known that sorting on characteristics is less noisy than sorting on covariances, making factors formed on covariance sorts less equipped to capture the underlying discount factor variation. In other words, even if past covariances are perfectly measured, they may not measure future conditional covariances well, and in particular characteristics often give a better proxy.*

of the return decomposition identity, assets will comove due to the comovement in their cash flow news, in their discount-rate news, or the interaction effects. As the time-variation of CoAnomaly is large in the short-term (quarterly autocorrelation coefficient around 0.5), it is unlikely the cash-flow component will be the main driver. Hence, I consider the channel of discount rate commonality across anomalies. As trading different assets together may induce comovement, it is natural to look at the trading behaviors from both retail investors and institutional investors. In Panel A of Table 1.2, I have already presented evidence that investor sentiment is the only strong predictor of CoAnomaly on the aggregate level: CoAnomaly will increase around 3% after a standard deviation increase in sentiment. This suggests that retail investors contribute to the time-variation of CoAnomaly. Here, I will explore further on anomaly level and see if I can find evidence that links CoAnomaly to arbitrage activities of institutional investors.

1.6.1 Arbitraging Capital

Are the arbitrageurs driving the correlations among these anomalies? My results suggest that the answer should be *yes but not entirely*, as I find mixed evidence for this question: following McLean and Pontiff (2016), I find that partial correlations with *publicly-known* strategies do increase after the publication of relevant research papers; using short interests as a proxy to the arbitrage capital allocated to different anomalies, I do not find that this proxy predicts the future change in the partial correlations. Here, I delve into the anomaly level and create a panel with returns, short interests, and partial correlations with other strategies for each anomaly.

Partial Correlation Change around Publication

(Insert Table 1.13)

$$\text{Part.Corr.}_{i,t}^{mp} = b_1 \times \text{Post-Publication Dummy} \quad (+b_2 \times \text{Trend}) \quad + a_i + e_{i,t}. \quad (1.7)$$

I use the publication year as a structural change and find that the partial correlation for each anomaly with *then-public* strategies increases after the publication of relevant academic research. Here, I only include anomalies with publication year no earlier than 1990 (27 anomalies), to make sure that I have enough existing anomalies to calculate the partial correlation. I conduct a regression approach to illustrate this point as shown in Panel A of Table 1.13. On average, the partial correlation is 13.2 percent higher (6.3 percent if controlling the trend) after the publication of academic papers, and this result echoes the findings in McLean and Pontiff (2016), where they find that once a predictor is published, its returns have a stronger beta with respect to other post-publication predictor portfolios.

Figure 1.4 shows graphically that, on average, the partial correlation increases around 10 percent on average across 27 anomalies, in a 10-year window around the publication date. Another interesting pattern worth noting is that the increasing trend in partial correlation emerges two years before the publication year, which coincides with the length of the circulating-submitting-revising period of academic publication.

(Insert Figure 1.4)

Short Interest to Proxy Arbitrage Capital To proxy the amount of arbitrage capital level, I use two specifications for the short interests: the plain average level on the short leg for each anomaly; and the coefficient on the dummy of the short leg in a regression following Hanson and Sunderam (2013). As argued by Hanson and Sunderam (2013), short interest is an excellent setting to empirically study the arbitrage capital because: first, short-sellers are mainly alpha-seeking sophisticated investors due to the complexity and the associated costs; second, the institutional holdings data on the long legs is contaminated by passive indexing. I run predictive panel regressions by regressing the partial correlations at on the lag of partial correlations, the anomaly return, and (the change in) the short interest of each anomaly:

$$\text{Part.Corr.}_{i,t}^{mp} = b_1 \times \text{Part.Corr.}_{i,t-1}^{mp} + b_2 \times \text{Ret}_{i,t-1} + b_3 \times (\Delta)\text{Short.Interest}_{i,t-1} + a_i + d_t + e_{i,t}. \tag{1.8}$$

My results, as shown in Panel B of Table 1.13, if any, do not find any predictive power from the more arbitrage capital to the higher the partial correlation under both specifications. If I acknowledge the short interest as a valid proxy for arbitrage capital, there should be other more important drivers of the time-varying correlations.

These two results together, suggest the following: the discovery of these anomalies do attract arbitrageurs to profit from them, consequently increasing the level of partial correlations among those anomalies; however, this tends to be a level shock on the correlations, since I do not find the anomaly-level short interests (as a proxy to the arbitrage capital) predicts the time-variation of the partial correlations.

1.6.2 Financial Intermediary and Endogenous Risk

$$\text{Shock.CoAnomaly}_t = b_1 \times \text{Intermediary-Level}_t + b_2 \times \Delta\text{Intermediary-Level}_t + \text{Controls}_t \quad (+\text{Trend}) \quad + e_t. \tag{1.9}$$

(Insert Table 1.14)

Finally, I connect CoAnomaly to the intermediary asset pricing literature, which has been attracting a lot of attention recently. Both Adrian et al. (2014) and He et al. (2017) find that the shocks to financial intermediaries' balance sheet can have strong asset pricing power. However, their results are somehow contradictory about the sign of the price of risk: Adrian et al. (2014) use leverage of securities broker-dealers (from the Federal Reserve *Flow of Funds* data) and He et al. (2017) use equity capital ratio of primary dealers (the holding companies of trading counterparties to the Federal Reserve Bank of New York), which is the reciprocal of the leverage. Nevertheless, both of them find positive risk premia for the shocks. Cho (2017) directly models that the intermediary-originated funding shocks to arbitrageurs will induce excess comovement (beyond fundamentals) in anomaly returns and hence generate endogenous risk[22]. This research directly links the CoAnomaly measure to the time-series variation of the intermediary balance sheet.

[22]Indeed, to infer the endogenous risk partially induced by the trading of sophisticated investors, ideally, researchers would like to observe their trading behaviors directly. However, the trading data and holding data of sophisticated arbitrageurs are both notoriously difficult to obtain in practice. Given the size of the institutional investors, their trading behaviors will pose a substantial price impact on any assets, hence generating comovements and price impacts across assets. This also motivates my study in the previous sections.

Table 1.14 reports the regression results of CoAnomaly news[23] on these financial intermediary time series. I find that CoAnomaly shock has a negative loading on both leverage shock and the capital ratio shock, which is consistent with the opposite signs in risk premia between CoAnomaly (negative) and the leverage shock (positive as in Adrian et al. (2014)) / the capital ratio shock (positive as in He et al. (2017)). This result is worth noting because these two intermediary measures are negatively correlated by construction because they are reciprocals of each other by construction[24], and the CoAnomaly shock has a consistent sign on both of them. However, I do not find any relationship with the term structure noise from Hu et al. (2013), which measures the illiquidity in the arbitrage of the treasuries across maturities. I also check if CoAnomaly shocks are correlated with real economy risk variables, but I find no strong relationships with financial uncertainty and macro uncertainty from Jurado et al. (2015) and *cay* variable from Lettau and Ludvigson (2001).

The evidence suggests that the CoAnomaly measure is partially linked to the intermediary asset pricing. It is also a support for Cho (2017)'s endogenous risk story that the aggregate shocks to the financial intermediaries will affect the funding liquidity conditions of arbitrageurs, and consequently, induce excess comovement in anomaly assets that these arbitrageurs are betting on.

1.7 Conclusion

Given the empirical fact that the variance of anomalies is quickly mean-reverting and does not predict future anomaly returns, I focus on the correlations among these anomalies. I propose a time-series risk measure *CoAnomaly* based on averaging the daily return correlations among 34 stock market anomalies to proxy the correlation risk faced by arbitrageurs, who are generally regarded as the main investors of these anomalies.

I find that CoAnomaly, (1) is an important component and a strong predictor of aggregate variance of the diversified anomaly portfolio, (2) positively predicts future

[23]CoAnomaly news is estimated from the intertemporal CAPM VAR, and it is the same time-series of the CoAnomaly-driven variance news since I assume the average variance does not change.

[24]They are not entirely negatively correlated as the two papers use different datasets and definitions about financial intermediaries. In my sample, the correlation coefficient is 0.09

average anomaly returns, and (3) carries a negative price of risk in the cross-section, which indicates that the loss of investment diversification outweighs the benefit of higher future anomaly returns. I also find that return patterns are consistent with the idea that arbitrageurs take the CoAnomaly risk into account. These results together highlight the importance of the comovement among anomaly assets. These results show that the anomaly return dynamics can be rationalized in a portfolio view from the perspective of anomaly investors.

The fact that CoAnomaly is robustly priced across different assets has a strong asset pricing implication. The impact of professional asset managers is substantial since the risk they care about is incorporated into the prices of many assets both in the time series as well as the cross-section. There are policy implications for the CoAnomaly measure as well: regulators can use it to evaluate the likelihood that the stock market arbitrageurs destabilize the market if there is a market-wide shock to the correlation structure. Based on this measure, future research can explore the mechanisms and rationales behind the behaviors of the arbitrageurs with substantial impacts, which may, in turn, lead to a better understanding of financial markets.

1.8 Appendix

1.8.1 34 Stock Market Anomalies and Robustness of Predictability

(Insert Table 1.15)

The full list of 34 stock market anomalies is reported in Table 1.15. The average monthly raw return of the equal-weighted anomaly return (E.A.R.) is 0.55%, with a standard deviation of 1.31%. This gives an annual Sharpe ratio 1.5.[25]

Too Many Anomalies, Duplicated Anomalies, and Sample back to 1963 There is some concern that some anomalies have high correlations by construction, as some composite anomalies incorporate information from other anomalies and some other anomalies are quite similar to each other, like the net issuance annual and net issuance monthly, PEAD with CAR3 (cumulative abnormal return) and PEAD with SUE (standardized unexpected earnings). However, I argue that this will not undermine my results since what I am exploiting is the time-variation of the correlations among anomalies. The time-variation will not be sabotaged by the unconditional high correlations among a few anomalies (as long as the predictability is not driven by these highly-correlated anomalies). To make sure this is the case, I also report results with fewer anomalies. As shown in Table 1.16, my main result of predictability is robust to two different anomaly sets: 23 anomalies studied in Novy-Marx and Velikov (2016) (NMV) and 11 anomalies studied in Stambaugh et al. (2012) (SYY) separately.

I also check a longer sample period, which dates back to 1963. So between 1963 and 1972, I calculate CoAnomaly, E.A.R., aggregate variance and average variance measure based on 27 anomalies which do not require valid *Report Date of Quarterly Earnings (RDQ)* (see Table 1.15 for 7 excluded anomalies). Later, I merge the early period data (1965-1972) with the original sample period data (1973-2017) and run the predictive regression together. To avoid the possible complication in levels since I am using different sets of anomalies, I also include a dummy for the pre-1973 quarters. The predictability shows up robustly.

[25] Assume there is no serial dependence, $\frac{0.55\% \times 12}{1.31\% \times \sqrt{12}} = 1.4544$

(Insert Table 1.16)

Separating Long Legs and Short Legs I conduct the same predictive regression separating the long legs and short legs across 34 anomalies and report the results in Table 1.17. Since the long-short portfolio is the long leg minus the short leg for each anomaly, the coefficients for specifications (1), (5) and (7) in the Panel B of Table 1.3 are the difference between the coefficients for long legs and short legs in Table 1.17. The predictability is robust for both legs. Please also note that the market beta for short legs is higher than long legs.

(Insert Table 1.17)

Different Horizons Table 1.18 reports predictive regression estimates of the equal-weighted anomaly returns (E.A.R.) for return intervals of one, six and twelve months using overlapping data. I find consistent results about the positive predictability of the CoAnomaly measure across different time horizons.

(Insert Table 1.18)

Mean-Variance Efficient Portfolio Since the trading size/capacity of different anomalies has no clear definition like market cap for different stocks in the stock market and is difficult to measure precisely, I remain agnostic about the relative composition of the 'optimal' portfolio in this anomaly-investing universe[26]. So my main analysis focus on the most naïve way to aggregate these anomalies - simple equally averaging them. This nonparametric approach should provide conservative results. However, it is worth exploring other specifications of aggregating these anomalies.

Table 1.19 reports the predictive regression results for the mean-variance efficient (MVE) portfolio. The predictability is stronger for the raw returns compared with E.A.R., which is not surprising considering the nature of the mean-variance efficient

[26] Recently there has been some literature studying this topic: Novy-Marx and Velikov (2016) find strategies based on size, value, and profitability have the greatest capacities to support new capital.

portfolio. It is the optimal portfolio given the sample covariance structure, which shares the information of CoAnomaly. MVE portfolio also shows no strong loadings on the market and size factors, which is intuitive as the optimization procedure will automatically choose not to load on the factors with a negligible risk premium. Note that I do not adjust the weight to calculate the average variance, aggregate variance and correlation (CoAnomaly) based on the MVE portfolio.

(Insert Table 1.19)

1.8.2 Estimating CoAnomaly: Estimation Errors, Calculation, and Other Proxy

Estimation Errors

(Insert Figure 1.5)

As mentioned before, the standard errors in the predictive regression require a caveat because they do not take into account the estimation uncertainty in CoAnomaly. Most of the previous studies on high-frequency asset correlations do not correct the standard error issue from the estimation of correlations.

Following Djogbenou et al. (2015), I conduct a double-layer block bootstrap: I first resample blocks of quarters from predictive regression, and I choose 4 quarters as a block; and then conditioning on each resampled block, I resample daily anomaly returns in the last quarter for each quarter in that block, and use these resampled daily anomaly returns to calculate a new CoAnomaly measure; and then use all quarters from the resampled blocks with resampled CoAnomaly to calculate the predictive coefficient.

Table 1.8 has shown that the t-stats with bootstrap standard errors is only marginally smaller than t-stats with Newey and West (1987) correction. The differences between these two are marginal and my results stay significant robustly.

I also conduct a bootstrap procedure to calculate the standard errors for the CoAnomaly time series. I resample daily cross-sections of anomaly returns within a quarter, which keeps the return structure across anomalies but also effectively assumes that there is time-series dependence across trading days. The standard errors for each quarter are based on resampling 10000 times.

As shown in Figure 1.5, I draw the original CoAnomaly time series together with its band of 90 percent confidence interval, and also the time series of the standard errors as well. The estimation errors (around 0.02) are small in relative magnitude compared with the time variation of CoAnomaly (with a standard deviation around 0.08), which suggests that the estimation is quite precise. This is because the estimation uses much more observations from high-frequency daily data. Effectively I am using more than 2000 data points (34 anomalies times 63 trading days in a quarter) to estimate one metric, which is CoAnomaly. These relatively small estimation errors suggest less serious problems for later inferences.

Principle Component I also calculate the first principal component in the 34 anomalies and find the variance of the first principal component does not contribute to my results. As the principal component calculation requires the full sample, which may be contaminated with some high volatility period, the no-effect result is not unexpected. This check follows Connor and Korajczyk (1986) and Connor and Korajczyk (1988), where they argue the first principal component for all stocks is a common risk factor[27].

Small and High iVol Stocks Considering the fact that the short leg of anomalies is the main driver of the time-variation in CoAnomaly, and another fact that the stocks in the short leg are small and volatile for most anomalies, a natural concern is that the comovement among these small and volatile stocks may drive the results. To rule out this mechanism, I conduct the following placebo test: I generate pseudo anomalies through picking up similar stocks with the same size and iVol characteristic scores for each anomaly. Then calculate a pseudo-CoAnomaly from these 34 pseudo anomalies. I find no effect of time-series predictability and cross-sectional pricing.

Too Many Anomalies and Lack of Dimension One concern about CoAnomaly is that among all these anomalies, some of them are highly correlated by construc-

[27] To some extent, if the volatility of the underlying single factor is higher, the correlations among these anomalies will be higher assuming that the loadings on the single factor do not change.

tion, for example, investment anomaly will mechanically be highly correlated with asset growth anomaly and net issuance anomaly. In the meantime, different strategies will have different size of trading capital in it. Following these two points, simple equal weighting different anomalies may overweight some anomalies and cannot fully exploit all information from the correlation structure, hence containing a certain level of noise.

I fully acknowledge this concern and conduct a simple robustness test, which produces results consistent with my findings. Instead of sorting months based on CoAnomaly, I sort all months by the correlation between two mispricing factors as in Stambaugh and Yuan (2016). In their study, they group 11 anomalies, which have been studied in Stambaugh et al. (2012) and Stambaugh et al. (2015), into two sets based on either cross-sectional correlations of stocks' rankings on the anomaly characteristics or time-series correlations of anomalies' long-short return spreads. Both methods yield the same clusters of anomalies in their work. I believe this measure will not suffer the problem of incorrectly-overweighting some set of anomalies due to mechanical high correlations (low dimensionality). As reported in Panel A of Table 1.20, there are two pieces of evidence supporting my results. First, the CoAnomaly measure in next period is also increasing across groups sorted by the correlation of two mispricing factors, which means that I am indeed catching up some component in the correlation among all these stock market strategies. Second, the pattern of future anomaly returns is consistent in both economic magnitude and statistical significance.

I also repeat the above procedure with the Fama and French (2015)'s five-factor model, Carhart (1997)'s four-factor model and Hou et al. (2015)'s q-factor model. My results are robust to these specifications. Moreover, in non-tabulated results, I also conduct Monte-Carlo simulations to randomly pick anomalies from the original 34 anomalies and then calculate the CoAnomaly within the newly picked 34 anomalies (with replacement). For $9,647$ out of $10,000$ simulations, my results remain qualitatively similar.

CoAnomaly Calculation Window As reported in Panel B of Table 1.20, the results remain qualitatively unchanged if I use the CoAnomaly measure calculated within one previous month. My main goal is proposing a new measure. In practice, money managers are facing different beta constraints and concentration limits, and they also have different assets in hand. So they can certainly choose their optimal anomaly / strategy set, weights, frequency, and sample window to calculate the correlation risk measure tailored for and based on their portfolio composition and other concerns.

(Insert Table 1.20)

Anomalies Sharing Same Stocks CoAnomaly will increase mechanically at times when anomalies are sharing more stocks that are '*mispriced*'. In the extreme case, if all anomalies are longing the same '*underpriced*' stocks and shorting the same '*overpriced*' stocks, CoAnomaly will be 1.

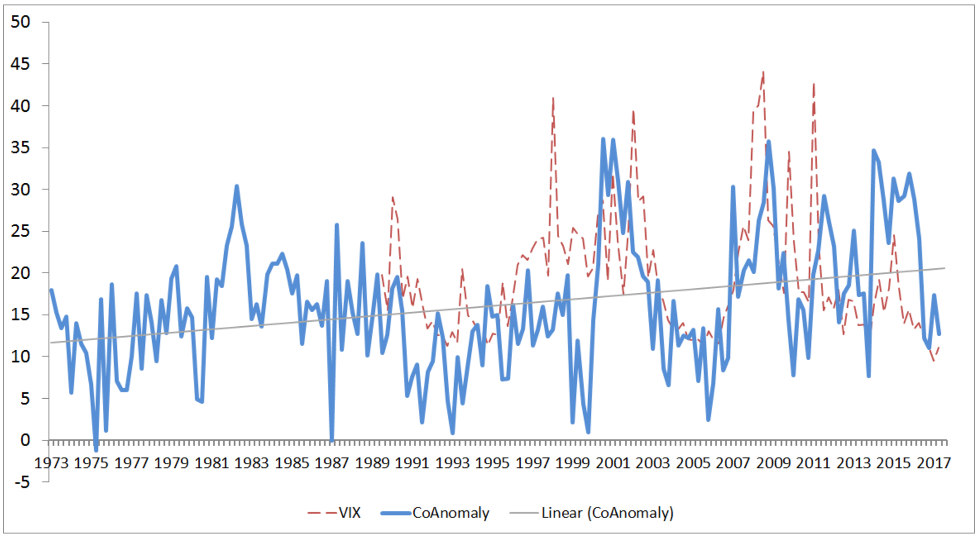
I propose a measure which I term the weighted anomaly score (WAS), which captures this effect. For each stock at each point of time (end of a month), I first calculate the short anomaly score (how many strategies/anomalies are shorting it), then divide it with the total number of anomalies (34 in my case) to normalize the score to one. I calculate the weighted average anomaly score for each anomaly and then take the simple mean across all anomalies. This procedure is repeated in each month, so a time series of WAS is generated. The same procedure can be implemented in the long leg as well.

(Insert Table 1.21)

The results in Table 1.21 show that the composition mechanism is indeed an important driver of CoAnomaly and in time-series evidence I also find the WAS predicts the E.A.R.; however, in the cross-sectional pricing tests, I do not find a significant price of risk.

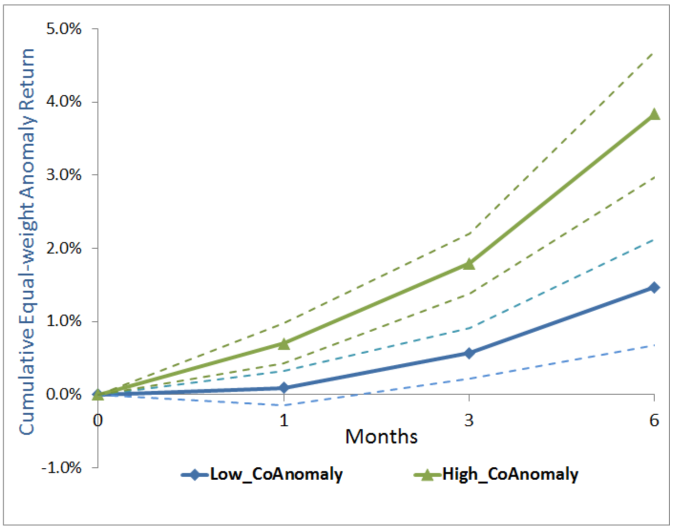
This finding is not surprising. Suppose there is a single asset which is strongly overpriced in the current period, it will be picked up by the short legs of strategies using price information. This effectively mechanically increases both the WAS and the CoAnomaly measure. In the next period(s), the price of the asset will go back to the rational value, which will generate predictability for both WAS and CoAnomaly. However, from arbitrageurs' point of view, a single deeply-mispriced asset will not be *systemically* important in terms of their fully diversified portfolio, so this WAS change is unlikely to induce a price of risk in the aggregate cross-section. Note that, this composition mechanism, as one driver of the time-varying CoAnomaly, does not undermine the trade-off story of arbitrageurs. Because these mispriced assets are mispriced in equilibrium, the fact that their mispricing was '*corrected*' in the next period but not in this period suggests that there are some limits which prevent arbitrageurs exploiting the mispricing fully, which is the main question I am exploring in this paper. One limit can be the lack of diversification, effectively captured by the CoAnomaly measure.

List of Figures



This figure plots the time-series of the long-short CoAnomaly measure based on 34 stock market anomalies. The blue line is the CoAnomaly measure with a time trend in grey, and the red dashed line is the CBOE VIX measure.

Figure 1.1: Time-Series of CoAnomaly



This figure reports the cumulative return (with 95% confidence intervals in dashed line) of E.A.R. for one, three, and six months after different CoAnomaly periods, for the second half of my sample spanning 1994 to 2017. It illustrates graphically the results in Panel B1 in Table 1.4.

Figure 1.2: Cumulative Equal-weighted Anomaly Returns after different CoAnomaly periods

Table 1.1: Summary Statistics of CoAnomaly and its Time-series Correlation with other Measures

This table reports the summary statistics of CoAnomaly measures and the time-series correlations among CoAnomaly for the long-short portfolios, long leg only, and short leg only as well as the time-series correlations among CoAnomaly measures and other contemporaneous market variables. Avg. Corr. is the average calculation in market portfolios following Pollet and Wilson (2010). MktRf is the market excess return from French's website. Realized Vol is the quarterly realized variance in the market, calculated from the daily returns. CBOE VIX is the key measure of market expectations of near-term volatility conveyed by S&P 500 stock index option prices. TED spread is the difference between the interest rates on interbank loans and on short-term U.S. government debt ("T-bills"). Avg_Liq_Level is the market liquidity measure by Pastor and Stambaugh (2003). HF Ret. is the Equity Market Neutral Index (HFRIEMNI) from Hedge Fund Research. Mispricing factors correlation is the correlation between two mispricing factors from Stambaugh and Yuan (2016). Sentiment index is taken from Jeffrey Wurgler's website and covers the period until September 2015.

	Panel A: Full Sample (1973-2017)			Panel B: Post-1994		
	CoAnomaly_LS	**CoAnomaly_S**	**CoAnomaly_L**	**CoAnomaly_LS**	**CoAnomaly_S**	**CoAnomaly_L**
Mean	0.17	0.30	0.15	0.17	0.32	0.17
St.D.	0.09	0.08	0.08	0.09	0.07	0.08
No. Oberv.	180	180	180	96	96	96
	Correlation			**Correlation**		
CoAnomaly_LS	1	0.72	-0.28	1	0.67	-0.26
CoAnomaly_S	0.72	1	-0.23	0.67	1	-0.23
CoAnomaly_L	-0.28	-0.23	1	-0.26	-0.23	1
Avg. Corr.	0.21	0.04	-0.18	0.31	0.20	-0.27
MktRF	0.02	-0.04	0.02	-0.06	-0.15	0.11
Mkt Realized Vol	0.31	0.15	-0.04	0.41	0.19	-0.17
VIX				0.40	0.18	-0.09
TED rate				-0.05	-0.19	0.03
Liquidity Level				-0.10	0.00	-0.02
HF Ret.				-0.28	-0.12	0.20
Mispri. Factors Corr.				0.57	0.47	-0.10
Sentiment				0.01	0.16	-0.01

Table 1.2: Determinants of CoAnomaly and Aggregate Variance

$$\text{CoAnomaly}_t = a + b \times \text{CoAnomaly}_{t-1} + \sum_p m_p \times \text{Controls}_{p,t-1} + t \times \text{Trend} + e_t.$$

$$\text{Aggr.Var}_{EAR,t} = a + b_1 \times \text{CoAnomaly}_t + b_2 \times \text{Avg.Var}_t + b_3 \times \text{CoAnomaly}_t \times \text{Avg.Var}_t + u_t$$

Panel A reports the predictive regression results of regressing CoAnomaly measures on the lag of the same CoAnomaly measure and other state variables in the last quarter (standard deviation normalized to 1). The coefficient on Trend is multiplied by 1000 and the coefficients on other non-CoAnomaly regressors are multiplied by 100 for readability. Panel B reports the regression results for different specifications for the aggregate variance of equal-weighted anomaly return (E.R.A.). The sample period is from 1973 to 2017, 180 quarters. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

Panel A: Determinent of CoAnomaly							
Dep. Var.	CoAnomaly_LS	CoAnomaly_S	CoAnomaly_L	CoAnomaly_LS			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
CoAnomaly t-1	0.52	0.49	0.56	0.51	0.41	0.55	0.42
	(6.02)	(6.81)	(8.85)	(5.73)	(5.50)	(6.46)	(5.06)
Sentiment t-1					3.12		2.67
					(4.13)		(3.36)
VIX t-1					2.10		1.83
					(3.08)		(1.78)
Trend T	0.09	0.13	-0.01	-0.05	0.37	0.27	0.37
	(2.39)	(3.83)	(-0.42)	(-0.39)	(3.60)	(2.96)	(3.78)
Trend T-squared				0.00			
				(0.94)			
Other Predictors	N	N	N	N	N	N	Y
Adj. R-squared	35.1%	36.6%	34.2%	36.4%	51.8%	51.5%	52.2%

Panel B: Dependent Variable: Aggregate Variance of Equal-weighted Anomaly Returns (E.A.R.) estimated at t				
	(1)	(2)	(3)	(4)
CoAnomaly t	0.741		0.123	
	(2.72)		(3.84)	
Average Var. t		0.171	0.164	
		(8.17)	(8.11)	
CoAnomaly*(Avg.Var.) t				0.643
				(12.63)
Adj. R-squared	21.0%	78.4%	81.3%	94.3%
CoAnomaly t-1	0.749			
	(2.45)			
CoAnomaly t-2		0.720		
		(2.37)		
CoAnomaly t-3			0.591	
			(2.25)	
CoAnomaly t-4				0.398
				(2.01)
Adj. R-squared	20.9%	19.2%	12.7%	7.5%
Avg.Var t-1	0.124			
	(4.53)			
Avg.Var t-2		0.075		
		(3.14)		
Avg.Var t-3			0.040	
			(2.41)	
Avg.Var t-4				0.022
				(1.66)
Adj. R-squared	41.4%	14.9%	3.8%	0.7%
Aggr.Var t-1	0.787			
	(6.94)			
Aggr.Var t-2		0.503		
		(6.89)		
Aggr.Var t-3			0.275	
			(2.43)	
Aggr.Var t-4				0.139
				(1.73)
Adj. R-squared	61.7%	24.9%	7.1%	1.4%

Table 1.3: Predictive Regression at Quarterly Level (E.A.R.)

$$\text{E.A.R.}_{t+1} = a + b \times \text{CoAnomaly}_t + \sum_p m_p \times \text{Other.Predictors}_{p,t} + t \times \text{Trend} \quad (+ \sum_j \beta_j \times Benchmark.Factors_{t+1}) +$$

The dependent variable is the equal-weighted anomaly returns (E.A.R.) for the next quarter $t+1$. All independent variables are measured in the quarter t. CoAnomaly is the average partial correlation among 34 stock market anomalies (long-short). Average realized variance is equally averaging the realized daily variances for the 34 stock market anomalies. Aggregate variance of the E.A.R. is measured as the realized variance of daily returns. Other predictors include TED spread (TED), market excess return, market average correlation, and E.A.R.. The standard deviations of all regressors are also normalized to 1 and returns are measured in percentage. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

	Panel A: Dependent Variable Quarterly E.A.R. at t+1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CoAnomaly t	0.71			0.71	0.78	0.51	0.80
	(2.59)			(2.58)	(2.69)	(1.97)	(2.85)
Average Var. t		0.04		0.03	0.39	0.02	0.35
		(0.18)		(0.11)	(0.75)	(0.03)	(0.69)
Aggregate Var. t			0.04		-0.37	-0.51	-0.82
			(0.18)		(-0.79)	(-1.19)	(-1.87)
Anomaly Value Spread t						0.29	0.48
						(1.13)	(1.89)
Sentiment t						1.52	1.80
						(3.81)	(4.02)
Other Predictors	N	N	N	N	N	N	Y
Trend T (negative)	Y	Y	Y	Y	Y	Y	Y
Adj. R square	8.1%	-1.8%	-1.8%	7.3%	7.2%	16.4%	27.8%
N	179	179	179	179	179	179	118

	Panel B: Dependent Variable Quarterly E.A.R. at t+1 with Benchmark Factor Adjustment						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CoAnomaly t	0.78			0.75	0.80	0.67	0.74
	(4.29)			(4.14)	(4.12)	(3.50)	(3.50)
Average Var. t		0.34		0.28	0.48	0.50	0.55
		(2.30)		(2.07)	(1.55)	(1.65)	(1.57)
Aggregate Var. t			0.34		-0.22	-0.37	-0.41
			(2.39)		(-0.71)	(-1.21)	(-1.20)
Anomaly Value Spread t						0.04	0.03
						(0.21)	(0.18)
Sentiment t						0.78	0.69
						(2.72)	(1.98)
MktRf t+1	-0.17	-0.14	-0.14	-0.15	-0.15	-0.14	-0.15
	(-7.40)	(-5.52)	(-5.72)	(-6.33)	(-6.32)	(-6.21)	(-6.08)
SMB t+1	-0.07	-0.10	-0.10	-0.09	-0.09	-0.09	-0.09
	(-2.63)	(-3.14)	(-3.18)	(-3.15)	(-3.05)	(-3.13)	(-3.13)
HML t+1	0.20	0.18	0.18	0.20	0.20	0.17	0.17
	(6.73)	(5.92)	(5.84)	(6.90)	(6.92)	(5.67)	(5.33)
UMD t+1	0.18	0.20	0.20	0.20	0.19	0.18	0.18
	(8.76)	(8.54)	(8.56)	(9.09)	(8.83)	(8.26)	(7.88)
Other Predictors	N	N	N	N	N	N	Y
Trend T (negative)	Y	Y	Y	Y	Y	Y	Y
Adj. R square	74.9%	68.7%	69.1%	74.8%	74.6%	76.5%	77.4%
N	179	179	179	179	179	170	118

Table 1.4: Monthly Sorting with Raw Returns

This table reports the mean of different measures after sorting all months into different groups based on A) single-sorted by the CoAnomaly measure in the last quarter or B) double-sorted by returns of quantitative equity hedge funds (HFRIEMNI) in the last quarter and then by the CoAnomaly measure in last quarter. Both measures are detrended to make sure my results are not driven by more recent periods. Then, I report the cumulative E.A.R. for the next 1, 3 and 6 months. T-stats, calculated with Newey and West (1987) correction for 4 lags, are shown in parentheses.

Panel A: Sort all months based on CoAnomaly (Full sample)

	CoAnomaly Group	No. Months		CoAnomaly t	CoAnomaly t+1	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
	1	162	.	0.10	0.11	0.45%	1.38%	2.91%
				(14.45)	(15.11)	(4.77)	(7.57)	(10.67)
	2	216	.	0.15	0.15	0.54%	1.69%	3.44%
				(26.14)	(25.38)	(7.31)	(12.09)	(16.53)
	3	162	.	0.22	0.20	0.73%	2.04%	4.11%
				(29.98)	(26.89)	(7.08)	(11.23)	(13.64)
	Diff 3-1			0.12	0.09	0.28%	0.66%	1.20%
				(11.69)	(8.37)	(2.01)	(2.57)	(2.95)

Panel B1: Sort all months based on CoAnomaly (Second Half sample)

	CoAnomaly Group	No. Months	HF Ret. t	CoAnomaly t	CoAnomaly t+1	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
	1	79	5.2%	0.12	0.13	0.09%	0.57%	1.47%
			(30.89)	(11.15)	(11.56)	(0.61)	(2.11)	(3.53)
	2	105	4.7%	0.17	0.17	0.40%	1.15%	2.09%
			(30.13)	(17.56)	(16.85)	(3.11)	(4.66)	(6.24)
	3	80	5.1%	0.24	0.23	0.70%	1.79%	3.83%
			(36.32)	(19.32)	(18.90)	(4.07)	(5.76)	(7.11)
	Diff 3-1		-0.04%	0.11	0.10	0.61%	1.23%	2.36%
			(-0.19)	(7.00)	(6.25)	(2.66)	(2.97)	(3.45)

Panel B2: First sort all months based on HF Ret., and then sort on CoAnomaly (Second Half sample)

HF Ret. Group	CoAnomaly Group	No. Months	HF Ret. t	CoAnomaly t	CoAnomaly t+1	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
	1	23	3.1%	0.15	0.18	-0.38%	-0.63%	-0.04%
1	2	32	3.1%	0.15	0.16	0.19%	1.11%	0.94%
	3	24	3.6%	0.25	0.25	1.23%	2.23%	3.45%
	Diff 3-1					1.61%	2.86%	3.50%
						(3.08)	(2.50)	(2.15)
	1	31	5.1%	0.12	0.12	0.52%	0.69%	1.56%
2	2	42	5.2%	0.18	0.16	0.24%	0.80%	1.94%
	3	32	5.2%	0.26	0.25	0.70%	1.74%	4.07%
	Diff 3-1					0.18%	1.05%	2.51%
						(0.53)	(2.07)	(2.92)
	1	24	6.5%	0.09	0.09	0.24%	1.44%	3.20%
3	2	32	6.5%	0.15	0.15	0.30%	1.37%	3.03%
	3	24	6.4%	0.22	0.21	0.56%	1.92%	4.17%
	Diff 3-1					0.32%	0.49%	0.97%
						(0.88)	(0.89)	(0.85)

Table 1.5: Monthly Sorting with Benchmark-Adjusted Returns

This table reports the benchmark-adjusted returns after sorting all months into different groups based on A) single-sorted by the CoAnomaly measure, or B) double-sorted by returns of quantitative equity hedge funds, and then by the CoAnomaly measure. Both measures are detrended to make sure my results are not driven by more recent periods. All returns are adjusted for either the single market factor or Carhart (1997) four factors. T-stats, calculated with Newey and West (1987) correction for 4 lags, are shown in parentheses.

Panel A: Sort all months based on CoAnomaly (Full sample)

	Market-Adjusted			Carhart-4-Adjusted		
CoAnomaly Group	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
1	0.48%	1.49%	3.03%	0.38%	1.20%	2.67%
	(5.08)	(7.84)	(10.57)	(4.05)	(6.70)	(9.79)
2	0.57%	1.78%	3.67%	0.58%	1.78%	3.42%
	(7.48)	(12.79)	(17.32)	(7.22)	(12.15)	(16.24)
3	0.75%	2.13%	4.27%	0.75%	2.13%	4.42%
	(6.81)	(11.28)	(14.16)	(7.65)	(11.81)	(14.49)
Diff 3-1	0.27%	0.64%	1.24%	0.37%	0.93%	1.75%
	(1.87)	(2.42)	(2.96)	(2.74)	(3.68)	(4.28)

Panel B1: Sort all months based on CoAnomaly (Second Half sample)

	Market-Adjusted			Carhart-4-Adjusted		
CoAnomaly Group	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
1	0.18%	0.81%	1.96%	0.08%	0.58%	1.47%
	(1.55)	(3.63)	(5.68)	(0.55)	(2.12)	(3.47)
2	0.51%	1.41%	2.58%	0.46%	1.29%	2.40%
	(5.32)	(8.04)	(10.57)	(3.50)	(5.08)	(6.53)
3	0.65%	1.76%	3.69%	0.71%	1.87%	4.00%
	(5.00)	(7.65)	(9.47)	(3.78)	(5.81)	(7.32)
Diff 3-1	0.47%	0.95%	1.73%	0.63%	1.28%	2.53%
	(2.73)	(2.95)	(3.32)	(2.63)	(3.03)	(3.64)

Panel B2: First sort all months based on HF Ret., and then sort on CoAnomaly (Second Half sample)

		Market-Adjusted			Carhart-4-Adjusted		
HF Ret. Group	CoAnomaly Group	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
	1	-0.18%	0.09%	1.11%	-0.41%	-0.69%	-0.14%
1	2	0.56%	1.68%	2.38%	0.20%	1.06%	0.99%
	3	0.81%	1.85%	3.42%	1.13%	2.09%	3.27%
	Diff 3-1	0.99%	1.75%	2.31%	1.55%	2.78%	3.41%
		(3.01)	(2.60)	(2.40)	(3.00)	(2.48)	(2.19)
	1	0.51%	0.74%	1.56%	0.42%	0.57%	1.62%
2	2	0.34%	1.15%	2.46%	0.34%	1.09%	2.47%
	3	0.64%	1.65%	3.69%	0.82%	1.94%	4.34%
	Diff 3-1	0.13%	0.91%	2.13%	0.40%	1.37%	2.72%
		(0.44)	(2.18)	(2.98)	(1.12)	(2.60)	(3.00)
	1	0.30%	1.47%	2.93%	0.34%	1.64%	3.09%
3	2	0.57%	1.56%	3.63%	0.39%	1.61%	3.54%
	3	0.49%	1.92%	4.03%	0.50%	1.96%	4.37%
	Diff 3-1	0.19%	0.46%	1.10%	0.16%	0.32%	1.27%
		(0.64)	(1.00)	(1.23)	(0.39)	(0.51)	(1.04)

Table 1.6: Monthly Sorting with Daily Raw Returns and Higher Moments

This table reports the statistics of daily raw E.A.R. returns after sorting all months into different groups based on the CoAnomaly measure. CoAnomaly measure is detrended to make sure my results are not driven by more recent periods. I first calculated the statistics within each tracking period (1 month or 3 months after sorting) for each month and take the simple average for each statistic across months within the same CoAnomaly group.

Panel A: Sort all months based on CoAnomaly (Full sample)							
		Daily raw E.A.R. returns after next 1 month			Daily raw E.A.R. returns after next 3 months		
CoAnomaly Group	No. Months	Average	St.D.	Skewness	Average	St.D.	Skewness
1	162	0.016%	0.160%	-0.044	0.016%	0.165%	-0.025
2	216	0.026%	0.151%	-0.073	0.025%	0.158%	-0.002
3	162	0.028%	0.216%	-0.101	0.029%	0.218%	-0.128

Panel B: Sort all months based on CoAnomaly (Second Half sample)							
		Daily raw E.A.R. returns after next 1 month			Daily raw E.A.R. returns after next 3 months		
CoAnomaly Group	No. Months	Average	St.D.	Skewness	Average	St.D.	Skewness
1	79	0.005%	0.178%	-0.071	0.009%	0.183%	-0.062
2	105	0.017%	0.177%	-0.079	0.015%	0.186%	-0.098
3	80	0.030%	0.276%	-0.151	0.028%	0.277%	-0.149

Table 1.7: Out-of-Sample Predictability

This table reports the results of the out-of-sample tests. The out-of-sample adjusted $\overline{R^2}$ is reported for each single predictors of E.A.R., with significance level calculated from MSE-F one-sided tests. All predictors and E.A.R. are detrended first to avoid complication. CoAnomaly is the average partial correlation among 34 stock market anomalies (long-short). The E.A.R. volatility is the aggregate variance of the E.A.R., measured as the realized variance of daily returns. Sentiment is the investor sentiment and anomaly value spread is the average book-to-market ratio difference between anomaly long legs and short legs. Numbers are reported in percentage.

	OOS $\overline{R^2}$	ΔRMSE
CoAnomaly	2.16**	0.28
E.A.R. Volatility	-1.21	-0.87
Sentiment	0.68*	0.07
Anomaly Value Spread	0.13	0.03

Table 1.8: Predictive Regression with Bootstrap T-stats

$$\begin{aligned}\text{E.A.R.}_{t+1} = \quad & a + b \times \text{CoAnomaly}_t + t \times \text{Trend} \\ & + \sum_p m_p \times \text{Other.Predictors}_{p,t} \quad (+\sum_j \beta_j \times \mathit{Benchmark.Factors}_{t+1}) + e_{t+1}.\end{aligned}$$

The dependent variable is the Equal-weighted Anomaly Returns (E.A.R.) for the next quarter $t+1$. All independent variables are measured in the quarter t. CoAnomaly is the average partial correlation among the whole long-short portfolio of 34 stock market anomalies. Average realized variance is equally averaging the realized daily variances for the 34 stock market anomalies. Aggregate variance of the Equal-weighted Anomaly Returns (E.A.R.) is measured as the variance of daily returns. E.A.R. is Equal-weighted Anomaly Returns and Anomaly Value Spread is the average value spread for all anomalies. The standard deviations of all regressors are also normalized to 1 and returns are measured in percentage. Results in the left panel do not adjust contemporaneous benchmark returns and results in the tight panel are adjusted for Carhart-4 factors. T-stats, shown in parentheses, are computed with: Newey and West (1987) correction for 4 lags in the first row; double-layer block bootstrap standard errors in the second row.

	Dependent Variable: Quarterly E.A.R. at t+1					
	Raw Return			Benchmark-Adjusted Return		
	(1)	(2)	(3)	(4)	(5)	(6)
CoAnomaly t	0.71	0.78	0.80	0.75	0.72	0.63
t-NW	(2.59)	(2.69)	(2.85)	(4.09)	(3.82)	(3.10)
t-bootstrap	(2.11)	(2.16)	(2.14)	(3.52)	(3.09)	(2.46)
Average Var. t		0.39	0.35		0.49	0.55
t-NW		(0.75)	(0.69)		(1.58)	(1.51)
t-bootstrap		(0.54)	(0.48)		(1.09)	(0.91)
Aggregate Var. t		-0.37	-0.82		-0.15	-0.32
t-NW		(-0.79)	(-1.87)		(-0.48)	(-0.92)
t-bootstrap		(-0.53)	(-1.56)		(-0.38)	(-0.56)
Anomaly Value Spread t			0.48			0.00
t-NW			(1.89)			(-0.01)
t-bootstrap			(1.82)			(-0.00)
Sentiment t			1.80			0.69
t-NW			(4.02)			(2.04)
t-bootstrap			(3.00)			(1.89)
Carhart-4 Factors t+1	N	N	N	Y	Y	Y
Trend	Y	Y	Y	Y	Y	Y
Controls	N	N	Y	N	N	Y
Adj. R square	8.1%	7.2%	27.8%	74.9%	74.6%	77.4%
N	179	179	118	179	179	118

Table 1.9: Simulation Results for Biased Estimators

$$\begin{aligned} E.A.R._{t+1} &= a + b \times CoAnomaly_t + e_{t+1} \\ CoAnomaly_{t+1} &= c + d \times CoAnomaly_t + u_{t+1} \end{aligned}$$

Panel A of the table reports the above VAR(1) estimates of the above system, together with the standard deviations and correlations of the estimated shocks. Panel B of the table reports the Monte Carlo simulation results based on different specifications, all conditioning on the null that there is no predictability $b = 0$. Auto Correlation in CoAnomaly and Shock Correlation are the parameters used for the Monte Carlo simulation. For each specification, the Monte Carlo simulation is based on 10000 draws, and within each draw, a time-series with equal length to the original sample (180 quarters) is randomly generated with the original estimated variance-covariance matrix. Summary statistics for the simulated coefficients are reported in the last three columns: the mean, standard deviation, and the probability that the simulated coefficient is larger than the original estimation (based on the normal distribution). T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

Panel A: VAR(1) with E.A.R. and CoAnomaly							
VAR Estimates				Standard Deviation of Shocks			
a	1.50	b	0.71	e	2.34	u	0.79
	(6.93)		(2.84)				
c	0.04	d	0.52	Correlation of Shocks			
	(0.47)		(7.36)	Corr(e,u)=	0.080		

Panel B: Bias Simulated Result: Given No Predictability (b=0)			Simulated Predictability Coefficient: b		
	Auto Correlation in CoAnomaly d	Shock Correlation $Corr(e,u)$	Mean	StD	Prob(Sim>Est)
Specification Estimated	0.52	0.08	-0.004	0.176	0.0024%
Specification 1	0.52	0.00	0.002	0.180	0.0040%
Specification 2	0.52	-0.99	0.034	0.178	0.0074%
Specification 3	0.8	-0.99	0.051	0.135	0.0000%
Specification 4	0.95	-0.99	0.060	0.087	0.0000%

Table 1.10: Pricing Test of CoAnomaly Risk - Anomaly Assets

$$\mathbb{E}[R_i^e] = \lambda_0 + \sum_j \widehat{\beta_{i,j}} \lambda_j + \xi_i.$$

This table reports the pricing results with 34 anomaly portfolios. In each row, I estimate the risk premia λ_j for each factor in each the pricing model by regressing the anomaly returns on estimated betas $\widehat{\beta}_j$ from time-series regression. For CoAnomaly, I use its innovation from a detrended AR(1). Broker Dealer shock follows Adrian et al. (2014) and uses leverage of securities broker-dealers. Primary dealer shock follows He et al. (2017) and uses the equity capital ratio of primary dealers. The estimated risk premia along with Fama and MacBeth (1973) t-stats and Shanken (1992) t-stats are reported. Cross-sectional R^2 statistics are also reported for each pricing model to show how much they can explain the average return dispersion of the test portfolios. Both returns and CoAnomaly are measured in percentage.

Pricing Models	Intercept	MktRf	**CoAnomaly**	Broker Dealer	Primary Dealer	Adj. R-squared
CAPM	1.48	0.94				3.0%
t-FM	(8.50)	(1.41)				
t-Shanken	(6.42)	(1.05)				
CoAnomaly	1.52		-5.99			9.3%
t-FM	(8.53)		(-2.99)			
t-Shanken	(6.49)		(-2.30)			
Broker Dealer Leverage Shock	1.52			11.80		1.1%
t-FM	(8.53)			(0.61)		
t-Shanken	(6.49)			(0.44)		
Primary Dealer Capital Ratio Shock	1.46				15.80	3.3%
t-FM	(8.64)				(1.43)	
t-Shanken	(6.58)				(1.14)	
CoAnomaly+CAPM	1.58	0.72	-5.07			11.3%
t-FM	(8.43)	(1.07)	(-2.46)			
t-Shanken	(6.14)	(0.73)	(-1.97)			
CoAnomaly+Leverage	1.49		-5.25	2.51		9.9%
t-FM	(7.67)		(-2.75)	(0.13)		
t-Shanken	(5.89)		(-2.23)	(0.07)		
CoAnomaly+Capital Ratio Shock	1.52		-4.50		11.44	11.7%
t-FM	(8.78)		(-2.42)		(1.05)	
t-Shanken	(7.49)		(-1.99)		(0.84)	

Table 1.11: Pricing Test of CoAnomaly Risk - Benchmark Assets

$$\mathbb{E}[R_i^e] = \lambda_0 + \sum_j \widehat{\beta_{i,j}} \lambda_j + \xi_i.$$

Panel A of this table reports the pricing results with 25 size and book-to-market portfolios, 10 momentum portfolios, 5 industry portfolios, and 6 treasury bond portfolios sorted by maturity. All of these test portfolios are downloaded from French Data Library. In each row, I estimate the risk premia λ_j of each factor in each the pricing model specification by regressing the portfolio excess returns on estimated betas $\widehat{\beta_j}$ from time-series regression. The estimated risk premia along with Fama and MacBeth (1973) t-stats and Shanken (1992) t-stats are reported. Cross-sectional R^2 statistics are also reported for each pricing model to show how much they can explain the average return dispersion of the test portfolios. Both returns and CoAnomaly are measured in percentage. In panel B, I exclude the 6 bond portfolios.

Pricing Models	Intercept	MktRf	**CoAnomaly**	Adj. R-squared
	Panel A: Equity Portfolios and Bond Portfolios			
CAPM	1.10	1.13		23.9%
t-FM	(3.92)	(1.57)		
t-Shanken	(2.96)	(1.25)		
CoAnomaly	1.83		-10.45	30.1%
t-FM	(3.75)		(-2.28)	
t-Shanken	(2.69)		(-1.97)	
CAPM + CoAnomaly	1.54	0.41	-7.83	33.4%
t-FM	(5.10)	(0.57)	(-2.99)	
t-Shanken	(3.54)	(0.40)	(-2.37)	
	Panel B: Equity Portfolios Only			
CAPM	3.31	-0.90		3.9%
t-FM	(3.82)	(-0.84)		
t-Shanken	(2.42)	(-0.68)		
CoAnomaly	2.14		-8.19	24.1%
t-FM	(3.43)		(-2.38)	
t-Shanken	(2.49)		(-2.02)	
CAPM + CoAnomaly	3.92	-1.86	-9.51	26.4%
t-FM	(4.48)	(-1.74)	(-3.47)	
t-Shanken	(3.14)	(-1.33)	(-2.84)	

Table 1.12: Value-Weighted Portfolios Sorted on Predicted CoAnomaly Betas

At each month-end between 1965 and 2017, eligible stocks are sorted into 5 portfolios according to pre-ranking CoAnomaly betas. The betas are constructed by regressing the excess stock returns on CoAnomaly innovations. The betas are estimated by regressing the value-weighted portfolio excess returns on the CoAnomaly innovation, Fama-French five factors, momentum factor, liquidity factor, and betting-against-beta factor. The first row of Panel A reports the time-series averages of the quintile portfolios' market capitalization (in million dollars). Rest of Panel A reports betas on different factors. Panel B reports the raw and adjusted excess returns based on different benchmark factors. The t-statistics are in parentheses.

Portfolios Sorted on CoAnomaly Beta						
	1 (low)	2	3	4	5 (high)	5-1
	Panel A: Portfolio Betas					
Market Cap (millions)	2046.7	3010.1	3210.5	3451.3	3195.1	
Beta CoAnomaly (post)	-0.018	-0.012	-0.001	0.008	0.013	0.030
	(-1.74)	(-1.06)	(-0.05)	(0.42)	(1.17)	(1.89)
Beta Market	1.05	0.96	0.97	1.00	1.08	0.03
	(51.98)	(65.32)	(79.37)	(80.56)	(56.39)	(0.94)
Beta SMB	0.15	-0.06	-0.02	-0.08	0.05	-0.10
	(4.54)	(-2.37)	(-0.99)	(-3.90)	(1.49)	(-1.91)
Beta HML	0.13	0.02	0.04	-0.03	-0.05	-0.18
	(3.10)	(0.84)	(1.74)	(-1.06)	(-1.42)	(-2.73)
Beta RMW	-0.06	0.02	0.10	0.17	0.01	0.07
	(-1.49)	(0.70)	(4.04)	(6.81)	(0.34)	(1.11)
Beta CMA	-0.10	0.03	0.08	0.05	0.03	0.13
	(-2.19)	(0.96)	(2.80)	(1.89)	(0.77)	(1.79)
Beta MOM	-0.05	-0.06	-0.03	-0.04	0.02	0.07
	(-2.84)	(-3.63)	(-2.39)	(-2.96)	(1.52)	(2.78)
Beta LIQ	-0.01	0.02	0.02	0.00	0.01	0.02
	(-0.54)	(1.14)	(1.26)	(0.26)	(0.38)	(0.55)
	Panel B: Monthly Alphas (in percentage)					
Raw Returns	0.65	0.60	0.62	0.62	0.62	-0.03
	(2.92)	(3.08)	(3.29)	(3.20)	(3.10)	(-0.14)
Alpha CAPM	0.08	-0.04	0.11	0.05	0.03	-0.05
	(1.07)	(-0.44)	(1.47)	(0.72)	(0.52)	(-0.87)
Alpha FF5	-0.02	-0.04	-0.07	-0.09	-0.09	-0.07
	(-0.24)	(-0.72)	(-1.39)	(-1.69)	(-1.58)	(-1.31)
Alpha FF5 + MOM	0.02	-0.01	-0.06	-0.07	-0.11	-0.13
	(0.23)	(-0.18)	(-1.17)	(-1.45)	(-1.81)	(-2.04)
Alpha FF5 + MOM + Liquidity	0.03	-0.05	-0.08	-0.10	-0.12	-0.15
	(0.32)	(-0.71)	(-1.58)	(-2.02)	(-2.15)	(-2.11)

Table 1.13: Panel Regression: Partial Correlation of each Anomaly

$$\text{Part.Corr.}_{i,t}^{mp} = b_1 \times \text{Post-Publication Dummy} \quad (+b_2 \times \text{Trend}) \quad + a_i + e_{i,t}.$$

$$\text{Part.Corr.}_{i,t}^{mp} = b_1 \times \text{Part.Corr.}_{i,t-1}^{mp} + b_2 \times \text{Ret}_{i,t-1} + b_3 \times (\Delta)\text{Short.Interest}_{i,t-1} + a_i + d_t + e_{i,t}.$$

Panel A of this table reports the regression estimates by regressing the partial correlations on the dummies of the post-publication periods (and a trend variable, with coefficient multiplied by 1000 for readability). For each anomaly, its post-publication dummy is equal to 1 if the quarter is in or after the publication year, and zero otherwise. Panel B of this table reports the regression estimates by regressing the partial correlation at time t on the lag of itself, the anomaly return and (the change in) the short interest (raw level or regression dummy) of each anomaly at time t-1. Panel A uses the sample for 27 anomalies (published on or after 1990) from 1990 to 2017. Panel B uses all anomalies and short interest data from mid-1988 following Hanson and Sunderam (2013). T-stats in the parentheses are calculated based on standard errors clustered in time and anomalies.

Panel A: Dependent Variable - Partial Correlation for each Anomaly t		
	(a)	(b)
Post-Publication dummy	0.132	0.063
	(3.83)	(2.72)
Trend T		0.034
		(2.31)
Anomaly Fixed Effect	Yes	Yes
No. Observations	3024	3024
Adj. R-squared	4.9%	8.3%

Panel B: Dependent Variable - Partial Correlation for each Anomaly t				
	Short Interest Level		Short Interest Dummy	
Part.Corr t-1	0.54	0.53	0.54	0.53
	(7.97)	(7.97)	(8.02)	(7.96)
Ret t-1	0.16	0.15	0.16	0.15
	(0.72)	(0.68)	(0.71)	(0.68)
Short.Int t-1	0.04		0.26	
	(1.79)		(0.52)	
Change.Short.Int t-1		-0.02		-0.24
		(-0.72)		(-0.40)
Time Fixed Effect	Yes	Yes	Yes	Yes
Anomaly Fixed Effect	Yes	Yes	Yes	Yes
No. Observations	3944	3944	3944	3944
Adj. R-squared	53.1%	51.4%	52.0%	51.0%

Table 1.14: Regressing the CoAnomaly Shocks and CoAnomaly Levels on Financial Intermediary Balance Sheet Levels and Shocks

$$\text{News.CoAnomaly}_t = b_1 \times \text{Intermediary-Level}_t + b_2 \times \Delta\text{Intermediary-Level}_t + \text{Controls}_t \quad (+\text{Trend}) \quad + e_t.$$

This table reports the contemporaneous quarterly regression estimates. CoAnomaly news is estimated using the third volatility specification in the ICAPM VAR. Financial Intermediary Leverage and Leverage shock are constructed as in Adrian et al. (2014) and Cho (2017). Capital Ratios and Shocks follow He et al. (2017). Term structure noise is from Hu et al. (2013), financial uncertainty and macro uncertainty from Jurado et al. (2015), *cay* variable from Lettau and Ludvigson (2001). I also control for seasonality. The sample period covers from 1973 to 2016 for financial intermediary information, and from 1987 to 2016 for other measures owing to data availability. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

Dependent Variable: CoAnomaly News						
	[1]	[2]	[3]	[4]	[5]	[6]
Leverage	0.001	0.001				
	(1.69)	(0.96)				
Leverage Shock	-0.012	-0.011				
	(-2.11)	(-1.82)				
Capital Ratio			-0.060	-0.034		
			(-0.24)	(-0.14)		
Cap.Ratio Shock			-0.105	-0.109		
			(-2.14)	(-2.25)		
Term Structure Noise					0.000	0.004
					(0.07)	(0.75)
Financial Uncertainty					0.022	0.013
					(0.39)	(0.23)
Macro Uncertainty					0.022	-0.057
					(0.21)	(-0.43)
Cay					-0.195	-0.059
					(-0.59)	(-0.17)
Seasonality	N	Y	N	Y	N	Y
N	172	172	172	172	124	124
Adj. R-Squared	3.2%	5.1%	2.7%	7.2%	-2.7%	1.6%

Table 1.15: 34 Stock Market Anomalies

This table reports the details of 34 stock market anomalies studied in this paper, together with the equal-weighted anomaly return (E.A.R.) by taking the simple average of all anomalies. Column *Source* represents the source of each anomaly: NMV stands for Novy-Marx and Velikov (2016), NMVa for the appendix of Novy-Marx and Velikov (2016), and SYY for Stambaugh et al. (2012). Next three columns report the average, t-stats and the standard deviation of raw monthly returns. Stars-* mean that the anomaly requires *Report Date of Quarterly Earnings (RDQ)* information. I also report the alphas and their OLS t-stats for different benchmark models, including CAPM, Fama-French three-factor model, and Carhart (1997) four-factor model. The sample period covers from 1973 to 2017, with 540 months in total.

No.	Label	Name	Publication Year	Authors	Source	Ret.	t-Ret.	StD-Ret.	CAPM	t-CAPM	FF3	t-FF3	Carhart 4	t-C4
1	acc	Accruals	1996	Sloan,	NMV, SYY	0.27%	(1.91)	3.27%	0.35%	(2.48)	0.28%	(2.09)	0.26%	(1.86)
2	atgrowth	Asset growth	2008	Cooper, Gulen, and Schill,	NMV, SYY	0.28%	(1.89)	3.44%	0.42%	(2.97)	0.13%	(1.07)	0.07%	(0.57)
3	ato	Asset turnover	2008	Soliman,	NMVa	0.36%	(2.22)	3.77%	0.30%	(1.86)	0.36%	(2.24)	0.34%	(2.05)
4	beta	Beta arbitrage	1972	Black,	NMVa	0.71%	(2.82)	5.82%	0.86%	(3.47)	0.57%	(2.47)	0.37%	(1.57)
5	ceissue	Composite equity issues	2006	Daniel and Titman,	SYY	0.41%	(2.74)	3.45%	0.62%	(4.82)	0.47%	(4.25)	0.42%	(3.76)
6	failprob*	Failure probability	2008	Campbell, Hilscher, and Szilagyi,	NMV, SYY	0.31%	(1.06)	6.91%	0.65%	(2.37)	1.07%	(4.21)	0.15%	(0.87)
7	gm	Gross margins	2008	Soliman,	NMVa	0.00%	(0.03)	3.25%	0.04%	(0.29)	0.28%	(2.34)	0.34%	(2.77)
8	hfcombo1	Industry mom. + Relative rev.	2016	Novy-Marx and Velikov,	NMV	1.27%	(8.20)	3.59%	1.21%	(7.84)	1.18%	(7.54)	1.10%	(6.92)
9	hfcombo2	Industry mom. + Relative rev. + Season.	2016	Novy-Marx and Velikov,	NMVa	1.14%	(7.15)	3.69%	1.05%	(6.70)	1.11%	(6.99)	1.00%	(6.21)
10	idiovol	Idiosyncratic volatility	2003	Ali, Hwang, and Trombly,	NMV	0.41%	(2.04)	4.67%	0.61%	(3.19)	0.52%	(3.64)	0.33%	(2.35)
11	indmom1m	Industry momentum	1999	Moskowitz and Grinblat,	NMV	0.61%	(2.48)	5.68%	0.70%	(2.84)	0.74%	(3.00)	0.49%	(1.96)
12	invest	Investment	2004	Titman, Wei, and Xie,	NMV, SYY	0.43%	(3.25)	3.05%	0.52%	(4.09)	0.36%	(2.96)	0.29%	(2.35)
13	mom12m	Momentum	1990	Jegadeesh,	NMV, SYY	1.12%	(3.97)	6.59%	1.23%	(4.33)	1.46%	(5.23)	0.23%	(2.09)
14	netissue_a	Net issuance annual	1995	Ikenberry, Lakonishok, and Vermaelen,	NMVa	0.64%	(5.01)	2.95%	0.74%	(6.05)	0.64%	(5.38)	0.57%	(4.71)
15	netissue_m	Net issuance monthly	1995	Ikenberry, Lakonishok, and Vermaelen,	NMV, SYY	0.52%	(3.73)	3.26%	0.64%	(4.72)	0.57%	(4.45)	0.54%	(4.15)
16	netoa	Net operating assets	2004	Hirshleifer, Hou, Teoh, and Zhang,	SYY	0.32%	(1.95)	3.79%	0.18%	(1.12)	0.53%	(4.10)	0.44%	(3.40)
17	ohlson*	Ohlson's O-score	1980	Ohlson,	NMVa, SYY	0.18%	(1.06)	4.03%	0.32%	(1.91)	0.61%	(4.84)	0.50%	(3.96)
18	peadcar3*	PEAD(CAR3)	2008	Brandt et al.,	NMV	0.75%	(5.78)	3.00%	0.78%	(6.00)	0.83%	(6.38)	0.66%	(5.17)
19	peadsue*	PEAD(SUE)	1982	Rendelman, Jones, and Latane,	NMV	0.67%	(4.67)	3.32%	0.68%	(4.72)	0.84%	(6.04)	0.51%	(4.14)
20	piotroski	Piotroski's f-score	2000	Piotroski,	NMV	0.42%	(2.41)	4.09%	0.56%	(3.29)	0.63%	(3.96)	0.54%	(3.32)
21	profit	Gross profitability	2010	Balakrishnan, Bartov, and Faurel,	NMV, SYY	0.26%	(1.76)	3.39%	0.22%	(1.52)	0.38%	(2.67)	0.35%	(2.37)
22	relrev1m	Industry relative reversals	1994	Asness, Porter, and Stevens,	NMV	0.73%	(4.33)	3.90%	0.55%	(3.51)	0.50%	(3.13)	0.73%	(4.70)
23	relrev1mlow	Short-run reversals low volatility	2016	Novy-Marx and Velikov,	NMV	1.11%	(7.32)	3.52%	1.02%	(6.84)	0.91%	(6.11)	1.03%	(6.93)
24	rev1m	Short-run reversals	1990	Jegadeesh,	NMV	0.26%	(1.15)	5.24%	0.07%	(0.31)	-0.01%	(-0.05)	0.28%	(1.32)
25	rev60m	Long-run reversals	1987	DeBondt and Thaler,	NMVa	0.26%	(1.36)	4.50%	0.32%	(1.64)	-0.12%	(-0.79)	-0.10%	(-0.61)
26	roa*	Return-on-assets	2010	Chen, Novy-Marx, and Zhang,	NMVa, SYY	0.39%	(2.19)	4.13%	0.52%	(3.01)	0.79%	(5.24)	0.51%	(3.60)
27	roe*	Return-on-book equity	1996	Haugen and Baker,	NMV	0.39%	(2.45)	3.71%	0.57%	(3.85)	0.63%	(4.85)	0.42%	(3.39)
28	rome*	Return-on-market equity	1977	Basu,	NMVa	0.98%	(5.07)	4.47%	1.13%	(6.04)	0.95%	(5.27)	0.78%	(4.29)
29	seasonal	Seasonality	2008	Heston and Sadka,	NMV	0.64%	(3.84)	3.85%	0.56%	(3.38)	0.68%	(4.19)	0.60%	(3.61)
30	size	Size	1981	Banz,	NMV	0.28%	(1.50)	4.40%	0.18%	(0.95)	-0.02%	(-0.26)	-0.16%	(-2.68)
31	valmom	Value + Momentum	2014	Novy-Marx,	NMV	0.70%	(3.21)	5.05%	0.84%	(3.93)	0.59%	(2.83)	-0.26%	(-2.33)
32	valmomprof	Value + Mom + Prof	2014	Novy-Marx,	NMV	0.94%	(4.80)	4.57%	1.02%	(5.19)	0.94%	(4.81)	0.27%	(1.95)
33	valprof	Value + Profitability	2014	Novy-Marx,	NMV	0.67%	(3.96)	3.93%	0.76%	(4.57)	0.45%	(3.16)	0.47%	(3.24)
34	value	Value	1985	Rosenberg, Reid, and Lanstein,	NMV	0.38%	(2.21)	4.02%	0.48%	(2.78)	-0.08%	(-0.79)	-0.11%	(-1.09)
	E.A.R.	Equal-weigted Anomaly Return				0.55%	(9.78)	1.31%	0.61%	(11.39)	0.58%	(11.09)	0.41%	(10.49)

Table 1.16: Predictive Regression at Quarterly Level (E.A.R.) Adjusted with Benchmark

$$\text{E.A.R.}_{t+1} = a+b\times\text{CoAnomaly}_t+\sum_p m_p\times\text{Other.Predictors}_{p,t}+t\times\text{Trend}+\sum_j \beta_j\times Benchmark.Factors_{t+1}+e_{t+1}$$

The dependent variable is the Equal-weighted Anomaly Returns (E.A.R.) for the next quarter $t+1$ for 23 anomalies in Novy-Marx and Velikov (2016) (NMV), 11 anomalies in Stambaugh et al. (2012) (SYY), or a time-series combination of 27 anomalies (1963-1972) and 34 anomalies (post-1973) separately. CoAnomaly, average variance, and aggregate variance are also measured within 23 anomalies in Novy-Marx and Velikov (2016), 11 anomalies in Stambaugh et al. (2012), or a time-series combination, respectively. The standard deviations of all regressors are also normalized to 1 and returns are measured in percentage. The coefficients on the trend variable and the pre-1973 dummy are multiplied by 1000 for readability. The investor sentiment data starts from mid-1965 and ends in 2014. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

	Dependent Variable: Quarterly E.A.R. at t+1								
	23 Anomalies from NMV			11 Anomalies from SYY			Sample back to 1963		
	(1)	(6)	(7)	(1)	(6)	(7)	(1)	(6*)	(7*)
CoAnomaly t	0.65	0.54	0.57	0.66	0.37	0.38	0.73	0.57	0.59
	(4.18)	(3.09)	(3.11)	(2.61)	(1.33)	(1.35)	(3.92)	(3.01)	(2.91)
Average Var. t		0.37	0.35		0.14	0.41		0.51	0.62
		(1.32)	(1.07)		(0.25)	(0.66)		(1.68)	(1.70)
Aggregate Var. t		-0.30	-0.24		-0.06	-0.38		-0.31	-0.37
		(-1.00)	(-0.76)		(-0.12)	(-0.71)		(-0.98)	(-1.06)
Anomaly Value Spread t		0.11	0.08		-0.15	-0.15		-0.03	-0.05
		(0.65)	(0.45)		(-0.47)	(-0.44)		(-0.17)	(-0.25)
Sentiment t		0.66	0.41		1.25	1.55		0.94	0.88
		(2.41)	(1.26)		(2.27)	(2.54)		(2.33)	(1.79)
Market Avg. Corr.t			-0.01			0.01			0.00
			(-0.87)			(0.20)			(-0.21)
E.A.R. t			-0.02			-0.49			-0.07
			(-0.14)			(-1.65)			(-0.41)
MktRf t			-0.23			-0.17			-0.11
			(-1.22)			(-0.50)			(-0.55)
TED Rate t			-0.16			-0.51			
			(-0.77)			(-1.27)			
Trend T	-0.02	-0.02	-0.02	0.01	0.05	0.04	-0.04	-0.02	-0.03
	(-1.52)	(-1.10)	(-0.81)	(0.46)	(1.29)	(0.75)	(-2.65)	(-1.29)	(-1.47)
Pre-1973 Dummy							-0.06	-0.05	-0.06
							(-1.89)	(-1.81)	(-1.83)
MktRf t+1	-0.10	-0.08	-0.07	-0.16	-0.14	-0.14	-0.12	-0.10	-0.10
	(-4.61)	(-3.51)	(-3.30)	(-3.96)	(-3.29)	(-3.11)	(-5.81)	(-4.90)	(-4.75)
SMB t+1	0.00	-0.01	-0.02	-0.19	-0.19	-0.20	-0.09	-0.11	-0.11
	(0.02)	(-0.48)	(-0.64)	(-4.22)	(-4.15)	(-4.24)	(-3.20)	(-3.78)	(-3.80)
HML t+1	0.22	0.20	0.21	0.04	0.01	0.00	0.18	0.16	0.16
	(8.34)	(7.14)	(6.87)	(0.88)	(0.24)	(0.07)	(6.08)	(5.25)	(4.90)
UMD t+1	0.23	0.22	0.23	0.31	0.31	0.30	0.20	0.20	0.20
	(11.76)	(11.00)	(10.75)	(9.38)	(8.89)	(8.06)	(9.47)	(9.33)	(8.89)
Adj. R square	73.0%	75.4%	76.4%	69.8%	72.1%	73.9%	68.1%	70.8%	72.0%
N	179	170	118	179	170	118	217	200	200

Table 1.17: Predictive Regression: Long leg and Short Leg

$$\text{E.A.R.leg}_{t+1} = a + b \times \text{CoAnomaly}_t + \sum_p m_p \times \text{Other.Predictors}_{p,t} + t \times \text{Trend} + \sum_j \beta_j \times Benchmark.Factors_{t+1} + e_t$$

The dependent variable is the long and short legs of Equal-weighted Anomaly Returns (E.A.R.) for the next quarter $t+1$. All independent variables are measured in the quarter t. CoAnomaly is the average partial correlation for the whole long-short portfolio of 34 stock market anomalies. Average realized variance is equally averaging the realized daily variances for the 34 stock market anomalies. The standard deviations of all regressors are also normalized to 1 and returns are measured in percentage. The coefficient on the trend variable is multiplied by 1000 for readability. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

	Dependent Variable: Quarterly E.A.R. at t+1					
	Long leg			Short leg		
	(L1)	(L5)	(L7)	(S1)	(S5)	(S7)
CoAnomaly t	0.24	0.29	0.36	-0.54	-0.51	-0.39
	(2.23)	(2.51)	(2.87)	(-3.63)	(-3.18)	(-2.18)
Average Var. t		0.33	0.38		-0.15	-0.17
		(1.78)	(1.86)		(-0.60)	(-0.57)
Aggregate Var. t		-0.27	-0.38		-0.05	0.03
		(-1.47)	(-1.91)		(-0.19)	(0.10)
Anomaly Value Spread t			0.24			0.21
			(2.24)			(1.35)
Sentiment t			0.18			-0.51
			(0.88)			(-1.74)
Market Avg. Corr.t			-0.01			0.00
			(-1.19)			(-0.17)
E.A.R. t			-0.16			0.01
			(-1.48)			(0.07)
MktRf t			0.05			0.24
			(0.41)			(1.37)
TED Rate t			-0.02			0.12
			(-0.13)			(0.64)
Trend T	-0.03	-0.03	-0.03	0.01	0.02	0.01
	(-2.66)	(-2.83)	(-2.14)	(1.06)	(1.20)	(0.36)
MktRf t+1	0.97	0.97	0.97	1.13	1.12	1.12
	(72.62)	(68.92)	(69.66)	(61.20)	(57.12)	(55.81)
SMB t+1	0.11	0.11	0.11	0.19	0.20	0.20
	(6.84)	(6.40)	(6.27)	(8.12)	(8.29)	(8.10)
HML t+1	0.10	0.10	0.08	-0.10	-0.10	-0.10
	(5.79)	(6.00)	(4.08)	(-4.06)	(-4.07)	(-3.52)
UMD t+1	0.04	0.04	0.03	-0.14	-0.15	-0.15
	(3.13)	(3.04)	(2.21)	(-8.44)	(-8.51)	(-7.86)
Adj. R square	98.0%	98.2%	98.3%	98.0%	98.1%	98.2%
N	179	170	118	179	170	118

Table 1.18: Predictive Regression: Different Horizons

The dependent variable is the Equal-weighted Anomaly Returns (E.A.R.) for the next 1 month or 6 months. All independent variables are measured in the quarter t, so there is overlapping data. CoAnomaly is the average partial correlation for the whole long-short portfolio of 34 stock market anomalies. Anomaly Value Spread is the average value spread for all anomalies. The coefficients are normalized to the quarterly specification as in Table 1.3, so they can be compared with each other. The standard deviations of all regressors are also normalized to 1 and returns are measured in percentage. The coefficient on the trend variable is multiplied by 1000 for readability. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

Dependent Variable: Cumulative Equal-weighted Anomaly Returns (E.A.R.) in the future									
	E.A.R. in next 1 month			E.A.R. in next 6 months			E.A.R. in next 12 months		
	(1)	(2)	(3)	(4)	(5)	(6)	(4)	(5)	(6)
CoAnomaly t	0.44	0.46	0.37	0.43	0.44	0.33	0.32	0.26	0.15
	(2.81)	(2.76)	(2.23)	(3.06)	(3.04)	(2.33)	(2.92)	(2.33)	(1.49)
Average Var. t		0.91	0.74		0.56	0.57		0.28	0.30
		(1.20)	(0.98)		(1.87)	(2.01)		(1.17)	(1.37)
Aggregate Var. t		-0.32	-0.42		-0.21	-0.37		0.07	-0.10
		(-0.46)	(-0.61)		(-0.70)	(-1.33)		(0.31)	(-0.49)
Anomaly Value Spread t			0.17			-0.03			0.21
			(0.97)			(-0.48)			(2.01)
Sentiment t			0.44			0.51			0.35
			(2.41)			(3.14)			(2.85)
Trend T	-0.03	-0.03	-0.03	-0.04	-0.04	-0.02	-0.04	-0.04	-0.03
	(-2.65)	(-2.79)	(-1.96)	(-2.33)	(-2.55)	(-1.53)	(-3.17)	(-3.25)	(-2.92)
Other Predictors	N	N	Y	N	N	Y	N	N	Y
Carhart-4 Factors t+1	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adj. R square	61.5%	62.5%	63.4%	67.5%	68.3%	71.6%	72.2%	75.5%	76.9%
N	539	539	511	534	534	506	528	528	500

Table 1.19: Predictive Regression: Mean-Variance Efficient Portfolio

$$\text{R.MVE}_{t+1} = a + b \times \text{CoAnomaly}_t + \sum_p m_p \times \text{Other.Predictors}_{p,t} + t \times \text{Trend} + \sum_j \beta_j \times Benchmark.Factors_{t+1} + e_{t+1}$$

The dependent variable is the in-sample mean-variance efficient portfolio return (MVE) for the next quarter $t+1$. All independent variables are measured in the quarter t. CoAnomaly is the average partial correlation for the whole long-short portfolio of 34 stock market anomalies. Average realized variance is equally averaging the realized daily variances for the 34 stock market anomalies. Aggregate variance of the E.A.R. is measured as the variance of daily returns. The coefficient on the trend variable is multiplied by 1000 for readability. T-stats, shown in parentheses, are computed with Newey and West (1987) correction for 4 lags.

	Dependent Variable: Quarterly MVE at t+1					
	(1)	(2)	(3)	(4)	(6)	(7)
CoAnomaly t	0.52	0.62	0.69	0.60	0.66	0.68
	(2.67)	(3.03)	(3.39)	(3.51)	(3.70)	(3.59)
Average Var. t		0.80	0.94		0.70	0.84
		(2.18)	(2.58)		(2.20)	(2.49)
Aggregate Var. t		-0.53	-0.89		-0.43	-0.65327
		(-1.61)	(-2.79)		(-1.48)	-2.1983
Anomaly Value Spread t			0.24			-0.65
			(1.31)			(-2.20)
Sentiment t			0.90			0.44
			(2.76)			(1.43)
Market Avg. Corr.t			-0.01			-0.01
			(-0.65)			(-0.71)
E.A.R. t			-0.28			-0.27
			(-1.53)			(-1.61)
MktRf t			0.32			-0.02
			(1.62)			(-0.08)
TED Rate t			-0.52			-0.58
			(-2.31)			(-2.83)
MktRf t+1				-0.01	-0.02	0.01
				(-0.56)	(-0.67)	(0.25)
SMB t+1				0.01	0.00	-0.01
				(0.50)	(0.07)	(-0.42)
HML t+1				0.13	0.12	0.09
				(4.72)	(4.19)	(3.31)
UMD t+1				0.08	0.09	0.09
				(4.37)	(4.38)	(4.21)
Trend T	-0.08	-0.09	-0.09	-0.07	-0.07	-0.09
	(-4.65)	(-4.88)	(-3.82)	(-4.53)	(-4.73)	(-3.98)
Adj. R square	17.7%	19.1%	27.6%	39.7%	43.0%	46.9%
N	179	179	118	179	179	118

Table 1.20: Robustness: Correlation between Two Mispricing Factors and 1-month CoAnomaly

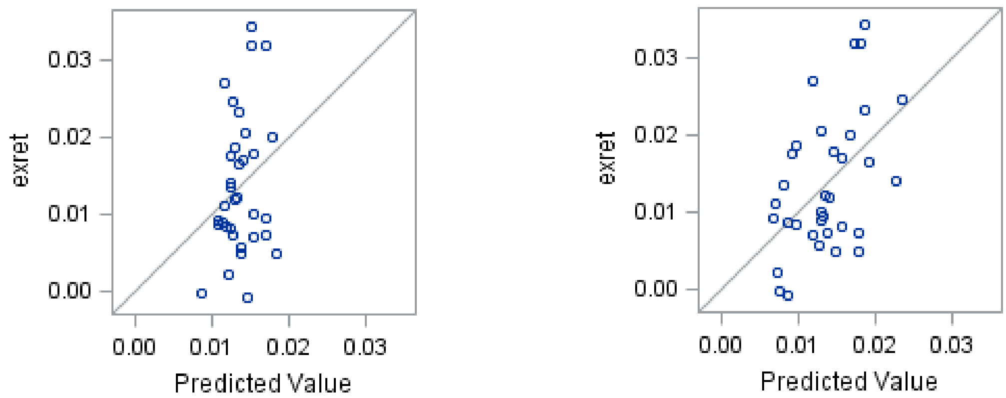
This table reports the robustness check results for Table 1.4. Instead of using the three-month CoAnomaly measure, I use the correlation between two mispricing factors in Stambaugh and Yuan (2016), or the one-month CoAnomaly measure.

HF Ret. Group	**Corr. Group**	No. Months	HF Ret. t	CoAnomaly t	CoAnomaly t+1	**E.A.R. t+1**	**E.A.R. t+3**	**E.A.R. t+6**
Panel A1: sorting on Correlation of two Mispricing Factors								
	1	79	0.49%	0.11	0.13	0.18%	0.67%	1.93%
			(4.26)	(9.09)	(11.65)	(1.17)	(2.26)	(4.35)
	2	105	0.62%	0.16	0.16	0.30%	0.81%	1.68%
			(6.67)	(17.77)	(16.15)	(1.83)	(2.44)	(3.10)
	3	80	0.41%	0.22	0.20	0.64%	1.84%	3.46%
			(3.67)	(16.53)	(13.25)	(2.39)	(4.05)	(4.84)
	Diff 3-1		-0.08%	0.11	0.07	0.46%	1.17%	1.54%
			(-0.51)	(6.43)	(3.87)	(1.48)	(2.14)	(1.82)
Panel A2: First sort all months based on HF Ret., and then sort on Correlation of two Mispricing Factors								
	1	23	-0.54%	0.15	0.16	0.03%	-0.22%	0.28%
1	2	32	-0.40%	0.19	0.18	-0.28%	-0.12%	-0.62%
	3	24	-0.38%	0.24	0.19	0.90%	2.03%	4.12%
	Diff 3-1					0.88%	2.25%	3.84%
						(1.45)	(1.70)	(2.04)
Panel B1: sorting on the 1-month CoAnomaly								
	1	79	0.58%	0.09	0.12	-0.01%	0.21%	0.98%
			(5.21)	(9.45)	(9.92)	(-0.07)	(0.64)	(1.88)
	2	105	0.48%	0.15	0.17	0.29%	1.15%	1.87%
			(4.84)	(16.73)	(15.83)	(1.78)	(3.62)	(3.63)
	3	80	0.51%	0.23	0.19	0.84%	1.83%	4.14%
			(4.74)	(19.49)	(15.08)	(3.19)	(4.14)	(6.24)
	Diff 3-1		-0.07%	0.14	0.08	0.85%	1.62%	3.16%
			(-0.39)	(9.01)	(4.32)	(2.79)	(2.93)	(3.75)
Panel B2: First sort all months based on HF Ret., and then sort on 1-month CoAnomaly								
	1	23	-0.50%	0.13	0.16	-0.81%	-1.17%	-0.92%
1	2	32	-0.40%	0.17	0.16	0.31%	0.83%	0.46%
	3	24	-0.41%	0.29	0.21	0.95%	1.75%	3.90%
	Diff 3-1					1.76%	2.92%	4.82%
						(2.47)	(2.08)	(2.34)

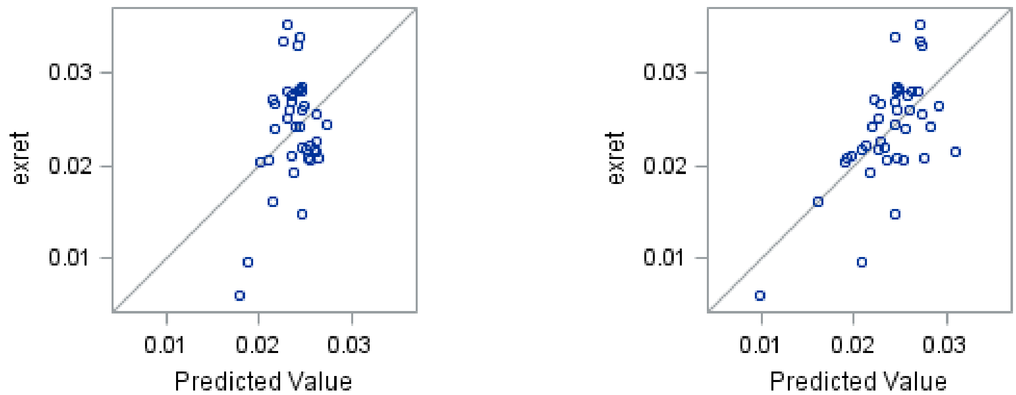
Table 1.21: Weigted Anomaly Score

This table reports the time-series properties of the Weighted Anomaly Score (WAS) for the long legs and the short legs of 34 stock market anomalies. WAS for the short leg (WAS_s) is calculated as the following: for each stock at each point of time (end of a month), I first calculate the short anomaly score (how many strategies / anomalies are shorting it), then divide it with the total number of anomalies (34 in my case) to normalize the score to one. I calculate the weighted average anomaly score for each anomaly and then take the simple mean across all anomalies. This procedure is repeated in each month, so a time series is generated. The same procedure can be done on the long leg to get the WAS for the long leg (WAS_l).

	WAS_l	WAS_s
MEAN	0.157	0.179
STD	0.009	0.017
Correlation with		
WAS_l	1	-0.375
WAS_s	-0.375	1
CoAnomaly_LS	-0.193	0.279
CoAnomaly_L	0.16	0.02
CoAnomaly_S	-0.315	0.406

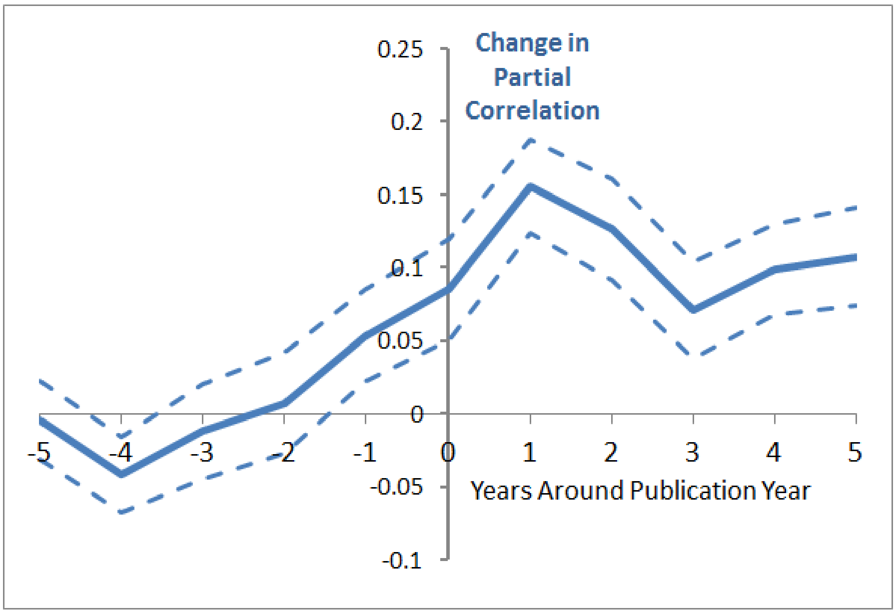


Anomaly Portfolios: This figure plots the quarterly realized mean returns of 34 anomaly portfolios against the predicted mean excess returns. In the left figure, I use CAPM to predict portfolio returns, and in the right figure, I use CAPM + CoAnomaly.



Standard Portfolios: This figure plots the quarterly realized mean excess returns of 40 equity portfolios (25 size- and book-to-market-sorted portfolios, 10 momentum-sorted portfolios, and 5 industry portfolios) against the predicted mean excess returns. In the left figure, I use CAPM to predict portfolio returns, and in the right figure, I use CAPM + CoAnomaly.

Figure 1.3: Realized versus Predicted Returns: Comparing CAPM versus CAPM + CoAnomaly.



This figure plots the average partial correlation (together with its confidence interval) relative for its all-sample mean for each anomaly around its academic publication year, with -5 to 5 years window, averaging across 27 anomalies. The partial correlations for each anomaly are calculated with respect to other post-publication strategies up to each year. Dashed lines indicate the 90% confidence interval.

Figure 1.4: Partial Correlation Change around the Publication Year

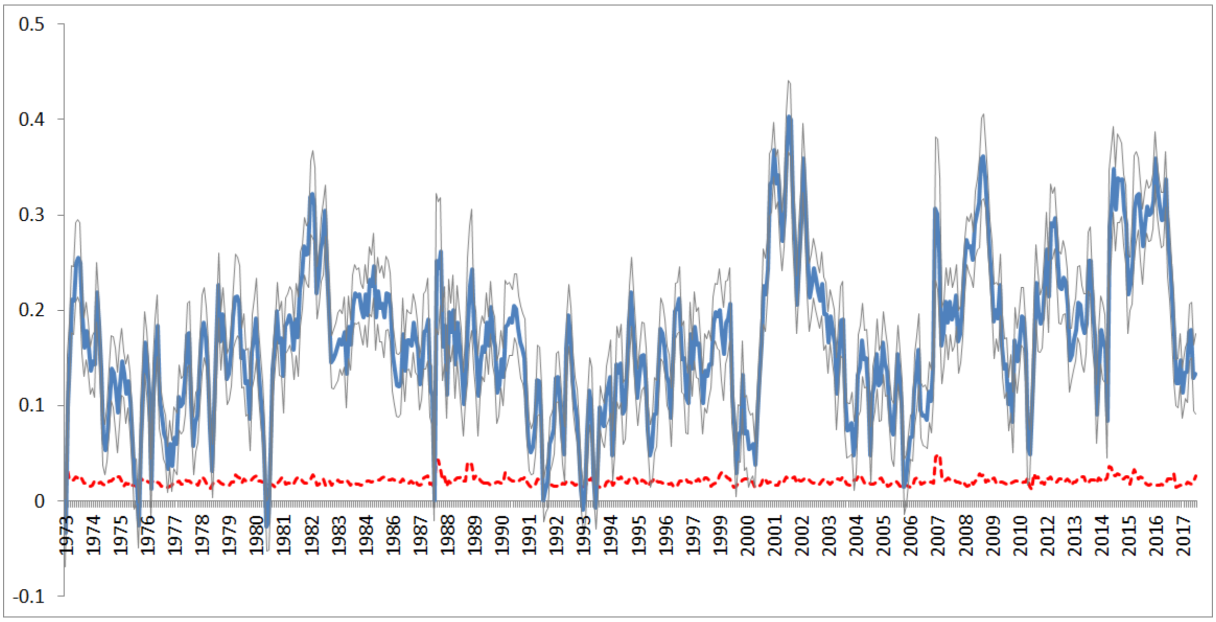


Figure 1.5: CoAnomaly (blue solid) and its Bootstrap Standard Error (red dashed, 10000 times)

Chapter 2

Decomposing Momentum Spread

Since momentum arbitrage activity, buying winners and selling losers, effectively enlarges the return spread between these two groups, I find that the momentum spread (the difference of the formation-period recent 6-month returns between winners and losers) negatively predicts future momentum profit in the long-term, but not in the following month. I further decompose the momentum spread into the spreads of old or young momentum stocks based on whether a stock has been identified as a momentum stock for more than three months. I show that the negative predictability is mainly driven by the old momentum spread. For the top 20% of the sample period associated with the highest values of old momentum spread, the momentum reversals happen sooner (only six months after formation) and stronger (more than 120 basis points per month from month 7 to month 24 after formation), relative to negligible momentum reversals observed following the bottom 20% period with low old momentum spread. As these old momentum stocks are more likely to be exploited by arbitrageurs, these findings suggest that momentum is amplified by arbitrage activity and excessive arbitrage destabilizes asset prices and generates strong reversals.

JEL-Classification: G12, G14

Keywords: Momentum, Return Spread, Underreaction and Overreaction, Destabilizing Mechanism of Excessive Arbitrage.

2.1 Introduction

Among all cross-sectional anomalies in the stock market, momentum is one of those that have been studied extensively but remains difficult to be justified in a rational model. The main reason is that it is difficult to explain two intertwisted phenomena, both the mid-term momentum and long-term reversal, in a single rational framework. Above that, momentum profits also shows strong time-series predictability (see Cooper et al. (2004), Stivers and Sun (2010) and Barroso and Santa-Clara (2015) among the others).

Another important feature of momentum is its vulnerability to destabilizing excessive arbitrage since arbitrageurs buy stocks whose prices rise (winners) and sell stocks whose prices fall (losers). The arbitrage activity will push the winners' price even higher and the losers' price even lower, which may attract more momentum arbitrageurs given that newcomers cannot distinguish the momentum price pattern coming from fundamental information or previous arbitrage activity. This loop is a typical case of *positive feedback*. If the change in form fundamental does not catch up with this speed, this will create a momentum bubble and bust since prices cannot deviate from fundamentals without a limit and will go back to the fair value in the long-term. Following this logic, excessively strong arbitrage activity in momentum will, first, drive up the momentum spread, which is the formation-period return difference between past winners and losers, and second, create strong reversals.

Is it the time-varying crowding in momentum arbitrage activity, captured by the momentum spread, driving the time-variation in momentum profit? My answer is yes. Based on this question, I first show that momentum spread *negatively* predicts momentum returns. Though similar empirical facts have been documented in the literature (see Stivers and Sun (2010) and Huang (2015)), I find that this negative predictability emerges mostly in the *long-term*, i.e. six months after the formation period, but not in the short-term. I argue that this is due to the bubble nature of the momentum effect. If mid-term momentum and long-term reversal result from the combination of *both* underreaction and overreaction, timing the switch between these two states, or equivalently timing the peak of momentum bubble, should be difficult. However, the long-term reversals will emerge ultimately and this is the part I focus on. This predictive effect is also robust to controlling other benchmark factors. To close the circle, I also provide evidence that the momentum spread is linked to the amount of momentum arbitrage capital.

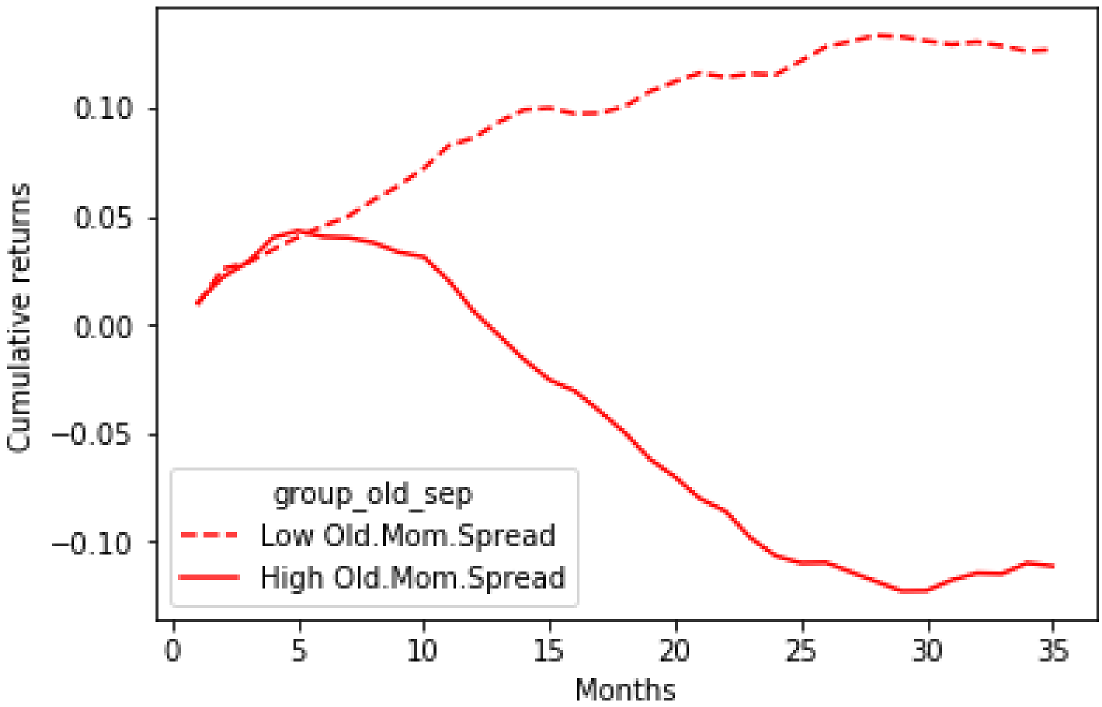


Figure 2.1: Different Momentum Strategy Performance

Time-series average of the cumulative returns for momentum portfolios (no balancing after formation) up to 60 months after their formation periods associated with high or low old momentum spreads. The dashed line shows the averaged momentum cumulative performance following months associated with a low old momentum spread (top 20%). The solid line shows the averaged momentum cumulative performance following months associated with a high old momentum spread (top 20%).

(Insert Figure 2.1)

I further decompose the momentum spread into old momentum spread and young momentum spread and provide evidence that the negative predictability is mainly driven by the old momentum spread. For every period, I classify all momentum stocks (separating winners and losers) into two subgroups based on how long they have been identified as a momentum stock. Old (Young) momentum spread is defined as the momentum spread within old (young) momentum stocks, which have been identified as momentum stocks for more than (less or equal to) three months. Due to the high correlation among spread-based predictors, I conduct a within-group-controlled analysis to show that old momentum spread has the strongest predictive power within the spread-based family. I use a horse-race regression to show the strong predictability of old momentum spread after controlling several well-known momentum predictors, including market state and volatility among the others. These empirical findings connect the literature on the time-series momentum signal (see Novy-Marx (2012)) and cross-sectional momentum stock composition (see Chen et al. (2009)).

To address why the long-term predictability of old momentum spread is so strong, I argue that it contains better information about the momentum arbitrage activity. Presumably, old momentum stocks are more likely to be traded by arbitrageurs because they need to observe momentum characteristics first. I provide evidence by first showing that the young and old momentum stocks are different: young momentum stocks show higher loadings to fundamental-driven variation, while the old momentum stocks are more sensitive to VIX and suffer unconditional sooner and stronger reversals going forward. I also show that momentum strategy with old momentum stocks, first, revert sooner and stronger than young momentum strategy unconditionally, and second, this reversal pattern can be strongly predicted by old momentum spread conditionally.

This mechanism is indirectly modeled by Hong and Stein (1999), which is so far one of the studies that successfully explain the mid-term momentum and long-term reversal in a unified model. In their model, the existence of momentum traders is the key to the result. These arbitrageurs will trade old momentum stocks more due to the *positive feedback*, so the information in old momentum spread would help me identify the *Destabilizing Mechanism of Excess Arbitrage*. I provide empirical evidence which is consistent with several implications of their paper. Among all momentum spread-based measures, the old momentum spread has the highest predictive power and generates a monotonic pattern across time. Overall, my findings suggest that the momentum effect is initiated by fundamental news (for young momentum stocks) and amplified by arbitrage activity (for old momentum stocks), which echoes the findings in Novy-Marx (2015) and Gargano and Rossi (2018).

Finally, I present some results of using the old momentum spread to predict returns of a group of 'contrarian' strategies, including value, low-beta and low volatility. The logic follows that if momentum bubble busts, I expect arbitrageurs to allocate capital to other strategies, which generates price impacts.

Related Literature The cross-sectional determinants of momentum have been widely studied in the last two decades and the time-variation of momentum is less explored, but recently it has caught more attention. Among them, Cooper et al. (2004) find that momentum premium falls when the market return in the past three years is low. Stivers and Sun (2010) find that momentum premium is negatively related to the recent cross-sectional dispersion and Wang and Xu (2015) find that market volatility has significant power to negatively forecast momentum payoffs. Instead of focusing on the market, Barroso and Santa-Clara (2015) find that the volatility of momentum is highly variable over time and predictable and it also predicts momentum returns. Daniel and Moskowitz (2016) document infrequent but repeating crashes of momentum strategy and link this phenomenon to significant time-varying exposure to systematic factors, which is explored by work such as Kothari and Shanken (1992) and Grundy and Martin (2001).

Arbitrageurs can contribute to higher liquidity and informational efficiency in the market, but they can also cause instability by taking high leverages and overcrowding as Stein (2009) points out. Jacobs and Levy (2014) document some stylized facts of smart beta investing, and also point out that overcrowding may lead to overvaluation and factor crash. Among all strategies, momentum is most prone to excess arbitrage since arbitrageurs are buying the winners and selling the losers, which will produce a price impact and broaden the momentum spread. If the arbitrageurs cannot infer how much capital is trading on this momentum-type '*unanchored*' strategy by Stein (2009), they may create a bubble of excessive arbitrage, which leads to the long-term reversal of the strategy, and may destabilize the market subsequently to a bigger extent. Barroso et al. (2017) show that this crowding mechanism can explain momentum tail risk with myopic arbitrageurs. In contrast, an '*anchored*' strategy like value, arbitrageurs are buying the high book-to-market (B/M) ratio stocks and selling the low book-to-market (B/M) ratio stocks, and pushing the value spread lower, which is a natural anchor for the value strategy.

Proxy the crowdedness of momentum arbitrage activity, several attempts have been made. Lou and Polk (2013) infer arbitrage activity from return correlations and find that the average pairwise correlation predicts reversals. Huang (2015) use momentum gap, the difference between the 75th and 25th percentiles of the distribution of cumulative stock returns, to proxy arbitraging activity and find that higher momentum activity is followed by stronger reversals.

I consider my result different from and more robust than Huang (2015) for two reasons: first, his result of predicting next-month momentum return is not robust in my sample, and as I argue the main reason for this could be the imperfect timing capability of the peak of momentum bubble; second, the momentum is not a linear strategy, trading the decile momentum stocks shows much higher return than trading quintile momentum stocks[1], so by construction, momentum arbitrage will have a larger effect on the momentum spreads between top decile and bottom decile.

Novy-Marx (2012) finds that momentum is primarily driven by the stocks' performances in the twelve to seven months from portfolio formation, which supports my choice of focusing on the recent performance spread of momentum stocks. Some other research supports the idea of decomposing momentum stocks in the cross-section: Chen et al. (2009) find that a two-way sorting based on long-term (two-year) and recent performance (one-year) can accommodate both momentum and long-term reversal effects by distinguishing between fresh and stale winners and losers, which is consistent with the story that investors mistakenly respond to shocks to firm fundamentals as if they are going to continue in the long run. Daniel et al. (2017) propose a 'Betting Against Winners' strategy that goes short the overpriced winners and long other winners generates a Sharpe-ratio of 1.08, due to the disagreement and short-sale constraints.

2.2 Momentum Spread

Given that momentum arbitrage will push winners and losers far away from each other, effectively broadening the past performance difference between these two groups of stocks, I first check whether this time-varying momentum spread can predict future momentum performance. I find that these spread-type measures have limited power of predicting momentum reversals in the short-term. However, the predictability emerges only if we focus on the long-term momentum performance.

[1] This can be easily checked by simply comparing the average return of UMD factor (top 30% - bottom 30%) and MOM factor (top 10% - bottom 10%) on French's website.

I show that if the momentum spread is particularly large, the momentum strategy will reverse sooner and stronger going forward.

2.2.1 Data and Spread Construction

I use the stock return data from the Center for Research in Security Prices (CRSP) and accounting information data from Compustat - Capital IQ. I include only stocks for companies listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11. To make sure my results are not driven by micro-cap stocks and other microstructure issues, I exclude stocks with prices below $5 per share or are in the bottom NYSE size decile. At the end of each month, I sort all remaining stocks into deciles based on the return in the last twelve months (skipping the most recent month, i.e. t-12 to t-2). The momentum strategy return is the value-weighted portfolio return difference between the top decile and bottom decile. As for the case of stock delisting, I follow Green et al. (2017) to add delisting returns back to returns in the spirit of Shumway and Warther (1999). That is, for the firms with dlstcd = 500 or dlstcd within [520, 584], I use dlret = -0.35 if exchcd = 1 or 2 and dlret = -0.55 if exchcd = 3.

I also download the hedge fund asset under management AUM data from the BarclayHedge. I focus on two categories: all hedge funds and the Equity Market Neutral[2]. Fama-French factors are from Kenneth French website[3].

Momentum spread I define the momentum spread as the difference of the weighted lagged *6-month* cumulative returns between the top and bottom decile stocks, and the weight is the same lagged market capitalization weighting the returns.

$$Mom.Spread_{t-1} = \overline{Past.Ret_{t-6,t-1,Winners}} - \overline{Past.Ret_{t-6,t-1,Losers}}.$$

[2]On the BarclayHedge website, they state the Equity Market Neutral as: *This investment strategy is designed to exploit equity market inefficiencies and usually involves being simultaneously long and short matched equity portfolios of the same size within a country. Market neutral portfolios are designed to be either beta or currency neutral, or both. Well-designed portfolios typically control for industry, sector, market capitalization, and other exposures. Leverage is often applied to enhance returns.*

[3]I show my sincere gratitude to Ken French for supplying the Fama-French factors.

where $\overline{Past.Ret_{t-6,t-1,Winners}}$ denotes the value-weighted average cumulative returns in the past six months for stocks in the winner decile at the end of each month t-1 and $\overline{Past.Ret_{t-6,t-1,Losers}}$ denotes the value-weighted average cumulative returns in the past six months for stocks in the loser decile at the end of each month t-1.

My momentum spread is not defined as the return spreads of the typical formation periods returns. The formation period momentum spread can be defined as $F.Spread_{mom} = \overline{Past.Ret_{t-12,t-2,Winners}} - \overline{Past.Ret_{t-12,t-2,Losers}}$, which is the difference of the weighted lagged 12-month cumulative return (skipping the most recent month) between the winners and losers. Note that, because of this difference in the definitions, the momentum spread might be negative for some periods since they are not spreads of the formation-period returns, which are strictly positive by construction.

I define the spread based on recent six-month returns mainly for two reasons: first, hypothetically, the spread in the recent periods is more likely to capture the effect from arbitraging activity as arbitrageurs need time to observe the momentum characteristics and then begin to trade momentum stocks, which may happen in the early half in the formation period; second, empirically, Novy-Marx (2012) shows that the momentum is primarily driven by firms' performance 12 to seven months prior to portfolio formation. On the other hand, Novy-Marx (2015) also shows that price momentum can be explained by the momentum in the change of firm fundamentals. So, in a relative sense, the price information in the first half of formation period is more fundamental-driven; on the other side of the coin, the price information in the second half is more likely to be none-fundamental-driven.

Another advantage of focusing the spreads in the periods different from the exact momentum formation periods is that it generates dispersion among the momentum stocks in the same batch of formation, i.e. the old momentum stocks and the young momentum stocks which I will discuss later. This allows me to bring the information in the cross-section of momentum stocks into the time-series predictability.

(Insert Figure 2.2)

Figure 2.2 plots the time-series of momentum spread. It is relatively stable and persistent, and its time-series average is 55.3% with a standard deviation of 27.3%

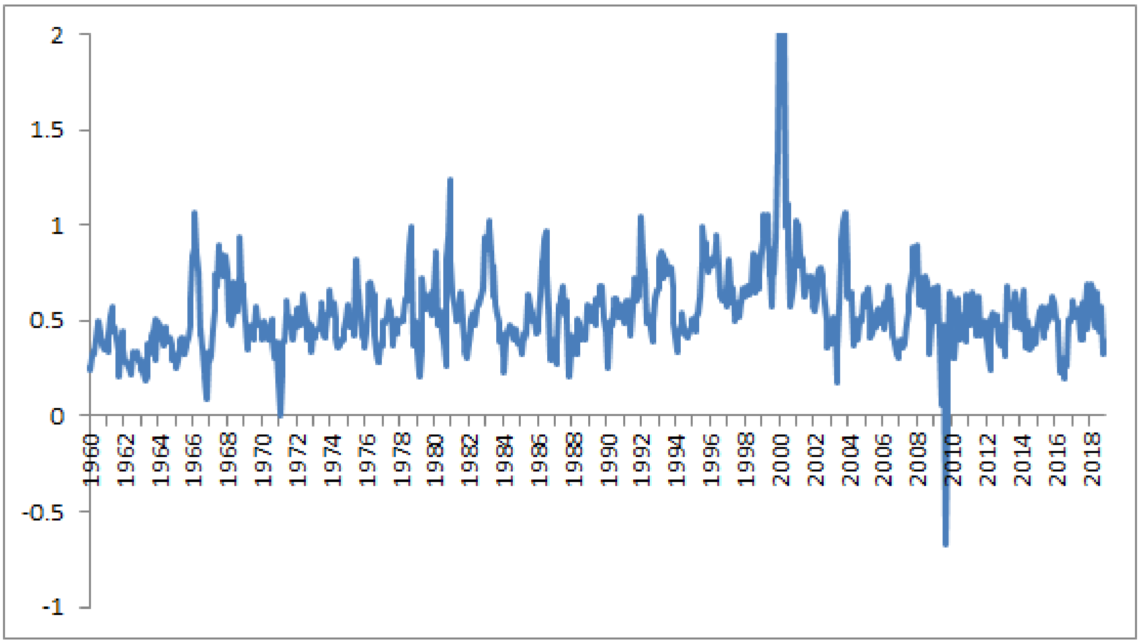


Figure 2.2: Time-series of Momentum Spread

as reported in Table 2.6. We can see that there is a peak in the 2000 dot-com bubble and it reaches a negative value after the financial crisis in 2008.

Formation Gap and Formation Spread I also construct two other spread-based measure, the formation period momentum gap (formation gap, for the rest of the paper) and the formation period momentum spread (formation spread, for the rest of the paper), for comparison in my analysis. Formation gap is the defined as Huang (2015) proposed, the difference between the 75th and 25th percentiles of the distribution of lagged 12-month cumulative return (skipping the most recent month, i.e. from month t-12 to t-2). Formation spread is the difference of the weighted lagged 12-month cumulative return (skipping the most recent month) between the winners and losers, as defined before.

2.2.2 Predictability of the Momentum Spread

I analyze different sets of monthly momentum portfolio returns based on how far they are from the formation period. Specifically, I study four sets of monthly returns: month 1-6, month 7-12, month 12-24, and month 25-36. It allows me to observe the heterogeneity in predicting momentum strategy by separating the future momentum performance.

Short-term Predictability: Negligible I first examine whether the momentum profits are predictable in the very short-run and track the profits on the momentum strategy over the next month subsequent to portfolio formation. All months are classified into five groups based on their different time-series measures. I consider four spread-type predictors, which are momentum spread, old momentum spread (constructed with old momentum stocks only, see section 2.3.1), formation gap and formation spread, as well as loser comomentum from Lou and Polk (2013)[4].

(Insert Table 2.1)

Table 2.1 shows the predictability of these five predictors. These results show that all four spread-based predictors have no strong predictability for momentum profits in the next month after portfolio formation and only the comomentum has a strong predictive power in the short-term.

This is not surprising considering that it is difficult to time the peak of momentum bubble. By trading momentum strategy, arbitrageurs are pushing the momentum spreads larger, which will be captured by these spread-based predictors. Excessive momentum arbitrage activities may push the winner stocks too high and loser stocks too low so that they will reverse *in the long-term*, as long as their fundamental does not change with the same magnitude. However, *in the short-term*, it is very difficult to time the peak of excessive trading and momentum bubble, so this predictive power will not show up.

Table 2.1 also shows the 1-month predictability result for two subsample periods, from 1960 to 1989 and from 1990 to 2018. The predictability of the formation gap documented by Huang (2015), if any, is mostly concentrated in the first half of the sample. If this predictability is indeed driven by excessive momentum arbitrage as he argues, this result is difficult to interpret given the fact that in the second half of my sample, momentum is publicly discovered by Jegadeesh and Titman (1993) and the professional managers are dominating the market.

However, the predictability of comomentum is stronger in the second half of my sample (with the discovery of momentum strategy and the growth in professional

[4] I are using the comomentum (the average pairwise partial return correlation) in the loser decile following their main analysis. Thanks to Dong Lou and Christopher Polk for sharing the time-series of comomentum.

arbitrageurs), which is consistent with Lou and Polk (2013)'s interpretation of inferring arbitrage activity from correlation among momentum stocks.

Long-term predictability: Strong

(Insert Table 2.2)

When looking into the long-term, Table 2.2 uses momentum spread to sort all months into groups. I show that in the short-term (within six months after the momentum portfolio formation), there is no consistent pattern across groups. However, I find strong predictive power for momentum strategy returns using momentum spread after six months following the formation of the momentum portfolio. Table 2.2 also shows that this predictive pattern preserves if I use other spread-type predictors, like formation gap or formation spread, though with a much smaller magnitude or a smaller significance level. Lastly, I show that the predictability of comomentum is strongest in the short-term.

In a nutshell, the momentum spread-type measures negatively predict the momentum portfolio returns in the long-term, but *not* in the short-term.

Table Details Here, I show how I construct the result statistics for all time-series sorting tables (Table 2.1, Table 2.2, and similar tables in following sections). All months are sorted into five groups based on their values of one time-series measure (for example, momentum spread) realized at the formation period. Then I create dummies for these months based on which group they are assigned to. Then I regress a set of future monthly returns (for example, all the monthly returns between 7 to 12 months after the formation) on the group dummies as well as contemporaneous Fama-French three factors. The average returns reported in each column in the tables are the regression coefficients on different dummies, $\alpha_{s,1}, \alpha_{s,2}, \alpha_{s,3}, \alpha_{s,4}, \alpha_{s,5}$, for each future set s (for example, 'month 1-6' means all the months within first half an year after the momentum formation).

$$r_{t,t+i} = \alpha_{s,1}\mathbf{1}_1+\alpha_{s,2}\mathbf{1}_2+\alpha_{s,3}\mathbf{1}_3+\alpha_{s,4}\mathbf{1}_4+\alpha_{s,5}\mathbf{1}_5+\beta_{M,s}\times MktRf_{t+i}+\beta_{S,s}\times SMB_{t+i}+\beta_{H,s}\times HML_{t+i}$$

for all formation months t and all actual months $t+i$ where $i \in s$ (for example, I treat each month between 7 to 12 months after each momentum formation as an observation). T-stats, shown in the parentheses, are based on the standard errors clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags.

To get the difference in the mean return between group 5 and group 1, I conduct similar analysis with the following modification: Instead of regressing on five dummy variables, I regress each set of future monthly returns (for example, all the monthly returns between 7 to 12 months after the formation) on a constant 1 and the dummies for group 2 to group 5, as well as contemporaneous Fama-French three factors:

$$r_{t,t+i} = \tilde{\alpha}_{s,1}+\tilde{\alpha}_{s,2}\mathbf{1}_2+\tilde{\alpha}_{s,3}\mathbf{1}_3+\tilde{\alpha}_{s,4}\mathbf{1}_4+\tilde{\alpha}_{s,5}\mathbf{1}_5+\beta_{M,s}\times MktRf_{t+i}+\beta_{S,s}\times SMB_{t+i}+\beta_{H,s}\times HML_{t+i}+\epsilon$$

The regression coefficients on the dummies are effectively the average return difference from group 1. For example, $\tilde{\alpha}_{s,5}$ measures the average return difference between group 5 and group 1 after adjusting contemporaneous Fama-French three factors. This number is also the one reported in the row '5_1' of the time-series sorting tables.

2.2.3 Link to Arbitrage Activity

As discussed before, the mechanism between momentum spread and momentum arbitrage capital is straight-forward: momentum arbitrageurs are buying winners and selling losers with the hope that they can profit from the slow dissemination of information and/or the underreaction of other investors. Meanwhile, their arbitraging behavior will impact the asset prices as well, which are reflected in the momentum spread. When the aggregate momentum arbitrage capital is excessively high, the prices would complete the convergence to the true value or even overshoot, which generates no profit or strong reversals in the future.

$$Mom.Spread_t = \beta_0+\beta_1\times HFAUM_{t-1}+\beta_2\times Mom(12-1)_{t-1}+\beta_3\times Comomentum_{t-1}+Controls_{t-1} \quad (2.1)$$

Table 2.3 connects momentum spread to several variables that proxy for the amount of capital allocated in momentum arbitrage. Specifically, I use several variables in quarter $t-1$ to forecast the momentum spread in quarter t. The forecasting variables I focus on here include the asset under management (AUM) of the hedge fund industry, the momentum returns in the most recent 12 months and the comomentum by Lou and Polk (2013). I would expect that more capital managed by hedge funds, or higher past momentum profit, or larger comovement among momentum stocks to proxy higher momentum arbitrage capital leads to a larger momentum spread.

(Insert Table 2.3)

In the sample period 2000-2018, where the aggregate hedge fund AUM data is available, I first find that when more asset is managed by hedge fund managers, who are the main investors in momentum, the future momentum spread is relatively high. The result from the momentum returns in the most recent 12 months is also consistent my above hypotheses: arbitrageurs will pour in more capital if they find the past performance of momentum is high. These results are robust when controlling the market state variables like the past 36-month market returns and the market dispersion. In the last column, I move to the sample period 1965-2010 studied in Lou and Polk (2013) and find that the momentum spread is also relatively high after high comomentum periods.

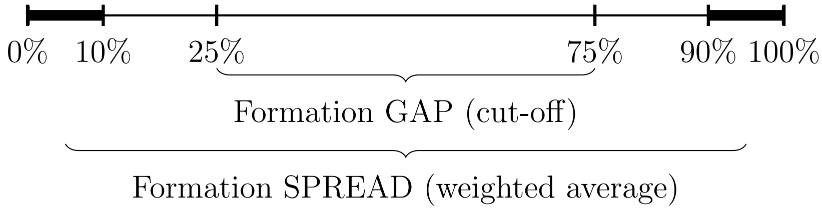
Higher Moments: Standard Deviation and Skewness

(Insert Table 2.4)

Table 2.4 checks the predictability of the momentum spread for higher moments of the momentum strategy. If the above excessive arbitrage argument is valid, I would expect periods with intensive momentum arbitrage activity would be followed by a higher standard deviation in momentum strategy and large crash risk going forward. Numbers in Table 2.4 confirms this hypothesis. The momentum strategy exhibits higher standard deviation after large momentum spread periods. As for the skewness, the crash-risk effect shows up after the first half-year and is particularly strong in the second year.

If we consider a mean-variance investor, though the average return is not predictable in the short-run, she might still benefit from timing the momentum strategy using the momentum spread due to the predictability in the standard deviation, the second moment.

Why the Momentum Spread is different from the Momentum Gap?



What is the economic motivation behind using the momentum spread instead of the formation gap? First, instead of using the cut-off in the distribution of past returns among stocks, I use the weighted average of past returns for the extreme winners and losers. Momentum is not a linear strategy and it is more profitable to trade the extreme winner and losers. This empirical fact will attract more arbitrage activity in these extreme stocks, so the spread from averaging the past performances of these stocks should be a better measure for arbitrage activity. Second, as argued by Novy-Marx (2012), momentum profit is mainly driven by the performance in the months from t-12 to t-6, so I use the performance from t-6 to t-1 to predicting the reversal - the non-profiting part of momentum.

Next, I will decompose momentum spread and focus on the momentum spread of the old momentum stocks because these stocks have been identified as momentum stocks for a longer length of time, and presumably, would be more heavily traded by arbitrageurs.

2.3 Decomposing Momentum Spreads

I now turn to the second main point of my paper: Can we refine the predictability from momentum spread by decomposing it? Recently, people have started to look into the cross-section of momentum stocks and find that by taking into account more information we can identify the true future winners and losers and hence improve the momentum performance. Chen et al. (2009) use a two-way sorting based on long-term (two-year) and recent performance (one-year) can accommodate both momentum and long-term reversal effects. Daniel et al. (2017) use institutional ownership and short interest data to filter out overpriced winners.

The method that I propose to distinguish momentum stocks is based on how many months a single stock has been identified as a momentum stock, which I term as momentum age. Since it takes time for the arbitrageurs to observe the momentum characteristics and to start to trade them, the stocks with more months identified as momentum stocks, i.e. old momentum stocks, are more likely to be traded by the arbitragers. So I focus on the momentum spread for these old momentum stocks and find that this old momentum spread has the strongest predictability for future momentum reversals across spread-based measures. I also provide evidence that these old momentum stocks are indeed more connected with arbitrage activity and less with fundamental information.

2.3.1 Momentum Age: Old and Young Momentum Stocks and Spreads

Momentum Age At the end of every month, after selecting the momentum stocks by their past performance, I classify them into subgroups based on their momentum age. I define momentum age as the number of months that a stock has been consecutively identified as a momentum stock in the past few months, both for winners and losers separately. For example, at the end of September, Microsoft shows up in the momentum portfolio as a winner (in the top decile), and it is not a winner in the last August, I assign age one to it. A month later, at the end of October, if Microsoft again stays as a winner (again in the top decile), then its momentum age will become two.

Table 2.5 reports the composition of momentum stocks in terms of their age, averaged across time. Stocks with momentum age larger than seven months consist of less than 20 percent of the momentum portfolio, which highlights the high turnover of momentum strategy. We can also see that the turnover are slightly higher for losers. In the second half of the whole sample, which is after the public finding of the momentum phenomenon and also witnesses the growth of sophisticated arbitrageurs like hedge funds, I find that the proportion of extreme old momentum stocks (with age 8 or plus) is smaller, which indicates that the turnover is higher possibly due to larger arbitrage activity. Apart from that, I do not find much difference in the pattern of the distribution.

(Insert Table 2.5)

In non-tabulated results, I do not find that there is any significant difference between old and young momentum stocks in terms of their market cap, market beta, book-to-market ratio or liquidity level.

Old and Young Momentum Stocks and Spreads After the formation of momentum portfolio at the end of each month, within the batch of momentum stocks, I further separate them into two groups based on the momentum age of a stock: young (1-3) and old (4 and above)[5] Both winners and losers are defined as the stocks in the top and bottom deciles based on their past performance of recent 12 months, but skipping the recent month.

I define the old momentum spread as the difference between the weighted average of the past cumulative returns for old winners and old losers, and stocks being '*old*' means that their momentum ages are above three months. Similarly, I can define the young momentum spread. Together with the momentum spread I defined earlier, these three momentum spreads are effectively the same, but calculated with all momentum stocks regardless of being old and young:

$$
\begin{aligned}
Mom.Spread_{t-1} &= \overline{Past.Ret_{t-6,t-1,Winners}} - \overline{Past.Ret_{t-6,t-1,Losers}}.\\
Old.Mom.Spread_{t-1} &= \overline{Past.Ret_{t-6,t-1,Winners,old}} - \overline{Past.Ret_{t-6,t-1,Losers,old}}.\\
Young.Mom.Spread_{t-1} &= \overline{Past.Ret_{t-6,t-1,Winners,young}} - \overline{Past.Ret_{t-6,t-1,Losers,young}}.
\end{aligned}
$$

Note that the momentum spread for all stocks can be (roughly) viewed as the weighted average of the old momentum spread and young momentum spread. Because the old and young stocks have different weights within the winner leg and the loser leg, this is an approximation relationship.

[5] I also tried other classifications, like young (1-4) and old (5 and above), and all my results hold qualitatively.

Again, there are mainly two reasons I am defining them in this way: first, the main reason I use the momentum spread, instead of momentum gap (the difference of the 75th and 25th cutoff), is that as mentioned before, momentum is not a linear strategy - trading the extreme decile momentum stocks shows much higher profit than trading quintile momentum stocks. If these spread-based measures are truly capturing momentum arbitrage activity as Huang (2015) argues in his paper, I would expect the momentum arbitrage pushes the extreme measures wider. In other words, the measures using extreme stocks (in the top or bottom decile) should contain more information about arbitrage activity, proxy the momentum arbitrage activity better, and hence predict the momentum reversal more strongly than the measures using the non-extreme stocks (at the 75th or 25th cut-off). However, as shown in his paper, if he is using another measuring, the difference between the 90th and 10th percentiles of the distribution of cumulative stocks returns, his results are less robust.

Second, I use the past cumulative returns in the last 6 months, $t-6$ to $t-1$, instead of $t-12$ to $t-2$, to calculate the momentum spread. The main reason I do this is that, since momentum ranking is based on the $t-12$ to $t-2$ returns and both of old and young momentum stocks need to meet the momentum criteria (cut-off) first, the correlation between old formation spreads and young formation spreads will be too high, conditioning on being calculated from the same ranking period returns. By using a shorter time horizon ($t-6$ to $t-1$) from the ranking horizon, the correlation between old and young momentum spreads decreases, which enables me to study them separately. This modification is also consistent with the result in Novy-Marx (2012): since he finds that momentum return with the one-year ranking period, is mainly positively driven by the first half of the year, one would expect the second half of the year is negatively driving the momentum return, in a relative sense. This naturally connects to the destabilizing excessive arbitrage.

High Correlation between Spread-based Predictors By construction, these spread-based measures will have high correlations among them because they share overlapping periods.

(Insert Table 2.6)

Table 2.6 reports the summary statistics and correlations among different predictors for momentum strategy returns. Momentum spread has a mean around 55% which lies between the old momentum spread and young momentum spread. Young momentum spread is slightly larger, which makes sense since young momentum stocks have to perform well in the recent half-year to show up as momentum stocks.

As for the correlations, the spread-based predictors are highly correlated by construction, which motives me to conduct a specific approach in the next step to find the strongest predictor. However, they are not particularly correlated with non-spread-based predictors like comomentum, past momentum volatility, past market return, volatility, and dispersion.

2.3.2 Strong Predictability of Old Momentum Spread

Long-term Predictability of Old Momentum Spread

(Insert Table 2.7)

Table 2.7 shows the time-series sorting results of predicting future momentum strategy returns with old momentum spread. Similar to the results in Table 2.2, old momentum spread preserves strong predictability for momentum reversals starting from the second half-year after the formation period and until two years after. This is not surprising considering the fact that old momentum spread is part of the momentum spread. A more interesting exercise is to disentangle and compare the predictive effect of two components of the momentum spread, i.e. old and young momentum spreads.

Comparing the Predictive Power between Old Momentum Spread and Other Spread-based Predictors Due to the high correlation between spread-based predictors as shown in Table 2.6, a direct regression analysis by throwing in all spread-based predictors will suffer multicollinearity problem and the results will be difficult and unreliable to interpret. To avoid this problem, I run a within-group-controlled regression as the following: for all the momentum strategy monthly returns in month $t+i$ within future set s,

$$r_{t,t+i} = \underbrace{\Sigma_{k=1}^{10}\alpha_k \mathbf{1}_k}_{\substack{\text{Dummy controls of} \\ \text{the first predictor}}} + \underbrace{\beta x_{t-1}}_{\text{Second predictor}} + FF.Factors_{t+i} + \epsilon_{t,t+i} \tag{2.2}$$

To compare the predictive power of two predictors, I first sort all formation months into deciles based on the first predictor and then generate dummies $\mathbf{1}_k$ for each formation months based on which group they are assigned to. Then I regress the future momentum returns on the second predictor x_{t-1} in their formation period and the dummies based on the first predictor, as well on contemporaneous Fama-French three factors, as Equation 2.2 shows. $r_{t,t+i}$ represent the momentum strategy return in month $t+i$ from the formation month t. I consider the months i after the momentum formation in groups: month 1-6, month 7-12, month 12-24 and month 25-26. This effectively constrains the regression coefficients in each group to be the same.

The regression coefficients from Equation 2.2 are effectively the same with Equation 2.3, which is easier to interpret. Equation 2.3 can be understood as that I first sort all months into ten deciles k based on their first predictor, and then within each group, I demean both the momentum returns $r_{t,t+i,k}$ on the left-hand side and the second predictor $x_{t-1,k}$ on the right-hand side. Put differently, this regression analyzes *the predictive power of the second predictor on top of the predictability of the first predictor.*

$$\underbrace{(r_{t,t+i,k} - \overline{r}_{t+i,k})}_{\substack{\text{Demean returns} \\ \text{within groups soted on} \\ \text{the first predictor}}} = \alpha + \underbrace{\beta(x_{t-1,k} - \overline{x}_k)}_{\substack{\text{Demean the second predictor} \\ \text{within groups soted on} \\ \text{the first predictor}}} + FF.Factors_{t+i} + e_{t,t+i}. \tag{2.3}$$

(Insert Table 2.8)

Results of these controlled-dummy regressions are shown in Table 2.8. I only report the point estimates of the coefficients for β for each specification and each future month set s: the first predictor as dummy controls, the second predictor as the regressor and which months each regression are focusing on. It shows that after controlling the old momentum spread, none of the young momentum spread, the formation gap, the formation spread or the comomentum shows strong and consistent predictive power for long-term reversals of momentum strategy. I also find that the comomentum still preserves strong predictability in the first half-year. Please note that the magnitude of the coefficients are comparable as I standardize all second predictors to have a standard deviation equal to 1.

However, if I switch the order and control any of these variables first, and then regress the long-term momentum strategy returns on the old momentum spreads, the predictability for reversals for periods between 6 and 24 months after momentum formation is significant and consistent. This result supports that old momentum spread shows stronger predictive power for momentum reversals than other spread-based predictors.

Predictive Regression: Comparing with more Predictors After ruling out the spread-based competitors, I can include other momentum predictors which have low correlation with old momentum spread.

Here I focus on five predictors well documented in the literature: past 36-month market return studied in Cooper et al. (2004), 3-month moving average of the cross-sectional return dispersion in 100 size-value portfolio studied in Stivers and Sun (2010), past 6-month realized market volatility studied in Wang and Xu (2015), past 6-month realized momentum volatility studied in Barroso and Santa-Clara (2015), and lastly, comomentum studied in Lou and Polk (2013).

(Insert Table 2.9)

Table 2.9 show the regression results for three group of months. In the short-term, i.e. in the first six months after the momentum formation, the predictability of old momentum spread is negligible and other competing predictors have a stronger forecasting power. However, if we focus on the long-term predictability, for both month 7-12 and month 13-24, the predictive power of old momentum spread remains strong and drives out the predictability of other predictors.

2.3.3 Old and Young Momentum Portfolios

Next, I provide evidence that supports the idea that old momentum spread is more connected to the momentum arbitrage activity.

Stronger Unconditional Reversals of Old Momentum Portfolios The first idea that I want to establish and elaborate here is that if a stock has been in the momentum portfolios for a longer horizon, it will be pushed further away from its fundamentals by the momentum traders. If this holds true, I would expect an unconditional stronger reversal for these old momentum stocks.

To see whether the above hypothesis holds in the data, I slit the momentum strategy into two parts: young momentum strategy which only consists of young momentum stocks and old momentum strategy which only consists of old momentum stocks.

$$\begin{aligned} MOM_{young} &= Ret\_Winners_{young} - Ret\_Losers_{young}, \\ MOM_{old} &= Ret\_Winners_{old} - Ret\_Losers_{old}, \end{aligned} \tag{2.4}$$

where $Ret\_Winners_{young} = \sum \omega_i R_i$ and $i \in$ Young Winners and the weight ω_i is the lagged market cap; and similar for Ret_Losers_{young}, $Ret\_Winners_{old}$, and Ret_Losers_{old}.

(Insert Figure 2.3)

In Figure 2.3, I replicate the classical figure 3 in Jegadeesh and Titman (2001) with the same sample period and plot the cumulative returns of the original momentum portfolio, as well as the young and old momentum portfolio separately. The figure shows that the old momentum portfolio will revert sooner and stronger, while the young momentum portfolio bearly reverse in the first 30 months after the portfolio formation. The pattern remains similar if I extend to the most recent periods. This pattern is consistent with the idea that old momentum stocks are pushed too far away from their fundamentals and their prices will revert to their fair value in the long run.

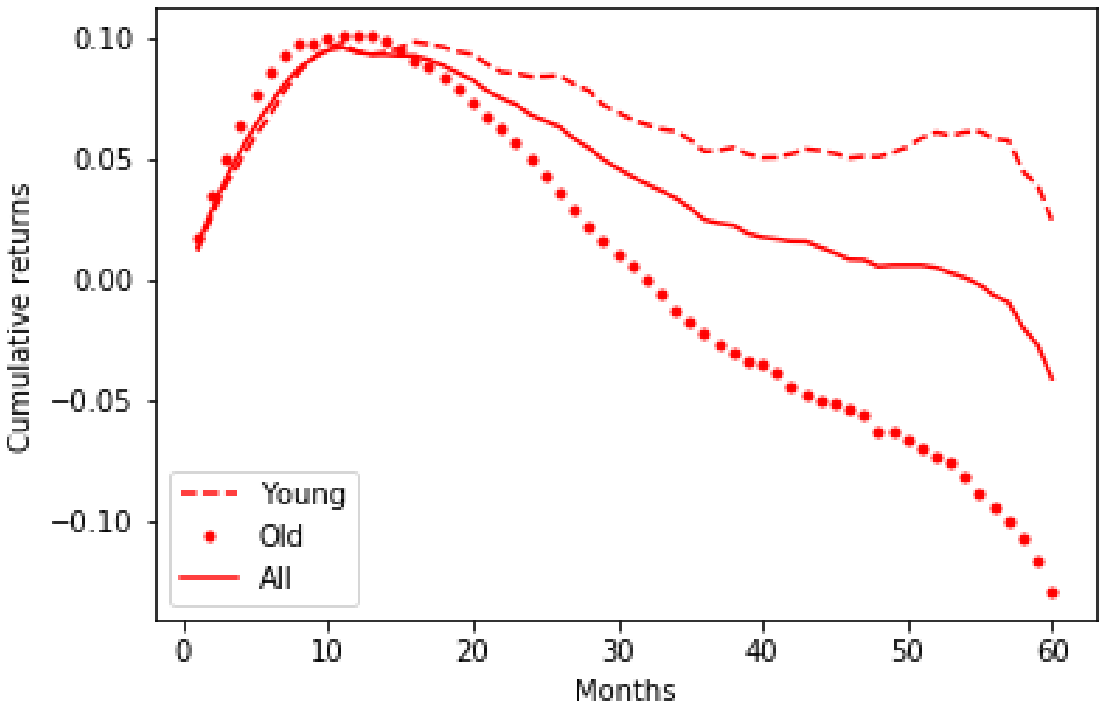


Figure 2.3: Momentum Strategy Performance: Old and Young

Cumulative momentum returns for old and young momentum portfolios (no balancing after formation) up to 60 months. The sample period in this figure is 1965-1997, which is the same with Jegadeesh and Titman (2001).

Another pattern in Figure 2.3 worth mentioning is that the old momentum stocks have better performance in the first ten months after the formation period. This fact is consistent with the idea of the coordination problem among momentum arbitrageurs: it is difficult for the market participants to infer how much capital in aggregate is allocated in the momentum arbitrage, so they may overchase the old and strong winners myopically in the short-term and generate the stronger reversals in the long-term. From another perspective, this also supports the empirical fact that the predictability from the spread will not emerge in the short-term but the long-term. Huang et al. (2016) also find a similar pattern in the beta arbitrage and argue that arbitrage activity instead generates booms and busts in the strategy. More specifically, they find that when activity is high, prices overshoot as short-run abnormal returns are much larger and then revert in the long run.

Different Factor Loadings of Old and Young Momentum Portfolios I also provide evidence that the old momentum strategy and young momentum strategy carries different loadings on certain factors. I run the following regressions and report the loadings β_k on different factors for old momentum strategy and young momentum strategy separately.

$$r_t = a + \Sigma_k \beta_k F_{k,t} + \epsilon_t \tag{2.5}$$

I consider the following factors. Post–earnings-announcement drift (PEAD) factor, which is constructed by longing the stocks in the top decile and shorting the stocks in the bottom decile based on the three-day cumulative abnormal returns around the most recent earnings announcement days. This PEAD factor proxies the profit from trading on the fundamental news. VIX, the CBOE Volatility Index, measures the market expectation of future volatility. HFI is the Equity Market Neutral Index (HFRIEMNI) from Hedge Fund Research website. I use the Equity Market Neutral Index (HFRIEMNI), which studies the quantitative equity funds and dates back to the beginning of 1990. Liquidity Level is the market liquidity measure by Pastor and Stambaugh (2003). MktRf, SMB, and HML are the standard three Fama-French factors.

(Insert Table 2.10)

Table 2.10 shows that young momentum strategy has a stronger loading on the PEAD factor, which supports the idea that young momentum is more about the fundamental information. It also shows that the old momentum strategy has stronger loadings on the VIX index and the long-short equity hedge fund returns. This observation is consistent with that old momentum stocks is more traded and affected by the arbitrageurs. The sign of the coefficient on VIX is also consistent with the above interpretation. When some unexpected negative shock hits the market, or when VIX goes up and margin (requirement) becomes tight, or when arbitrageurs unwind part of their bettings on (old) momentum positions and generate negative price impacts. One final observation is that the negative correlation between momentum and value is mostly driven by old momentum stocks.

These results consolidate the idea that, in a relative sense, the information in old momentum stocks is more associated with arbitrage activity and the information in young momentum stocks is more associated with the fundamental news. With this evidence in mind: if momentum stocks with different ages are behaving differently after the formation, it is natural to expect the momentum stocks (and their spreads) with different ages are also carrying different information at the formation time. This drives me to treat the old momentum spread and young momentum spread differently.

Predicting Old and Young Momentum Portfolios Finally, I show that the predictability to the old momentum portfolio is stronger in magnitude. Since, first, the old momentum spreads are the main driver of the predictability, second, the old momentum stocks are suffering unconditional long-term reversal the most (from Figure 2.3), and third, there is supportive evidence that old momentum stocks are more traded by arbitrageurs and pushed further away from their fundamentals, combining these three facts should yield stronger predictability of old momentum spread on old momentum stocks than on young momentum stocks.

(Insert Table 2.11)

Table 2.11 checks the predictability of old momentum spread for old and young momentum portfolios separately. Clearly, the long-term predictability (month 6-24) of old momentum spread exists for both types of momentum portfolios. More importantly, it also shows that this predictive effect is much stronger for old momentum portfolios in terms of both the economic magnitude and statistical significance.

2.3.4 Implication from Hong and Stein (1999)'s Model

Now I borrow the model in Hong and Stein (1999) to link the old momentum spread to excessive arbitrage activity, which may push the winners too high and losers too low and lead to a mechanical long-term reversal back to their fundamentals.

In the classical paper by Hong and Stein (1999), *A unified theory of underreaction, momentum trading, and overreaction in asset markets*, they build an equilibrium model with bounded rationality. In their model, there are two types of investors, *News-watchers* and *Momentum traders*, both risk-averse with CARA utility. Their rationality is bounded in a way: for news-watchers, they only process fundamental information; and for momentum traders, they only trade on the past price change.

In a simplified version of their model, the price follows a process:

$$P_t = P_{t-1} + \Delta I_t + X_t - X_{t-12}, \tag{2.6}$$

where ΔI_t is the change in the fundamental information and X_t is the price impact from momentum trading activity, which follows

$$X_t = \phi \Delta P_{t-1}, \tag{2.7}$$

ϕ the equilibrium demand elasticity, and 12 is the holding period: after 12 periods, momentum traders will close their position, which brings in an opposite price impact with a lag of 12, $-X_{t-12}$.

In their original model, they assume that the size of momentum traders is constant over time, which is not true in the real world. In practice, the arbitrage capital has a huge time-variation in practice, and meanwhile, the exact amount to capital allocated in arbitrage activity is unknown to arbitrageurs. Stein (2009) argue that without knowing the exact amount of arbitrage activity if all arbitrageurs are just optimizing their own leverage without considering the externality, financial markets will suffer possible destabilization. He denotes this as a coordination problem, which comes from the inability of traders to condition their behaviour on current market-wide arbitrage capacity.

If the time-variation of momentum arbitrage capital γ_t was introduced Hong and Stein (1999)'s model, which comes from the time-variation in either the original asset under management, the leverage on original arbitrage capital, or the momentum traders' risk aversion. Most importantly, this is **not** observed by other momentum traders. The new momentum trade follows:

$$X_t = \gamma_t \phi \Delta P_{t-1}, \tag{2.8}$$

and in their original model, $\gamma_t = 1$.

With this modification, there is an **amplification effect** in the **momentum**

positive feedback:

$$\frac{dX_t}{d\gamma_{t-1}}|_{\gamma_t} = \gamma_t\phi\frac{d\Delta P_{t-1}}{d\gamma_{t-1}} = \gamma_t\phi(\phi\Delta P_{t-2}), \quad \text{(positive)}. \tag{2.9}$$

Momentum positive feedback captures that more momentum arbitrage capital in last period γ_{t-1} will generate a larger price impact ΔP_{t-1}, and this will attract more momentum trading in this period with the same direction. In Hong and Stein (1999), they explore the change in risk tolerance, which has the same effect here.

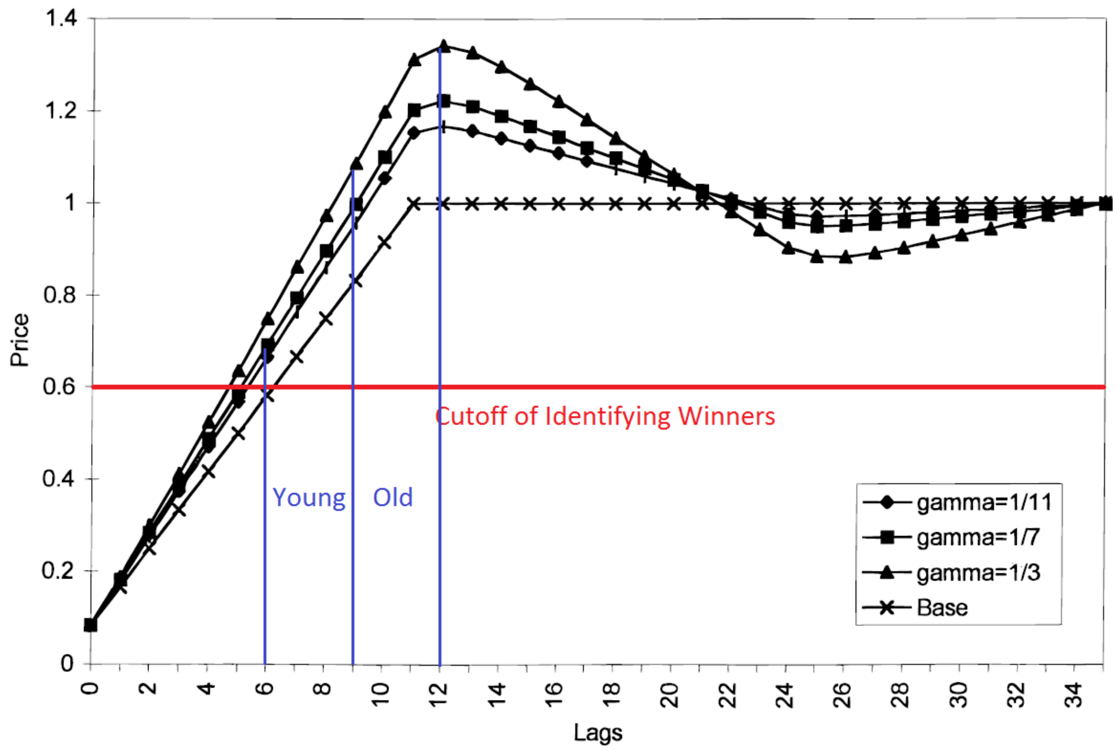


Figure 2.4: Cumulative impulse response in Hong and Stein (1999)'s model

Hong and Stein (1999)'s model with different risk tolerance from momentum traders, or equivalently different amount arbitrage capital in trading momentum. The higher the gamma, the higher the risk tolerance, the more intense the momentum arbitrage.

(Insert Figure 2.4)

Figure 2.4 shows the impulse response function of prices if there is a fundamental shock. It takes the benchmark parameterization in Hong and Stein (1999), with the $z = 12$ measuring the (linear) rate of information flow and higher values of z implying slower information diffusion, and $j = 12$ the momentum traders' horizon, and the volatility of the new shocks 0.5.

The momentum age can be recovered in this figure. In this case, let me suppose the cross-sectional winner cutoff[6] is 0.6, an arbitrary number for illustrative purpose. The price before the fundamental information shock (t=0) will be zero. After a positive fundamental information shock to the stock, its price will trend up but it will not be identified as a winner immediately because its price is below the cross-sectional winner cutoff, 0.6. After 6 lags of the initial positive shock, its price goes above the winner cutoff and this stock will appear as a young winner stock for the first time. After eight lags, its price is still trending up and it has been identified as a winner stock for more than three periods, in other words, an old winner stock. It is clear in the figure that the higher the momentum age, the larger the formation spread (the price between the current price and the price 12 periods ago). In the figures, the formation spread is the vertical distance between the price and the horizontal axis if we focus on the first 12 periods after the initial fundamental shock. This is because the price is zero before the shock.

On the other hand, in Figure 2.4, the momentum traders' risk tolerance takes three values $\gamma = 1/3$, $\gamma = 1/7$ and $\gamma = 1/11$, which is one potential source of the time-varying momentum arbitrage capital. The figure clearly shows that with more momentum arbitrage capital (in this case, the momentum traders are more risk-tolerant), the momentum spread is larger and the following long-term reversal is stronger. Moreover, due to the positive feedback in trading momentum, the absolute difference in the momentum spread between low and high scenarios will be larger if we focus on the old momentum stocks. This observation provides a theoretical foundation of using the old momentum spread to predict future reversals.

I formulate three hypotheses below and show that my empirical findings consistent with them:

- **Hypothesis 1**: With the increase in the momentum age, the momentum formation spread (the difference of the past cumulative returns between the winners and losers) will be larger.

- **Hypothesis 2**: Conditioning on same age (a given point on the horizontal axis), more momentum arbitrage capital will drive up the momentum spreads

[6]Momentum arbitrageurs need time to observe the momentum characteristics in the cross-section and then start to trade these stocks.

and result in stronger reversals in the long-term.

- **Hypothesis 3**: Because of the positive feedback effect, the price impact coming from momentum trading will increase with the momentum age, relative to the fundamental part. Old momentum spreads will carry relatively more information about the momentum arbitrage activity than young momentum spreads. Old momentum spreads will predict the long-term reversal of momentum more strongly.

(Insert Table 2.12, Table 2.2 and Table 2.8)

In Table 2.12, I find evidence consistent with the **Hypothesis 1**. With higher momentum arbitrage activity, proxied by larger old momentum spreads, the past return spreads (from t-12 to t-2) between winners and losers are larger. This effect is larger for stocks with higher momentum ages, the old momentum stocks.

Previous results shown in Table 2.2 support that the larger the momentum spread is, the stronger the reversal will be, consistent with the **Hypothesis 2**. In Table 2.8, I find the old momentum spread has higher predictability than the young one, which is consistent with **Hypothesis 3**.

2.4 Robustness

Momentum Arbitrage in the Formation Period One concern is that arbitrageurs will begin trading momentum stocks only after the momentum characteristics are observed. That concern is mitigated by the fact that I measure the momentum characteristic based on a relatively long ranking period up to one year while other studies have shown that using ranking periods as short as 3 or 6 months (see Jegadeesh and Titman (1993)) also produces mid-term momentum effect. Presumably, momentum traders are trading based on these short-term momentum characteristics and thus would have been pushing winners and losers apart from each other, which would be picked up my measure, the (old) momentum spread. Nevertheless, if I instead measure the (old) momentum spread in the post-ranking period, most of my results continue to hold since, first, the reversal that I am trying to explain is not happening in the short-run (within half of a year), and second, the (old) momentum spread measure is persistent.

Subsample Analysis: Stronger Predictability in the Second Half

(Insert Table 2.13)

Table 2.13 reports the time-series sorting results in two subsample periods. The predictability is stronger in the second half of the sample, from 1990 to 2018, in terms of both the economic magnitude and the statistical significance. This supports the argument of the destabilizing mechanism of excessive arbitrage since the momentum arbitrage is larger in the second half of the sample with the momentum effect being publicly studied around the 90s and the large growth in asset management industry since then.

Out-of-Sample Predictability Though the main point of this paper is not to propose a way to exploit the predictability in momentum strategy, it is still worth to explore whether the in-sample relation between the old momentum spread and future reversals is stable when extending out of sample (OOS). Welch and Goyal (2007) have highlighted the problem of out-of-sample predictability and argued that in-sample correlations conceal a systematic failure of these variables out of sample. They find that none of the well-known market predictors are able to beat a simple forecast based on the historical average stock return. Campbell and Thompson (2007) show that forecasting variables with significant forecasting power in-sample generally have a better out-of-sample performance than a forecast based on the historical average return, once imposing restrictions on the signs of coefficients and return forecasts. I follow Campbell and Thompson (2007) to construct the out-of-sample R-squared:

$$R^2_{OOS} = 1 - \frac{\Sigma(r_t - \hat{r}_t)}{\Sigma(r_t - \overline{r}_t)} \tag{2.10}$$

where r_t is the future momentum strategy return, $\hat{r}_t$ is the fitted value from a predictive regression estimated in the rolling look-back (training) periods, and $\overline{r}_t$ is the historical average return estimated in the look-back periods. For example, if the look-back years equal to 10, I use the data from 2001 to 2010 to train the model and then use it to predict the momentum returns starting in 2011.

(Insert Table 2.14)

Table 2.14 reports the out-of-sample returns with three predictors, Old Momentum Spread, Formation Gap and Comomentum. For each predictor, I further split their predictability for the future three sets of months: month 1-6, month 7-12 and month 13-24. I also checked the stability in the choice of the look-back (training) period by reporting the different look-back periods. As the table shows, in the short-term, the past average outperforms all predictors with all out-of-sample R-squared being negative, including the comomentum which has the highest in-sample predictability for the first six months. When we look at the month 7-12, the old momentum spread shows strong out-of-sample predictability for several different look-back periods.

2.5 Cross Predictability and Factor Timing Strategy

Another observation from Table 2.2 is that, after controlling Fama-French three factors, the predictability of long-term reversal decreases in terms of magnitude (though still robust). Put differently, the momentum spread can predict other factors (size and value factors in this case) as well.

There are several strategies or factors consistent with the idea of low-risk defensive investing. Asness et al. (2014) state that '*low-risk investing that focus on various measures: market beta (Black (1972); Frazzini and Pedersen (2014)), total volatility (e.g., Baker et al. (2011)), residual volatility (e.g., Falkenstein 1994; Ang et al. (2006); Blitz and Van Vliet (2007)), the minimum-variance portfolio, and other related measures (for connections between these measures, see Clarke et al. (2013)).*' These types of strategies are all longing the relatively low-risk and high-value stocks with high book-to-market ratio, low idiosyncratic volatility and low market beta, and shorting the counterparts.

On the other hand, momentum is more '*aggressive*' by chasing the winners and losers, and for most of the time, coming with relatively higher risk. There is also anecdotal evidence suggesting that hedge funds, especially high-frequency trading funds, prefer to hold stocks with high iVol and large beta, for intraday trading on both directions.

If the momentum spread, proxying the momentum arbitrage capital, forecasts the bust of momentum bubble, I would expect that arbitrageurs or the investors in momentum will reallocate capital to other investment strategies if they find that momentum is not profitable anymore going forward.

I simply consider an equal-weighted low-risk combo factor for value, low-iVol and low-beta strategy, as in Equation 2.11. I use a moving average (MA) strategy to time the factor returns: comparing the '*old momentum spread*$\perp$' with its MA in the last 12 months. '*old momentum spread*$\perp$' is the part orthogonal to the cumulative market returns and the market volatility in the last 36 months, which have been found being able to predict momentum returns by Cooper et al. (2004) and Wang and Xu (2015) respectively.

$$Combo_t = \frac{1}{3}(HML_t + VOL_t + BAB_t). \tag{2.11}$$

(Insert Table 2.15 and Table 2.16)

Table 2.15 and Table 2.16 show that with high old momentum spread (a proxy for high momentum arbitrage), the combo strategy performs poorly in the short-term, but relatively well in the long-term. It is logically the same story in the momentum predictability, but with opposite direction: in the short-run, arbitrageurs are still riding on momentum, without knowing the exact time the bubble will burst. However, in the long-term, momentum performs badly and the combo strategy performs well. It is plausible that when arbitrageurs find momentum is not profiting anymore, they decide to move their capital to pursue other strategies. This 'substitute' effect (without identification) is stronger in the recent 20 years, possibly with the increasing market share of professional institutional investors, allocating capital across strategies from time to time.

2.6 Conclusion

Different from prior studies, I find that the momentum spread negatively predicts momentum returns in the long-term, but not in the following month. This predictability is strong and robust after controlling several other known momentum predictors as well as the contemporaneous Fama-French three factors. I provide supportive evidence and argue that excessive momentum arbitrage activity drives up the momentum spread and leads to lower future momentum returns or even reversals; more importantly, these reversals will not show up instantly, but rather in the long-term (mostly after the first six months).

I decompose the momentum spread into old and young ones based on the momentum age of stocks and find that the old momentum spread drives the predictability. Both before and after adjusting Fama-French three factors, the old momentum spread has the highest predictive power for long-term reversals of momentum strategy among the momentum spread-based family and it also drives out the predictability of past market performance, volatility, cross-sectional dispersion, and momentum volatility. Given the evidence that suggests these old momentum stocks are more likely to be traded by arbitrageurs and the predictability to long-term reversals is stronger in the recent thirty years, these facts help to pin down the destabilizing mechanism of excessive arbitrage. I also connect the old momentum spread to the excessive momentum arbitrage activity based on Hong and Stein (1999)'s canonical model and test several implications that fit my hypotheses.

Finally, I find cross-predictability using the old momentum spread and argue that this could be due to the time-varying allocation of capital by the arbitrageurs. Though it is not the main point of this paper, further analysis is highly recommended in this direction to solidify the argument.

In a nutshell, these empirical patterns highlight the importance of arbitrageurs in setting asset prices and excessive arbitrage activity might destabilize certain asset prices due to the coordination problem in crowding.

Table 2.1: Forecasting one-month Momentum Returns with Momentum spread, Old momentum spread, Formation gap, Formation Spread, and Comomentum

This table reports the Fama-French three-factor adjusted returns of the momentum portfolio in the first month after formation. All months are sorted into five groups based on the realized characteristics (Momentum Spread, Old Momentum Spread, Formation Gap, Formation Spread, and Comomentum, respectively) in the last month. Momentum Spread, Old Momentum Spread are the difference of the past 6-month weighted-average returns between all or old winner and loser stocks. Formation Gap is the difference between 75th and 25th percentiles for momentum formation-period return (t-12 to t-2). Formation Spread is the difference between the momentum formation-period weighted-average returns (t-12 to t-2) of winners and losers. Comomentum is the average pairwise correlation among loser stocks in the last 12 months. The sample period is from 1960 to 2018. Due to data availability, the sample period for comomentum analysis is 1965-2010. 5_1 means the difference between the average returns in the month group 5 with largest characteristics and the month group 1 with the smallest. Due to data availability, the sample periods for comomentum analysis is 1965-2010. T-stats, shown in the parentheses, are based on the standard errors adjusted with Newey and West (1987) correlation for 6 lags.

Panel A: Full Sample

	Momentum Spread			Old Momentum Spread			Formation Gap			Formation Spread			Comomentum		
Group Ranking	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat
1	30.9%	1.6%	(3.97)	25.3%	1.2%	(3.14)	29.2%	1.9%	(4.82)	83.9%	1.2%	(2.32)	6.8%	2.2%	(4.64)
2	43.8%	2.0%	(4.17)	41.0%	1.3%	(2.61)	34.4%	0.9%	(1.56)	105.4%	0.9%	(1.83)	9.1%	2.1%	(4.20)
3	51.4%	1.1%	(2.29)	50.7%	1.9%	(4.00)	38.3%	0.4%	(0.65)	124.1%	1.6%	(4.13)	10.7%	1.8%	(2.92)
4	61.1%	0.6%	(1.11)	61.1%	1.2%	(1.96)	43.8%	1.9%	(3.59)	147.7%	1.6%	(2.88)	13.0%	0.8%	(1.16)
5	89.1%	1.3%	(1.74)	87.8%	1.1%	(1.60)	60.1%	1.1%	(1.39)	235.8%	1.3%	(1.72)	19.2%	-0.3%	(-0.30)
5_1	58.2%	-0.2%	(-0.28)	62.5%	-0.1%	(-0.08)	30.9%	-0.8%	(-0.95)	151.9%	0.1%	(0.16)	12.4%	-2.5%	(-2.32)

Panel B: 1st Half Sample 1960-1989

Group Ranking	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat
1	28.8%	1.9%	(3.56)	25.9%	1.7%	(3.19)	27.4%	1.9%	(3.96)	76.4%	1.9%	(3.75)	6.8%	1.6%	(2.66)
2	39.6%	1.7%	(2.73)	38.0%	2.1%	(3.68)	31.9%	1.2%	(1.53)	92.9%	0.9%	(1.27)	9.1%	2.1%	(3.30)
3	46.7%	2.2%	(4.11)	45.8%	1.6%	(2.70)	36.6%	2.4%	(3.93)	109.5%	2.5%	(6.04)	10.6%	2.1%	(3.17)
4	54.6%	1.2%	(1.70)	55.9%	1.5%	(2.46)	42.2%	2.2%	(3.37)	130.3%	2.5%	(4.80)	12.4%	2.1%	(2.73)
5	76.9%	1.2%	(1.73)	78.1%	1.3%	(1.86)	55.8%	0.2%	(0.21)	183.4%	0.4%	(0.61)	16.7%	0.0%	(0.03)
5_1	48.1%	-0.7%	(-0.80)	52.1%	-0.4%	(-0.43)	28.4%	-1.8%	(-1.97)	106.9%	-1.5%	(-1.78)	9.9%	-1.6%	(-1.40)

Panel C: 2nd Half Sample 1990-2018

Group Ranking	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat	Ranking var.	Returns	t-stat
1	34.6%	0.5%	(0.88)	25.2%	0.1%	(0.14)	32.1%	1.1%	(2.21)	99.7%	-1.0%	(-0.94)	7.1%	2.5%	(2.97)
2	48.5%	1.1%	(1.57)	44.8%	2.1%	(3.02)	36.1%	0.0%	(0.03)	120.6%	0.8%	(1.04)	9.3%	2.2%	(2.51)
3	56.8%	1.0%	(1.27)	55.4%	-0.3%	(-0.31)	39.6%	-0.4%	(-0.42)	136.9%	1.2%	(1.78)	11.2%	1.8%	(1.64)
4	66.0%	-0.2%	(-0.15)	65.1%	1.9%	(2.24)	45.3%	1.9%	(2.23)	167.8%	2.5%	(3.04)	14.7%	1.5%	(1.27)
5	99.9%	2.4%	(2.12)	96.6%	1.1%	(0.93)	63.8%	2.2%	(1.74)	279.5%	1.5%	(1.16)	21.4%	-2.7%	(-1.74)
5_1	65.3%	2.0%	(1.50)	71.3%	1.0%	(0.78)	31.7%	1.1%	(0.82)	179.8%	2.5%	(1.40)	14.2%	-5.3%	(-2.88)

Table 2.2: Forecasting Momentum Returns with Momentum Spread and Other Predictors

This table reports the average raw and Fama-French 3-factor adjusted monthly returns of the momentum portfolio for the 1-6, 7-12, 13-24 and 25-36 months after their formation (without any rebalancing). All months are sorted into five groups based on their momentum spread realized at the formation period. 5-1 means the difference between the average returns in the group with the largest momentum spread and the group with the smallest. The last panel reports a similar analysis with other predictors, including Formation Gap, Formation Spread, and Comomentum. The sample period is from 1960 to 2018. T-stats, shown in the parentheses, are based on the standard errors clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags.

Panel A: Raw Returns								
TS Group Rank	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	0.96%	(3.99)	0.80%	(4.11)	0.16%	(0.97)	-0.06%	(-0.38)
2	1.22%	(4.66)	0.15%	(0.65)	0.06%	(0.33)	0.09%	(0.47)
3	0.81%	(3.06)	0.19%	(0.83)	-0.32%	(-1.79)	0.11%	(0.52)
4	0.95%	(3.24)	0.57%	(2.14)	-0.27%	(-1.15)	-0.33%	(-1.30)
5	0.63%	(1.15)	-0.58%	(-1.22)	-1.02%	(-3.45)	-0.01%	(-0.02)
5_1	-0.32%	(-0.55)	-1.39%	(-2.71)	-1.18%	(-3.55)	0.06%	(0.19)

Panel B: Fama-French 3-factor Adjusted Returns								
TS Group Rank	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	1.24%	(5.02)	0.88%	(4.32)	0.36%	(2.34)	0.10%	(0.55)
2	1.49%	(5.28)	0.51%	(2.06)	0.18%	(1.18)	0.26%	(1.48)
3	1.16%	(4.63)	0.50%	(2.34)	-0.23%	(-1.42)	0.23%	(1.17)
4	1.22%	(4.27)	0.63%	(2.62)	-0.15%	(-0.70)	-0.15%	(-0.60)
5	0.91%	(1.95)	-0.09%	(-0.23)	-0.55%	(-2.08)	0.23%	(0.85)
5_1	-0.33%	(-0.64)	-0.97%	(-2.28)	-0.91%	(-3.03)	0.14%	(0.44)

Panel C: Fama-French 3-factor Adjusted Returns with Other Predictors								
TS Group Rank	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
	Predictor: Formation Gap							
1	1.21%	(3.62)	0.62%	(1.92)	0.18%	(0.86)	0.27%	(1.41)
5	1.15%	(2.17)	0.24%	(0.59)	-0.53%	(-2.02)	-0.11%	(-0.43)
5_1	-0.07%	(-0.11)	-0.37%	(-0.73)	-0.71%	(-2.13)	-0.38%	(-1.22)
	Predictor: Formation Spread							
1	1.03%	(2.57)	0.78%	(2.67)	0.18%	(0.89)	0.16%	(0.72)
5	0.97%	(1.91)	0.18%	(0.46)	-0.30%	(-1.22)	-0.22%	(-0.87)
5_1	-0.06%	(-0.10)	-0.61%	(-1.27)	-0.48%	(-1.50)	-0.38%	(-1.15)
	Predictor: Comomentum							
1	1.55%	(5.66)	0.59%	(2.51)	0.29%	(1.52)	0.34%	(1.50)
5	0.04%	(0.07)	-0.02%	(-0.05)	-0.27%	(-0.85)	1.13%	(3.64)
5_1	-1.51%	(-2.34)	-0.61%	(-1.32)	-0.56%	(-1.56)	0.78%	(2.09)

Table 2.3: Forecasting Momentum Spread

$$Mom.Spread_t = \beta_0 + \beta_1 \times HFAUM_{t-1} + \beta_2 \times Mom(12-1)_{t-1} + \beta_3 \times Comomentum_{t-1} + Controls_{t-1} + e_t$$

This table reports the quarterly predictive regressions of momentum spread on variables related to momentum arbitrage capital. $HF.AUM_{t-1}$ is the log asset under management in hedge funds obtained from BarclayHedge in quarter t-1. $Mom(12-1)_{t-1}$ is the 1-year cumulative momentum strategy ending in quarter t-1. Due to data availability, the sample period for the first two regressions is from 2000 to 2018 and the sample period for the third regression is from 1965 to 2010. T-stats, shown in the parentheses, are based on the standard errors with Newey and West (1987) correlation for 4 lags.

Determinants of Momentum Spread			
	[1]	[2]	[3]
HFAUM t-1	0.77	0.78	
	(2.63)	(2.63)	
Mom(12-1) t-1	0.18	0.17	-0.04
	(2.85)	(2.67)	(-0.56)
MktRf(36-1) t-1		-0.05	0.37
		(-0.87)	(5.00)
CrossS.Disp(MA-3) t-1		0.00	0.02
		(0.78)	(2.84)
Comomentum t-1			0.99
			(2.14)
Trend	Yes	Yes	Yes
Sample	2000-2018	2000-2018	1965-2010

Table 2.4: Predicting Higher Moments of Momentum: Standard Deviation and Skewness

This table reports the standard deviation and skewness of the momentum monthly returns for the 1-6, 7-12, 13-24 and 25-36 months after their formation (without any rebalancing). All months are sorted into five groups based on their momentum spread realized at the formation period. The sample period is from 1960 to 2018.

	Panel A: Standard Deviation of Momentum Monthly Returns			
Group Rank	Month 1-6	Month 7-12	Month 12-24	Month 25-36
1	4.7%	3.8%	4.1%	4.4%
2	5.2%	4.7%	4.6%	4.9%
3	5.6%	4.9%	5.0%	5.8%
4	6.3%	5.7%	6.3%	6.8%
5	9.2%	7.6%	6.7%	6.4%

	Panel B: Skewness of Momentum Monthly Returns			
Group Rank	Month 1-6	Month 7-12	Month 12-24	Month 25-36
1	-0.02	0.01	-0.25	-0.08
2	-0.10	-0.13	0.03	0.26
3	-0.95	-0.14	-0.16	0.32
4	-0.35	0.20	-0.94	0.14
5	0.34	-0.28	-1.74	0.33

Table 2.5: Momentum Age Distribution

This table reports the percentage distribution of momentum stocks for different momentum ages. At the end of every month, after selecting the momentum stocks by their past performance (t-12 to t-2), I classify them into subgroups based on their momentum age, which is the number of months that a stock has been consecutively identified as a momentum stock in the past few months, both for winners and losers separately. The numbers below are the time-series average of the composition of momentum stocks.

Momentum Stocks' Composition based on Momentum Age								
	Full Samle: 1960-2018							
	Young (1 - 3)			Old (4 - 8+)				
Winners	56.6%			43.4%				
Losers	59.0%			41.0%				
Momentum Age	1	2	3	4	5	6	7	8+
Winners	28.2%	16.7%	11.8%	8.9%	7.1%	5.7%	4.7%	17.0%
Losers	30.0%	17.2%	11.9%	8.9%	6.9%	5.5%	4.4%	15.4%
	1960-1989							
Winners	28.4%	16.6%	11.5%	8.8%	6.9%	5.6%	4.6%	17.6%
Losers	30.4%	16.9%	11.6%	8.7%	6.8%	5.3%	4.3%	16.1%
	1990-2018							
Winners	27.9%	16.8%	12.0%	9.1%	7.2%	5.8%	4.7%	16.4%
Losers	29.4%	17.4%	12.1%	9.1%	7.0%	5.6%	4.5%	14.7%

Table 2.6: Time-Series Summary Statistics and Correlations among Different Predictors

This table reports the summary statistics and correlations among different predictors for momentum strategy returns. Momentum Spread, Old Momentum Spread and Young Momentum Spread are the difference of the past 6-month weighted-average returns between all, old and young winner and loser stocks. Formation Gap is the difference between 75th and 25th percentiles for momentum formation-period return (t-12 to t-2). Formation Spread is the difference between the momentum formation-period weighted-average returns (t-12 to t-2) of winners and losers. Comomentum is the average pairwise correlation among loser stocks in the last 12 months. MktRf(36-1) is the cumulative market excess returns in the last 36 months. MktVol(6-1) is the past 6-month realized market volatility and MomVol(6-1) is the past 6-month realized momentum strategy volatility, both normalized to monthly level. CrossS.Disp(MA-3) is the 3-month moving average of cross-sectional monthly return dispersion within 100 size-value portfolios. The sample period is from 1960 to 2018. Due to data availability, the sample period for comomentum analysis is 1965-2010.

	Time-series Summary Statistics									
	Mom. Spread	Old Mom. Spread	Young Mom. Spread	Formation Gap	Formation Spread	Comomentum	MktRf(-36)	MktVol(-6)	MomVol(-6)	CrossS.Disp
N	708	708	708	708	708	559	708	708	708	708
Mean	55.3%	53.2%	57.3%	41.2%	139.4%	11.8%	20.6%	4.1%	5.0%	4.4%
Standard Deviation	27.3%	28.1%	29.6%	12.0%	67.9%	4.6%	28.2%	2.0%	3.6%	3.7%
	Correlation Table									
Mom. Spread	1	0.94	0.97	0.49	0.72	0.11	0.20	0.07	0.06	0.11
Old Mom. Spread	0.94	1	0.85	0.45	0.66	0.08	0.22	0.00	-0.04	0.08
Young Mom. Spread	0.97	0.85	1	0.49	0.72	0.13	0.19	0.10	0.12	0.13
Formation Gap	0.49	0.45	0.49	1	0.82	0.24	0.05	0.14	0.21	0.35
Formation Spread	0.72	0.66	0.72	0.82	1	0.16	0.14	0.10	0.17	0.15
Comomentum	0.11	0.08	0.13	0.24	0.16	1	-0.21	0.54	0.64	0.20
MktRf(36-1)	0.20	0.22	0.19	0.05	0.14	-0.21	1	-0.28	-0.19	-0.10
MktVol(6-1)	0.07	0.00	0.10	0.14	0.10	0.54	-0.28	1	0.77	0.21
MomVol(6-1)	0.06	-0.04	0.12	0.21	0.17	0.64	-0.19	0.77	1	0.26
CrossS.Disp(MA-3)	0.11	0.08	0.13	0.35	0.15	0.20	-0.10	0.21	0.26	1

Table 2.7: Forecasting Momentum Returns with Old Momentum Spread

This table reports the average raw and Fama-French 3-factor adjusted monthly returns of the momentum portfolio for the 1-6, 7-12, 13-24 and 25-36 months after their formation (without any rebalancing). All months are sorted into five groups based on their momentum spread realized at the formation period. 5-1 means the difference between the average returns in the group with the largest momentum spread and the group with the smallest. The sample period is from 1960 to 2018. T-stats, shown in the parentheses, are based on the standard errors clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags.

Panel A: Raw Returns								
	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
TS Group Rank	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	0.77%	(2.98)	0.69%	(3.28)	0.25%	(1.44)	0.05%	(0.28)
2	0.99%	(3.69)	0.11%	(0.50)	-0.24%	(-1.50)	-0.01%	(-0.08)
3	0.83%	(3.25)	0.29%	(1.21)	-0.04%	(-0.25)	0.20%	(0.99)
4	1.29%	(4.35)	0.61%	(2.48)	-0.40%	(-1.94)	-0.41%	(-1.81)
5	0.68%	(1.37)	-0.58%	(-1.27)	-0.96%	(-3.22)	-0.04%	(-0.13)
5_1	-0.09%	(-0.17)	-1.27%	(-2.55)	-1.20%	(-3.58)	-0.09%	(-0.26)

Panel B: Fama-French 3-factor Adjusted Returns								
	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
TS Group Rank	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	1.00%	(3.99)	0.87%	(3.94)	0.36%	(2.21)	0.15%	(0.85)
2	1.29%	(4.62)	0.56%	(2.67)	-0.08%	(-0.54)	0.13%	(0.74)
3	1.34%	(5.06)	0.38%	(1.66)	0.08%	(0.48)	0.35%	(1.81)
4	1.40%	(5.00)	0.70%	(3.00)	-0.21%	(-1.17)	-0.17%	(-0.76)
5	0.99%	(2.31)	-0.09%	(-0.24)	-0.53%	(-1.90)	0.21%	(0.77)
5_1	-0.01%	(-0.03)	-0.96%	(-2.31)	-0.88%	(-2.83)	0.06%	(0.19)

Table 2.8: Comparing the Predictability of Old Momentum Spreads and Other Predictors

$$r_{t,t+i} = \underbrace{\Sigma_{k=1}^{10}\alpha_k \mathbf{1}_k}_{\text{Dummy controls of the first predictor}} + \underbrace{\beta x_{t-1}}_{\text{Second predictor}} + FF.Factors_{t+i} + \epsilon_{t,t+i}$$

This table reports the coefficients β of regressing monthly momentum strategy returns on different pairs of predictors in 1-6, 7-12, 13-24 and 25-36 months after each momentum formation batch. Each number in the table represents a point estimate of beta given one regression specification (one pair of predictors for one set of months), with its t-stat in parentheses right below it. $r_{t,t+i}$ denotes the return in month $t+i$ of the momentum strategy formed in month t. The first predictor will be used in generating dummy controls. $\mathbf{1}_k$ is the dummy which takes value 1 if the control series' value falls into the k^{th} decile and 0 otherwise. The second predictor x is normalized to have a standard deviation equal to 1. Standard errors are clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags. All second predictors are normalized to have a standard deviation equal to 1. Fama-French three factors include $MktRf_{t+i}$, SMB_{t+i}, and HML_{t+i} in month $t+i$. 5% statistical significance is indicated in bold.

Panel A: Predictive Regression Results of **Other Predictors** after controlling Old Momentum Spread

Dummy controls	**Predictor x**	Month 1-6	Month 7-12	Month 12-24	Month 25-36
Old Mom. Spread	Young Mom. Spread	-0.26%	**-0.53%**	-0.13%	0.05%
		(-0.90)	**(-2.74)**	(-1.43)	(0.40)
Old Mom. Spread	Formation Gap	-0.17%	-0.13%	-0.17%	-0.16%
		(-0.62)	(-0.67)	(-1.54)	(-1.41)
Old Mom. Spread	Formation Spread	-0.15%	**-0.43%**	-0.14%	-0.13%
		(-0.44)	**(-1.96)**	(-1.10)	(-0.93)
Old Mom. Spread	Comomentum	**-0.54%**	-0.18%	-0.10%	0.34%
		(-1.99)	(-0.99)	(-0.61)	(1.88)

Panel B: Predictive Regression Results of **Old Momentum Spread** after controlling Other Predictors

Dummy controls	**Predictor x**	Month 1-6	Month 7-12	Month 12-24	Month 25-36
Young Mom. Spread	Old Mom. Spread	0.13%	**-0.48%**	**-0.22%**	0.01%
		(0.64)	**(-2.64)**	**(-2.37)**	(0.10)
Formation Gap	Old Mom. Spread	-0.06%	**-0.58%**	**-0.23%**	0.09%
		(-0.27)	**(-3.37)**	**(-2.51)**	(0.69)
Formation Spread	Old Mom. Spread	-0.16%	**-0.52%**	**-0.25%**	0.12%
		(-0.67)	**(-3.17)**	**(-2.79)**	(1.00)
Comomentum	Old Mom. Spread	-0.10%	**-0.55%**	**-0.29%**	-0.02%
		(-0.40)	**(-2.95)**	**(-2.96)**	(-0.17)

Table 2.9: Forecasting the Monthly Momentum Strategy Returns in the First Half-Year, the Second Half-Year, and the Second Year

$$r_{t,t+i} = \beta_i Old.Mom.Sprd_{t-1} + \Sigma_n \beta_{i,n} Predictor_{n,t-1} + Controls + \epsilon_{t,t+i}$$

This table reports the coefficients β of regressing monthly momentum strategy returns on the predictors in 1-6, 7-12, and 13-24 months after each momentum formation batch. Each row represents the point estimates of betas for a specification. Panel A, B, and C reports the results based on 1-6, 7-12, and 13-24 months respectively. $r_{t,t+i}$ denotes the return in month $t+i$ of the momentum strategy formed in month t. Standard errors are clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags. All predictors are normalized to have a standard deviation equal to 1. The controls include contemporaneous Fama-French three factors, an intercept, and a trend variable.

	Panel A: Predicting Monthly Momentum Returns for Month 1-6					
Predictors	[1]	[2]	[3]	[4]	[5]	[6]
Old Mom. Sprd	-0.31%	-0.17%	-0.18%	-0.03%	-0.08%	-0.33%
	(-1.33)	(-0.69)	(-0.73)	(-0.12)	(-0.31)	(-1.46)
MktRf(36-1)	**0.74%**					0.39%
	(2.81)					(1.50)
MktVol(6-1)		**-1.13%**				**-0.82%**
		(-3.94)				**(-2.44)**
MomVol(6-1)			**-0.96%**			-0.41%
			(-3.18)			(-1.13)
CrossS.Disp(MA-3)				-0.42%		0.14%
				(-1.40)		(0.49)
Comomentum					**-0.52%**	0.20%
					(-1.99)	(0.79)

	Panel B: Predicting Monthly Momentum Returns for Month 7-12					
Predictors	[1]	[2]	[3]	[4]	[5]	[6]
Old Mom. Sprd	**-0.53%**	**-0.54%**	**-0.53%**	**-0.51%**	**-0.51%**	**-0.53%**
	(-3.07)	**(-3.12)**	**(-3.05)**	**(-2.99)**	**(-2.91)**	**(-3.19)**
MktRf(36-1)	0.06%					-0.06%
	(0.28)					(-0.28)
MktVol(6-1)		**-0.34%**				-0.46%
		(-1.97)				(-1.83)
MomVol(6-1)			-0.20%			0.10%
			(-0.92)			(0.33)
CrossS.Disp(MA-3)				-0.05%		0.06%
				(-0.21)		(0.26)
Comomentum					-0.13%	0.01%
					(-0.70)	(0.06)

	Panel C: Predicting Monthly Momentum Returns for Month 12-24					
Predictors	[1]	[2]	[3]	[4]	[5]	[6]
Old Mom. Sprd	-0.16%	**-0.29%**	**-0.30%**	**-0.27%**	**-0.27%**	**-0.18%**
	(-1.88)	**(-3.36)**	**(-3.45)**	**(-2.73)**	**(-3.01)**	**(-1.97)**
MktRf(36-1)	**-0.38%**					**-0.56%**
	(-2.03)					**(-2.94)**
MktVol(6-1)		-0.31%				-0.39%
		(-1.41)				(-1.52)
MomVol(6-1)			-0.29%			-0.28%
			(-1.32)			(-1.10)
CrossS.Disp(MA-3)				0.01%		0.15%
				(0.06)		(0.97)
Comomentum					-0.11%	0.08%
					(-0.71)	(0.53)

Table 2.10: Different Factor Loadings of Old and Young Momentum Portfolios

$$r_t = a + \Sigma_k \beta_k F_{k,t} + \epsilon_t$$

In this table, each column reports the regression coefficients β_k of regressing old or young momentum strategies on different contemporaneous factors. Old and young momentum strategies are formed with only old and young momentum stocks separately. PEAD is the return of long-short strategy based on the three-day cumulative abnormal returns around the most recent earnings announcement days following Brandt et al. (2008). VIX is the CBOE Volatility Index. HFI is the Equity Market Neutral Index (HFRIEMNI) from Hedge Fund Research website. Liquidity Level is the market liquidity measure by Pastor and Stambaugh (2003). MktRf, SMB, and HML are the three Fama-French factors. The sample period is from 1960 to 2018. T-stats, shown in the parentheses, are based on the standard errors adjusted for Newey and West (1987) correlation for 6 lags.

Regressing Old and Young Momentum Strategy Returns on Contemporaneous Factors				
Sample periods	1960-2018		1990-2017	
Factors	Old_Mom	Young_Mom	Old_Mom	Young_Mom
PEAD	**0.68**	**0.78**	**0.41**	**0.55**
	(2.85)	**(3.64)**	**(2.25)**	**(3.37)**
VIX			**-0.20**	-0.05
			(-2.25)	(-0.49)
HFI			**4.75**	**4.23**
			(7.01)	**(5.32)**
Liquidity Level			-0.03	0.01
			(-0.34)	(0.15)
MktRf	-0.20	**-0.40**	**-0.56**	**-0.64**
	(-1.18)	**(-2.65)**	**(-3.54)**	**(-3.90)**
SMB	0.23	0.59	0.12	0.51
	(0.83)	(1.65)	(0.46)	(1.53)
HML	**-0.76**	-0.42	**-1.02**	**-0.58**
	(-2.53)	(-1.21)	**(-4.42)**	**(-2.13)**

Table 2.11: Forecasting the Monthly Momentum Strategy Returns in the First Half-Year, the Second Half-Year, and the Second Year

This table reports the average Fama-French 3-factor adjusted monthly returns of the old and young momentum strategy for the 1-6, 7-12, 13-24 and 25-36 months after their formation (without any rebalancing). Old and young momentum strategies are formed with only old and young momentum stocks separately. All months are sorted into five groups based on their old momentum spread realized at the formation period. 5-1 means the difference between the average returns in the group with the largest old momentum spread and the group with the smallest. The sample period is from 1960 to 2018. T-stats, shown in the parentheses, are based on the standard errors clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags.

Panel A: Future Returns of Old Momentum Stocks								
Group Rank	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	1.18%	(4.24)	0.81%	(3.16)	0.45%	(2.20)	-0.19%	(-0.78)
2	1.71%	(5.66)	0.59%	(2.43)	-0.14%	(-0.79)	0.08%	(0.37)
3	1.63%	(5.35)	0.56%	(1.96)	-0.16%	(-0.76)	0.08%	(0.31)
4	1.64%	(5.02)	0.68%	(2.35)	-0.44%	(-1.73)	-0.45%	(-1.48)
5	1.23%	(2.58)	-0.29%	(-0.69)	-0.66%	(-1.86)	0.67%	(1.94)
5_1	0.04%	(0.08)	-1.10%	(-2.29)	-1.10%	(-2.79)	0.86%	(2.11)

Panel A: Future Returns of Young Momentum Stocks								
Group Rank	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	0.93%	(3.64)	0.84%	(3.85)	0.22%	(1.40)	0.27%	(1.55)
2	1.12%	(3.93)	0.65%	(2.72)	0.02%	(0.13)	0.12%	(0.72)
3	1.23%	(4.80)	0.34%	(1.50)	0.14%	(0.84)	0.43%	(2.13)
4	1.30%	(4.77)	0.68%	(2.81)	-0.01%	(-0.08)	-0.04%	(-0.20)
5	1.01%	(2.43)	0.10%	(0.30)	-0.44%	(-1.61)	-0.26%	(-0.99)
5_1	0.09%	(0.19)	-0.74%	(-1.83)	-0.66%	(-2.15)	-0.53%	(-1.74)

Table 2.12: Percentage of Momentum Age in Momentum Stocks.

This table reports the formation spreads (t-12 to t-2) for all momentum stocks, young momentum stocks and old momentum stocks. I sort all months into three groups based on their old momentum spread and then I calculate the formation spreads for different groups of momentum stocks.

	Formation Spreads (t-12 to t-2)		
Old Mom. Spread Group	All	Young	Old
<30%	98.2%	87.1%	113.9%
30%-70%	127.4%	110.4%	151.0%
>70%	182.1%	151.9%	217.2%

Table 2.13: Subsample Analysis: Predicting Momentum Performance with Old Momentum Spread

This table reports the average raw monthly returns of the momentum portfolio for the 1-6, 7-12, 13-24 and 25-36 months after their formation (without any rebalancing) for the two subsample periods. All months are sorted into five groups based on their momentum spread realized at the formation period. 5-1 means the difference between the average returns in the group with the largest momentum spread and the group with the smallest. T-stats, shown in the parentheses, are based on the standard errors clustered on the actual return months $t+i$ and across time with Newey and West (1987) correlation for 6 lags.

Panel A: Subsample 1960-1989								
	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
TS Group Rank	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	1.12%	(3.14)	0.73%	(2.57)	0.11%	(0.49)	-0.33%	(-1.60)
2	1.45%	(4.57)	0.34%	(1.29)	-0.03%	(-0.15)	-0.18%	(-0.84)
3	1.37%	(4.10)	0.14%	(0.45)	-0.21%	(-1.05)	-0.49%	(-2.36)
4	1.36%	(4.13)	0.38%	(1.17)	-0.22%	(-0.90)	0.13%	(0.47)
5	0.63%	(1.20)	0.15%	(0.27)	-0.72%	(-2.28)	-0.72%	(-2.22)
5_1	-0.49%	(-0.80)	-0.58%	(-0.99)	-0.83%	(-2.24)	-0.38%	(-1.03)

Panel B: Subsample 1990-2019								
	Month 1-6		Month 7-12		Month 12-24		Month 25-36	
TS Group Rank	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
1	0.31%	(0.81)	0.68%	(2.04)	0.15%	(0.54)	0.47%	(1.57)
2	0.51%	(1.48)	0.01%	(0.02)	0.00%	(0.02)	0.58%	(1.82)
3	0.32%	(0.66)	0.23%	(0.64)	-0.29%	(-0.94)	0.10%	(0.27)
4	1.31%	(2.64)	0.39%	(1.02)	-0.25%	(-0.80)	-0.22%	(-0.70)
5	0.69%	(0.81)	-0.83%	(-1.09)	-1.32%	(-2.80)	0.45%	(1.03)
5_1	0.38%	(0.42)	-1.51%	(-1.84)	-1.46%	(-2.85)	-0.02%	(-0.03)

Table 2.14: Out-of-Sample R-squared for Different Predictors

$$R^2_{OOS} = 1 - \frac{\Sigma(r_t - \hat{r}_t)}{\Sigma(r_t - \overline{r}_t)}$$

This table reports the Out-of-Sample R-squared for three predictors: Old Momentum Spread, Formation Gap and Comomentum. The look-back years report how many years of data is used for training the predictive model. The Out-of-Sample R-squared statistic is calculated following Campbell and Thompson (2007).

	Predictor: Old Momentum Spread		
Look-back Years	Month 1-6	Month 7-12	Month 13-24
5	-0.64%	3.28%	-2.73%
10	-0.22%	1.87%	-1.20%
20	-0.22%	2.45%	-0.11%

	Predictor: Formation Gap		
Look-back Years	Month 1-6	Month 7-12	Month 13-24
5	-7.12%	1.69%	-5.28%
10	-1.50%	-1.66%	-0.31%
20	-0.46%	0.05%	0.07%

	Predictor: Comomentum		
Look-back Years	Month 1-6	Month 7-12	Month 13-24
5	-2.47%	-4.47%	-3.33%
10	-1.43%	-3.51%	-2.40%
20	-0.17%	-1.14%	-0.03%

Table 2.15: Combo Factor Returns.

This table reports the average returns of the combo strategy, for 1-6, 7-12, 13-24 and 25-36 months. All months are sorted into two groups based on their old momentum spread ⊥ realized in the last month, and its relative value compared to its MA in the last 12 month.

1965-2010					
<MA12			>MA12		
Variable	**Mean**	**t Value**	**Variable**	**Mean**	**t Value**
Combo_1_3	**1.0%**	**3.22**	**Combo_1_3**	0.5%	1.64
Combo_4_6	**1.2%**	**3.73**	**Combo_4_6**	0.4%	1.2
Combo_7_12	**1.7%**	**3.99**	**Combo_7_12**	**1.7%**	**3.17**
Combo_13_24	**3.4%**	**5.09**	**Combo_13_24**	**4.1%**	**5.18**
Combo_25_36	**2.6%**	**3.6**	**Combo_25_36**	**5.3%**	**7.46**

Table 2.16: Combo Factor Returns in 1990-2010.

This table reports the average returns of the combo strategy, for 1-6, 7-12, 13-24 and 25-36 months. All months are sorted into two groups based on their old momentum spread $\perp$ realized in the last month, and its relative value comparing to its MA in the last 12 month.

1990-2010					
<MA12			>MA12		
Variable	**Mean**	**t Value**	**Variable**	**Mean**	**t Value**
Combo_1_3	0.7%	1.24	**Combo_1_3**	-0.3%	-0.4
Combo_4_6	0.9%	1.65	**Combo_4_6**	-0.4%	-0.6
Combo_7_12	0.1%	0.24	**Combo_7_12**	1.2%	1.15
Combo_13_24	0.0%	0.01	**Combo_13_24**	**3.7%**	**2.37**
Combo_25_36	0.9%	0.69	**Combo_25_36**	**3.0%**	**2.27**

Chapter 3

Anomaly Investing: Out-of-Sample Performance and Intertemporal Considerations

I first show that the naïve equal-weighted 1/N investing in the set of 34 stock market anomalies is a robust implementation for out-of-sample diversification. Two types of popular portfolio optimization methods, including Sharpe-Ratio-optimizing with weight constraints and Dimension-Reduction with machine learning techniques, do not achieve robustly higher out-of-sample performance. Further to explore the gains and risks in investing stock market anomalies, I take this equal-weighted anomaly portfolio to an intertemporal CAPM framework with stochastic volatility to understand the investment considerations of a specific anomaly investor. Based on my estimation, only the correlation-induced volatility news carries a significant risk premium, which highlights the economic importance of the comovement in anomaly asset prices.

JEL-Classification: G11, G23.

Keywords: Portfolio Strategy, Diversification, Machine Learning, Hedging Demand of Sophisticated Investors.

3.1 Introduction

One important question in the anomaly investing is how to construct the optimal out-of-sample portfolio when combining all the anomalies. The groundbreaking research of Markowitz (1952) shows that investors with mean-variance preferences should allocate capital based on the expected returns and covariance, but it has notoriously unstable out-of-sample behaviors as the estimation error in the moments tends to lead to extreme portfolio weights. Though different methods and modifications have been proposed to overcome the measurement error issue, the equal-weighting 1/N rule still serves as a robust benchmark as argued in DeMiguel et al. (2007).

In the sample of 34 stock market anomalies, I explore two sets of mean-variance-efficient-optimized (MVE-optimized) portfolios: Sharpe-Ratio-based portfolios which are designed to maximize the Sharpe ratio, and Dimension-Reduction-based portfolios which are proposed to shrink the cross-section of a large number of anomaly portfolios. I find that they do achieve a higher out-of-sample Sharpe ratio in a long sample between 1983 to 2017 comparing to the 1/N equal-weighted anomaly (EAR) portfolio, but this outperformance is coming with a higher portfolio turnover among anomalies. This implies the outperformance is not likely to survive the transaction costs. Moreover, in the second half of my sample, i.e. from 2000 to 2017, I find that EAR actually achieves the mean-variance efficiency since regressing other MVE-optimized portfolios on the EAR does not leave significantly positive intercepts. This motivates me to take the conservative but robust EAR portfolio to explore the intertemporal consideration in anomaly investing in the second half of this paper.

Guo (2018) shows that the time-varying average correlation among stock market anomalies, which he terms as CoAnomaly, predicts the future anomaly returns as well as the future variance of the aggregate anomaly portfolio. In other words, this state variable, CoAnomaly, predicts two directions in the change of the investment opportunity for anomaly investors at the same time: the benefit from higher expected returns and the loss from large aggregate variance. To explore which of these two effects is stronger, I study the intertemporal hedging demand of a particular arbitrageur who seeks market-neutral absolute returns and invests in these long-short stock market anomalies in a more general framework.

Motivated by Campbell et al. (2017), I use vector autoregressions (VAR) and estimate an intertemporal CAPM with stochastic volatility for factor investing with the focus on 34 stock market anomalies. To accommodate the rebalancing nature of these anomaly portfolios, I construct the cash-flow and discount-rate news following a bottom-up approach, i.e. news estimated on stock levels and then aggregated to anomaly level. After aggregating the news to the portfolio level, I find that anomalies returns are mainly driven by cash-flow news.

Since the aggregate portfolio volatility can be decomposed into the average variance of the constituents and the correlations among the constituents, I study the pricing effects of the different volatility proxies separately: aggregate variance, average-variance-driven variance, and CoAnomaly-driven variance. Of the three, I find evidence that only the news on CoAnomaly-driven variance gets robustly priced in my estimation.

These facts are consistent with the negative risk premium of CoAnomaly documented in Guo (2018) since I show that, in my estimation, the anomaly investors care about the CoAnomaly-driven volatility risk premium and the risk premium on the discount rate news in close to zero. Moreover, I find that these anomalies are profiting by loading on the CoAnomaly-driven volatility risk, which might explain why their 'abnormal' returns are sustained in equilibrium instead of being fully eliminated. This could be due to the limited diversification on these correlated anomalies, which is in line with the limits of arbitrage literature.

The interpretation and extrapolation of the above results should be done with care since the intertemporal CAPM is a partial equilibrium model and the estimation of the model is based on the assumption that the specific anomaly investor is holding the EAR portfolio. That being said, the results clearly show that the comovement within the investment set of anomaly investors will affect their marginal utility.

Related Literature The optimal diversification question has been a central research topic in the finance literature. As the estimation error in the standard mean-variance optimization will generate unstable and extreme weights on assets, both Bayesian approaches and non-Bayesian approach has been proposed to address this issue[1]. DeMiguel et al. (2007) surveyed 14 different models and compared their performance with the simple 1/N equal-weighting diversification and find that the latter is a robust benchmark within their samples. That being said, they argue that their study is '*not to advocate the use of the 1/N heuristic as an asset-allocation strategy, but merely to use it as a benchmark to assess the performance of various portfolio rules proposed in the literature*'. Compared with their overwhelming results based on model parameters estimated from monthly returns, I find that daily returns will deliver less estimation error and more robust performance for the MVE-optimized models in the long sample.

Recently, the finance literature has started to explore the impacts and efficiency of machine learning techniques which have been used in practice for years. Gu et al. (2018) synthesize several machine learning techniques with predicting asset and factor returns and find that they do add value in terms of achieving positive out-of-sample positive return prediction R^2. Kozak et al. (2017) consider a robust stochastic discount factor (SDF) that summarizes the joint explanatory power of a large number of cross-sectional stock return predictors with the motivation from machine learning literature. However, I do not find these techniques outperform the 1/N strategy robustly, especially in the recent sample periods.

Many researches have documented that anomaly returns are predictable empirically (see Cohen et al. (2003), Stambaugh et al. (2012), Greenwood and Hanson (2012), and Novy-Marx (2014) among the others). Guo (2018) has shown that the time-varying average correlation among anomalies has strong predictive power to both the future average returns and the future volatility of these anomalies. It is worth to explore the intertemporal hedging considerations of these anomaly investors given the investment opportunity is time-varying. The intertemporal CAPM with stochastic volatility developed by Campbell et al. (2017) provides a comprehensive framework to study the intertemporal considerations as both the first and second moments of asset returns are time-varying. This topic has also been explored from a theoretical perspective (see Kondor and Vayanos (2019) among the others).

[1] Both Pástor (2000) and Pástor and Stambaugh (2000) establish their prior based on asset-pricing models. Stutzer (1995) and Ghosh et al. (2016) construct their stochastic discount factors based on the information criterion

3.2 Robust Anomaly Investing

In this section, I first explore whether an anomaly investor can achieve better performance by deviating from the naïve 1/N diversification. I found that the simple equal-weighted portfolio by averaging all anomalies is not dominated by other more advanced techniques of searching for the mean-variance-efficient-optimized portfolios, including the Sharpe-Ratio-based optimization with norm constraints, and Dimensional-Reduction-based efficient portfolios with machine learning methods.

The 34 anomalies I consider here are from the same set studied in Guo (2018), which is a union set of anomalies studied in Stambaugh et al. (2012) and Novy-Marx and Velikov (2016) and includes Accruals (acc), Asset growth (atgrowth), Asset turnover (ato), Beta arbitrage (beta), Composite equity issues (ceissue), Failure probability (failprob), Gross margins (gm), Industry mom. + Relative rev. (hfcombo1), Industry mom. + Relative rev. + Season. (hfcombo2), Idiosyncratic volatility (idiovol), Industry momentum (indmom1m), Investment (invest), Momentum (mom12m), Net issuance annual (netissue_a), Net issuance monthly (netissue_m), Net operating assets (netoa), Ohlson's O-score (ohlson), PEAD(CAR3) (peadcar3), PEAD(SUE) (peadsue), Piotroski's f-score (piotroski), Gross profitability (profit), Industry relative reversals (relrev1m), Short-run reversals low volatility (relrev1mlow), Short-run reversals (rev1m), Long-run reversals (rev60m), Return-on-assets (roa), Return-on-book equity (roe), Return-on-market equity (rome), Seasonality (seasonal), Size (size), Value + Momentum (valmom), Value + Mom + Prof (valmomprof), Value + Profitability (valprof), Value (value). The labels for the anomalies are reported in parentheses.

The whole sample is from 1973 to 2017 with both monthly returns as well as daily returns. Anomalies are based on monthly, quarterly or annually rebalancing and please see Guo (2018) for more details.

3.2.1 Equal-Weighted and Sharpe-Ratio-Based Diversification

How to construct the diversified anomaly portfolio optimally is certainly a key question for factor investors in practice. This is also important for finance academics as the mean-variance efficient portfolio contains the same pricing information as the stochastic discount factor (SDF, see Hansen and Jagannathan (1991)). Different from the standard investment question in the aggregate market, anomaly portfolios do not have a natural equilibrium weight like the market capitalization. For academic researchers as well as arbitrageurs, it is extremely difficult to infer how much capital was allocated in each strategy at each point of time. On the other hand, the variance-based efficient portfolio back to Markowitz (1952) suffers the problem of measurement error in the expected returns and the covariance matrix, and it also ignores other practical issues like price impacts. Here I propose several weighting candidates on anomalies and compare their out-of-sample performances.

Naïve 1/N Diversification A natural starting point is to put equal weight on each of these anomalies. Overweighting some small and illiquid stocks is less of concern as I construct these anomalies from stocks based on value-weighting and excluded the stocks which are in the bottom decile of NYSE breakpoint. In this case of 34 anomalies, I assign a weight 1/34 to each one of them.

Sharpe-Ratio-Based Mean-Variance Efficient Portfolios If we believe that the past covariance structure is precisely estimated and persistent in the near future, we can form the mean-variance efficient portfolio based on the estimated expected returns and the covariance matrix.

The standard mean-variance optimization is notorious for producing extreme weights which is quite unstable across time, mainly due to the estimation error in the first and second moments. Attempts have been made to handle the measurement errors and to improve the out-of-sample performance of the Markowitz model. The vast literature on this issue includes the Bayesian approach (diffuse priors and shrinkage estimators), robust portfolio allocation rules, imposing short-selling and other constraints, etc.

Here I consider a simple box-constraints by bounding the weight on each anomaly within a cerntain positive range to avoid extreme weights. I solve the following maximization problem,

$$
\begin{aligned}
\max_{w} \quad & (w'\mu - r_f)/(w'\Sigma w) \\
s.t. \quad & w'\vec{1} = 1 \\
\text{and} \quad & w_i \in [w_{min}, w_{max}] \quad \text{for all } w_i \text{ in } w
\end{aligned}
$$

where w is the vector of portfolio weights, μ is the vector of expected return, r_f the risk-free rate, and Σ is the total covariance matrix.

3.2.2 Dimension-Reduction-Based Mean-Variance Efficient Portfolios

Machine learning, or statistical learning, has been introduced to the investment process and now has been widely used across the industry. Due to the versatility of different machine learning techniques and the validation of the hyperparameters, these methods normally generate strong out-of-sample predictability. However, the shortcoming is evident given that these methods are purely statistical and data-based without using explicit instructions.

Recently, finance academics have started to look into the performance and impacts of machine learning methods. Among different studies, Kozak et al. (2017) consider a robust stochastic discount factor (SDF) that summarizes the joint explanatory power of a large number of cross-sectional stock return predictors. As the SDF naturally maps to the mean-variance efficient portfolio which achieves the highest Sharpe ratio, this method fits the purpose of identifying the optimal weights in the anomaly portfolio.

Starting in a Bayesian setting to estimate an SDF in the linear span of factor returns F_t, $M_t = 1 - b'(F_t - \mathbb{E}F_t)$, they propose a prior on the expected future factor returns μ,

$$\mu \sim \mathcal{N}(0, \frac{\kappa^2}{\tau}\Sigma^2), \tag{3.1}$$

where $\tau = tr[\Sigma]$ and κ is a scaling constant. Following the eigendecomposition of the covariance matrix $\Sigma = QDQ'$, where Q is the matrix of eigenvectors and D is the diagonal matrix of decreasing-ordered eigenvalues, the above prior is equivalent to

$$D^{-\frac{1}{2}}\mu_P \sim \mathcal{N}(0, \frac{\kappa^2}{\tau}D), \tag{3.2}$$

where $\mu_P = Q'\mu$ is the expected returns of the principal component portfolios. Effectively, this prior implies that we are more uncertain about the non-zero Sharpe ratio (the left-hand side of Equation 3.2) on the principal components associated with large eigenvectors and consequently we are more willing to accept that the principal components with large eigenvectors carry a large Sharpe ratio estimated from the data. As Kozak et al. (2018) argue, absence of near-arbitrage opportunities implies that the high Sharpe ratio PCs must coincide the PCs with high eigenvalues.

Moreover, they Kozak et al. (2017) show that this Bayesian estimator maps into a L-2 penalized machine learning estimator which minimizes the HJ distance,

$$\hat{b} = arg \min_b \quad \{(\overline{\mu} - \Sigma b)'\Sigma^{-1}(\overline{\mu} - \Sigma b) + \gamma b'b\}, \tag{3.3}$$

where $\gamma = \frac{\tau}{\kappa^2 T}$ and T is the sample periods. As $\kappa = \mathbb{E}[\overline{\mu}'\Sigma^{-1}\overline{\mu}]^{-\frac{1}{2}}$, which is the expected maximum Sharpe Ratio under the prior, Kozak et al. (2017) bring economic interpretation to the penalty hyperparameter γ in this machine learning field. I also entertain the above machine learning problem with sparsity, which introduces a penalty on the L-1 norm and helps to select a subset of relevant factors.

$$\hat{b} = arg \min_b \quad \{(\overline{\mu} - \Sigma b)'\Sigma^{-1}(\overline{\mu} - \Sigma b) + \gamma_1 b'b + \gamma_2 \Sigma_{i=1}^N |b_i|\}. \tag{3.4}$$

This method is known as *elastic net* by Zou and Hastie (2005).

Implementation I estimate two sets of portfolio weights, one based on the anomaly portfolios and the other based on the principal components calculated from the anomaly portfolios. I start from the HJ-distance in the optimization problem, which can be rewritten as

$$
\begin{aligned}
(\overline{\mu} - \Sigma b)'\Sigma^{-1}(\overline{\mu} - \Sigma b) &= (\overline{\mu} - \Sigma b)'QD^{-1}Q'(\overline{\mu} - \Sigma b) \\
&= [D^{-\frac{1}{2}}Q'(\overline{\mu} - \Sigma b)]'[D^{-\frac{1}{2}}Q'(\overline{\mu} - \Sigma b)] \\
&= [D^{-\frac{1}{2}}Q'\overline{\mu} - D^{-\frac{1}{2}}Q'\Sigma b)]'[D^{-\frac{1}{2}}Q'\overline{\mu} - D^{-\frac{1}{2}}Q'\Sigma b)] \\
&= [D^{-\frac{1}{2}}Q'\overline{\mu} - D^{-\frac{1}{2}}Q'\Sigma Q b_P)]'[D^{-\frac{1}{2}}Q'\overline{\mu} - D^{-\frac{1}{2}}Q'\Sigma Q b_P)].
\end{aligned}
$$

Effectively, I can estimate b by using penalized linear regression with $D^{-\frac{1}{2}}Q'\overline{\mu}$ on $D^{-\frac{1}{2}}Q'\Sigma$ (termed as DR_raw), and estimate b_P by using penalized linear regression with $D^{-\frac{1}{2}}Q'\overline{\mu}$ on $D^{-\frac{1}{2}}Q'\Sigma Q$ (termed as DR_pca). Finally, I normalize the coefficient to have a summation of weight 1, $w = \frac{1}{\Sigma_{i=1}^{N} b_i} b$. Similar normalization can be done starting from $b = Qb_P$ for DR_pca.

3.2.3 Comparing Different Weighting Methods

Estimation with Past Information Only: Look-back Choices I am interested in the performance of different mean-variance efficient-optimized (MVE-optimized) portfolios out-of-sample. So at the beginning of each month, I calculate the weights based on only the past information.

I consider three look-back choices: (1) using the past 10 years as a rolling window to estimate the mean and covariance of anomaly portfolios (following DeMiguel et al. (2007) as a comparable benchmark), (2) using the past 10 years to estimate the covariance and using all past periods to estimate the mean, and (3) using all past periods to estimate the mean and covariance.

Due to the estimation in the look-back periods with a minimum of 10 years, the results are based on anomaly returns starting from 1983 to 2017. Different from DeMiguel et al. (2007), I use daily, instead of monthly, anomaly returns to estimate model parameters. This choice of high-frequency daily returns will produce more precise estimation in the second moments as argued in Kozak et al. (2017).

Since I am studying the long-short zero-cost anomaly portfolios, I start with the risk-free rate $r_f = 0$, though my results are robust of using other specifications. For the maximizing Sharpe ratio problem, I consider three different box constraints for each weight: relatively *unconstrained* $w_i \in [-1, 1]$ (SR_[-1,1]), *non-negative* weights

$w_i \in [0, 1]$ (SR_[0,1]), and *capped* weights $w_i \in [0.01, 0.1]$ (SR_[0.01,0.1]). The first *unconstrained* case (SR_[-1,1]) is relative in the sense that I still place a constraint that requires the absolute value of every weight being no greater than 1. If this mild constraint is removed, the standard mean-variance optimization will generate unreasonably extreme weights and huge turnover in most cases, for example, a monthly turnover as large as 606594.36% for industry portfolios as shown in DeMiguel et al. (2007). As for the dimension reduction with penalized regression of raw anomaly portfolios (DR_raw) or principal components (DR_pca), I do not tune the hyper-parameters γ_1 and γ_2. I choose the $\kappa = 3$ as the prior belief about the maximum Sharpe ratio in the economy and the weight to sparsity penalty 0.05. Note that, in the potential validation process, it does not necessarily require to maintain the temporal order as Bergmeir and Benítez (2012) point out.

I use the daily anomaly returns to estimate the weights and I first orthogonalize all anomaly returns with respect to the market using betas estimated within the past estimation period. Given the estimated weight $\hat{w}$ at the beginning of each month, I construct the daily time-series of mean-variance efficient portfolios with $P_t = \hat{w}_t F_t$ within that month. I repeat this step every month.

I also report the monthly turnover on the anomaly level for each MVE-optimized portfolio. The turnover is defined as

$$\text{Turnover} = \frac{1}{T}\Sigma_{t=1}^{T}\Sigma_{n=1}^{N}|\hat{w}_{n,t+1} - \hat{w}_{n,t^+}|, \tag{3.5}$$

where $w_{n,t+1}$ is the weight of anomaly n for month $t + 1$ after rebalancing and w_{n,t^+} is the weight for month $t+1$ before balancing, which is the weight from month t times the aggregate return of anomaly n in month t.

Out-of-Sample Performance

(Insert Table 3.1)

Table 3.1 reports the time-series average of weights on each anomaly that the five different MVE methods assigned based on the look-back choice (2) of using the past 10 years to estimate the covariance and using all past periods to estimate the mean. At the end of the table, I report the simple cross-anomaly average of the standard deviation, max, and min. Effectively, the average standard deviation is a proxy for the turnover on anomaly level of each MVE portfolios. Sharpe-Ratio-based MVE methods (SR_m1top1, SR_0to1, and SR_001to01) clearly tend to have higher turnovers than the Dimension-Reduction-based MVE methods (DR_raw and DR_pca). It is also worth mentioning that these two families of methods disagree on certain anomalies. For example, size gets relatively high weights from Sharpe-Ratio-Based MVE methods while Dimension-Reduction-Based MVE methods put higher weights on 1-month industry momentum.

(Insert Table 3.2)

Panel A of Table 3.2 reports the annualized average returns and Sharpe ratio of the five MVE methods from three different look-back choices. The monthly turnover is also reported in the last column. I also normalize the weights so that each portfolio has the summation of the weights on 34 anomalies up to 1. If the MVE weights are only based on the past 10 years to estimate the mean and covariance of anomaly portfolios as in look-back choice (1), the unconstrained the Sharpe-Ratio-Based MVE method (SR_m1top1), as well as Dimension-Reduction-based MVE methods produce low out-of-sample Sharpe ratios, all below 1, with high turnover ratios. On the contrary, two Sharpe-Ratio-Based MVE methods with constraints on weights (SR_[0,1] and SR_[0.01,0.1]) produce high out-of-sample Sharpe ratios.

If I use the past 10 years to estimate the covariance and use all past periods to estimate the mean as in look-back choice (2), the poor performances of the unconstrained the Sharpe-Ratio-Based MVE method (SR_m1top1), as well as Dimension-Reduction-based MVE methods, are improved quite a lot. If I use all past periods to estimate the mean and the covariance as in look-back choice (3), the out-of-sample performances of all methods are highest, though the improvement is marginal from look-back choice (2).

This result highlights the time-varying average returns in the cross-section of anomalies and shows that it is better to use all available information to estimate the mean. It partially shares the same insight with Kozak et al. (2017) since their results focus on the uncertainty or the estimation error in the means of returns. It also demonstrates the superiority of using daily data in the estimation which is ignored by DeMiguel et al. (2007), at least within my sample of 34 anomalies.

Panel B of Table 3.2 reports the same results with the focus on the second half of my sample, 2000-2017. The pattern discussed above remains similar to the whole sample while the magnitude of returns and Sharpe ratio are almost halved for most cases.

Table 3.2 also reports the annualized average returns and Sharpe ratio of the equal-weighted anomaly portfolio (EAR) at the end of each panel. It clearly shows that its Sharpe ratio is relatively low compared to other MVE portfolios. Next, I evaluate different MVE portfolios with respect to the EAR portfolio statistically.

Do they outperform the 1/N weighting significantly? Now I focus the on the pricing ability of the naïve equal-weighted anomaly (EAR) portfolio. If its pricing power strong, or in other words, if it is close to the mean-variance efficient portfolio, then the intercept of regressing other MVE-optimized portfolios on EAR shall not be significantly different from zero. Here I run the following regression:

$$R_t = \alpha + \beta_m MktRf_t + \beta_e EAR_t + \epsilon_t,$$

where R_t is the daily return of each of the MVE-optimized portfolios. Here I focus on the MVE-optimized portfolios based on the look-back choice (2), using the past 10 years to estimate the covariance and using all past periods to estimate the mean.

(Insert Table 3.3)

Table 3.3 reports the results of the above regression for the whole sample 1983-2017 as well as the second half-sample. The regression is based on daily returns and the intercepts are reported as annualized returns. The standard errors with Newey and West (1987) correction for 10 lags are reported in the parentheses.

It clearly shows that in the full sample, the EAR, together with the market excess return, cannot price the five MVE-optimized portfolios as large positive intercepts are left unexplained relative to the standard errors. However, if we look at the second half sample, EAR has stronger pricing power in the last two decades. The Sharpe-Ratio-based unconstrained portfolio (SR_[-1,1]) is the only MVE-optimized portfolio that has a significant alpha left unexplained. This is not surprising considering the high turnover of this MVE-optimized portfolio (see Table 3.1 and Table 3.2) and I did not take into account the transaction cost of trading these anomalies. If the transaction costs are accounted for, it is not clear that the Sharpe-Ratio-based unconstrained portfolio (SR_[-1,1]) can outperform the EAR.

In a nutshell, I find that in my 34-anomaly-investing setting, the naïve 1/N EAR portfolio serves as a robust benchmark out-of-sample. This finding echoes the results in DeMiguel et al. (2007), where they conduct a more comprehensive analysis by comparing 14 models on 7 empirical datasets. Next, I will take this EAR benchmark to study the intertemporal consideration of a specific anomaly investor.

3.3 Intertemporal CAPM for Market-Neutral Investing

Guo (2018) provides evidence that the time-varying average correlation among anomalies, CoAnomaly, is a strong predictor of the anomaly returns, and on the other hand, it has a mechanical link to the portfolio volatility and predicts the future variance of trading these anomalies as well. From an intertemporal perspective, the first fact is an increase in the investment opportunity as the level of future return is higher, so CoAnomaly should carry a positive price of risk, holding other things constant (discount rate channel); however, the second fact is a deterioration of investment opportunity as the future variance is higher, so CoAnomaly should be negatively priced across assets (volatility channel). This section makes a step further to study the intertemporal hedging demand of the anomaly investors in a more general framework, and most importantly, to see: first, which risk matters for these arbitrageurs, and second, why CoAnomaly carries a negative price of risk.

As Maio and Santa-Clara (2012) argue, the ICAPM places restrictions on the behavior of the state variable: if a state variable forecasts positive changes in the investment opportunities, its innovation should carry a positive price of risk. On the other hand, if the state variable forecasts the increase in the volatility, its price of risk should be negative. Given the empirical fact that CoAnomaly innovation is negatively priced in the cross-section as shown in Guo (2018), it suggests that the loss of investment diversification exceeds the benefit of higher future anomaly returns, which is confirmed in my analysis: the discount-rate news carries a negligible risk premium, similar to the findings in aggregate stock market by Campbell and Vuolteenaho (2004), and CoAnomaly-driven variance news has a large and negative risk premium.

3.3.1 Stochastic Volatility Setting

Campbell et al. (2017) consider a investor with Epstein and Zin (1991) recursive utility and can write the investor's value function as

$$U_t = \left\{(1-\delta)C_t^{\frac{1-\gamma}{\theta}} + \delta E_t[U_{t+1}^{1-\gamma}]^{\frac{1}{\theta}}\right\}^{\frac{\theta}{1-\gamma}}$$

where γ is the relative risk aversion (RRA) parameter, $\theta = \frac{1-\gamma}{1-1/\psi}$ and ψ is the intertemporal elasticity of substitution (IES). RRA measures the willingness to substitute consumption across states of nature, and IES measures willingness to substitute over time.

Epstein and Zin (1991) show that this utility specification leads to the Euler equation

$$E_t\left[\delta^\theta \left(\frac{C_{t+1}}{C_t}\right)^{\frac{\theta}{\psi}} \left(\frac{1}{R_{W,t+1}}\right)^{1-\theta} R_{t+1}\right] = 1,$$

where $R_{W,t+1} = W_{t+1}/(W_t - C_t)$ is the return on a claim to the wealth, Epstein and Zin (1991) use stock market index return as a proxy. The corresponding stochastic discount factor can be written as

$$M_{t+1} = \delta^\theta \left(\frac{C_{t+1}}{C_t}\right)^{\frac{\theta}{\psi}} \left(\frac{W_t - C_t}{W_{t+1}}\right)^{1-\theta} \tag{3.6}$$

Campbell and Vuolteenaho (2004) assume homoscedasticity of market returns, so all discount rate shocks are coming from shocks to the risk-free rate and they cannot generate time-varying risk premium. Campbell et al. (2017) expand this to heteroscedasticity by considering time-varying volatility. They rewrite the innovation in the log SDF as

$$m_{t+1} - \mathbb{E}_t[m_{t+1}] = \frac{\theta}{\psi}(h_{t+1} - \mathbb{E}_t[h_{t+1}]) - \gamma(r_{t+1} - \mathbb{E}_t[r_{t+1}]) \tag{3.7}$$

where $h_{t+1} = ln(W_{t+1}/C_{t+1})$. Solving forward with the assumption that asset returns and all state variables are jointly normal and the SDF should price the return on the wealth portfolio $0 = ln\mathbb{E}_t[exp\{m_{t+1} + r_{t+1}\}] = \mathbb{E}_t[m_{t+1} + r_{t+1}] + \frac{1}{2}Var_t[m_{t+1} + r_{t+1}]$, they get

$$\begin{aligned} h_{t+1} - \mathbb{E}_t[h_{t+1}] &= (\psi - 1)(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &+ \frac{1}{2}\frac{\psi}{\theta}(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j Var_{t+j}[m_{t+1+j} + r_{t+1+j}] \\ &= (\psi - 1)N_{DR,t+1} + \frac{1}{2}\frac{\psi}{\theta}N_{RISK,t+1}, \end{aligned} \tag{3.8}$$

where ρ is the loglinearization parameter[2], the discount rate news $N_{DR,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}$, and the news of future risk $N_{RISK,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j Var_{t+j}[m_{t+1+j} + r_{t+1+j}]$.

Rearrange above two equations with the present-value identity that $r_{t+1} - \mathbb{E}_t[r_{t+1}] = N_{CF,t+1} - N_{DR,t+1}$ (see Campbell and Shiller (1988), Campbell (1991) and Campbell and Vuolteenaho (2004)), they get

$$\begin{aligned} m_{t+1} - \mathbb{E}_t[m_{t+1}] &= -\gamma[r_{t+1} - \mathbb{E}_t[r_{t+1}]] - (\gamma - 1)N_{DR,t+1} + \frac{1}{2}N_{RISK,t+1} \\ &= -\gamma N_{CF,t+1} - (-N_{DR,t+1}) + \frac{1}{2}N_{RISK,t+1}, \end{aligned} \tag{3.9}$$

where the cash-flow news $N_{CF,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) = \sum_{j=1}^{\infty} \Delta d_{t+1+j}$ with Δd the log dividend growth.

Campbell et al. (2017) further write the news about the future risk into the news about the future volatility by assuming that there is a common risk component that governs time-variation in all shocks, which will drive the log SDF in a linear way in

[2] Campbell et al. (2017) show that $\rho \approx 1 - C/W$, and the economic meaning of which is how much of wealth in proportion is reinvested each period. I set it to be 0.95 per annual, equivalently 0.98726 per quarter.

their log-linear model.

$$\begin{aligned}
News_{Risk,t+1} \quad &= (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j Var_{t+j}[m_{t+1+j} + r_{t+1+j}] \\
&= (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j (\omega \sigma_{t+j}^2) \\
&= \omega(\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j (\sigma_{t+j}^2) \\
&= \omega News_{Volatility,t+1} \quad \text{(or} \quad \omega News_{V,t+1}).
\end{aligned} \tag{3.10}$$

where the ω solves

$$\omega \sigma_t^2 = (1-\gamma)^2 Var_t[N_{CF,t+1}] + \omega(1-\gamma) Cov_t[N_{CF,t+1}, N_{V,t+1}] + \frac{1}{4}\omega^2 Var_t[N_{V,t+1}]. \tag{3.11}$$

Following the moment condition that the SDF prices all assets $\mathbb{E}_t[M_{t+1}R_{i,t+1}] = 0$, the asset pricing equation in a beta representation form can be written as:

$$\begin{aligned}
\mathbb{E}_t[R_{i,t+1} - R_{f,t+1}] &= \gamma Cov_t[r_{i,t+1} - r_{f,t+1}, News_{CF,t+1}] \\
&\quad + Cov_t[r_{i,t+1} - r_{f,t+1}, -News_{DR,t+1}] - \frac{1}{2}\omega Cov_t[r_{i,t+1} - r_{f,t+1}, News_{V,t+1}] \\
&= \gamma Var_t(r_{M,t+1}) \frac{Cov_t[r_{i,t+1} - r_{f,t+1}, News_{CF,t+1}]}{Var_t(r_{M,t+1})} \\
&\quad + Var_t(r_{M,t+1}) \frac{Cov_t[r_{i,t+1} - r_{f,t+1}, -News_{DR,t+1}]}{Var_t(r_{M,t+1})} \\
&\quad - \frac{1}{2}\omega Var_t(r_{M,t+1}) \frac{Cov_t[r_{i,t+1} - r_{f,t+1}, News_{V,t+1}]}{Var_t(r_{M,t+1})} \\
&= \gamma Var_t(r_{M,t+1}) \times \beta_{i,t,CF_M} \\
&\quad + Var_t(r_{M,t+1}) \times \beta_{i,t,DR_M} \\
&\quad - \frac{1}{2}\omega Var_t(r_{M,t+1}) \times \beta_{i,t,V_M},
\end{aligned} \tag{3.12}$$

where the betas are defined as following,

$$\begin{aligned}
\beta_{i,t,CF_M} &\equiv \frac{Cov_t[r_{i,t+1} - r_{f,t+1}, News_{CF,t+1}}{Var_t(r_{M,t+1})} \\
\beta_{i,t,DR_M} &\equiv \frac{Cov_t[r_{i,t+1} - r_{f,t+1}, News_{V,t+1}]}{Var_t(r_{M,t+1})} \\
\beta_{i,t,V_M} &\equiv \frac{Cov_t[r_{i,t+1} - r_{f,t+1}, News_{V,t+1}]}{Var_t(r_{M,t+1})}.
\end{aligned} \tag{3.13}$$

The above beta representation can be conditioned down to obtain the unconditional implications, which are estimated and tested by Campbell et al. (2017).

3.3.2 Volatility Decomposition and Specification

I further divide the variance of the diversified portfolio into variances and covariances of its constituents. Here, I derive the case of equal-weighted constituents; however, this can be extended to value-weighted and other-weighted easily[3].

$$\begin{aligned}
News_{Volatility,t+1} &= (\mathbb{E}_{t+1} - \mathbb{E}_t)\sum_{j=1}^{\infty}\rho^j Var_{t+j}[\frac{1}{N}\sum_{n=1}^{N} r_{n,t+j+1}] \\
&= (\mathbb{E}_{t+1} - \mathbb{E}_t)\sum_{j=1}^{\infty}\rho^j (\sum_{n=1}^{N}\frac{1}{N^2}Var_{t+j}[r_{n,t+j+1}] \\
&\quad + \sum_{l\neq m}\frac{1}{N^2}Cov_{t+j}[r_{l,t+j+1}, r_{m,t+j+1}]) \\
&= (\mathbb{E}_{t+1} - \mathbb{E}_t)\sum_{j=1}^{\infty}\rho^j \quad (\frac{1}{N}\frac{\sum_{n=1}^{N} Var_{t+j}[r_{n,t+j+1}]}{N} \\
&\quad + \frac{N-1}{N}\frac{\sum_{l\neq m} Cov_{t+j}[r_{l,t+j+1}, r_{m,t+j+1}]}{N(N-1)}) \\
&= (\mathbb{E}_{t+1} - \mathbb{E}_t)\sum_{j=1}^{\infty}\rho^j (\frac{1}{N}\text{Avg.Variance}_{t+j+1} + \frac{N-1}{N}\text{Avg.CoVariance}_{t+j+1}).
\end{aligned} \tag{3.14}$$

This shows that volatility news can be decomposed into two parts — news about average variance and news about average covariance. Moreover, as the equation above shows, when the portfolio is fully diversified (i.e. when N is large enough), the effect from the average variance converges to zero and the average covariance

[3] A simple example: consider a portfolio with two constituents, A and B, with weights 1/3 and 2/3, respectively. It can be viewed as an equal-weighted portfolio with constituents A, $\frac{1}{2}$ B, and $\frac{1}{2}$ B.

dominates the whole effect. As the covariance can be decomposed into average variance and average correlation, I make further assumptions about the constant variances or the constant correlations across time and assets to study the different effects from either the average variance or the correlation. This *imperfect* simplification ignores the potential comovement between the level of average variance and the level of average correlation, but it may help me study the two effects separately or at least, provide suggestive evidence about which time-varying component in the aggregate variance plays a stronger role. Therefore, when N is large,

$$\begin{aligned}
&News_{Volatility,t+1} && \approx (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j \text{Avg.CoVariance} \\
&\textit{Specification 2}\text{: Constant corr.} && = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j \times \overline{\text{Avg.Correlation}} \times \text{Avg.Variance}_{t+j+1} \\
&\textit{Specification 3}\text{: Constant var.} && = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j \times \overline{\text{Avg.Variance}} \times \text{Avg.Correlation}_{t+j+1}.
\end{aligned} \tag{3.15}$$

In the estimation of the intertemporal CAPM for market neutral investing, I present results for three specifications: *Specification 1*, I use the time-varying variance of aggregate portfolio return as the stochastic volatility measure; *Specification 2*, I assume that the correlation structure between anomalies is constant through time and across asset pairs and use the time-varying average variance for anomalies as the stochastic volatility measure; *Specification 3*, I assume the average variance for anomalies is constant through time and across assets, and use the time-varying average correlation (*CoAnomaly*) as the stochastic volatility measure.

Note that through these specifications, instead of perfectly modeling the volatility dynamic, I am effectively searching for a better[4] proxy for the marginal utility (SDF m_t) of the investor from the data, since the definition of news of future risk is $N_{RISK,t+1} = (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=1}^{\infty} \rho^j Var_{t+j}[m_{t+1+j} + r_{t+1+j}]$ (see Equation 3.9). In other words, I am exploring which variation is of concern to the investor, by observing the price dynamics of the assets held by the investor.

In the appendix, I take another approach to isolate the effect on the aggregate

[4]In the sense that it has a stronger power to price the assets.

volatility coming from CoAnomaly, by constraining the VAR estimation and shutting down the feedback loop of the variance. I find that the two approaches generate consistent results.

VAR Specification and Estimation Procedure for Volatility

$$\mathbf{x}_{t+1} = \overline{\mathbf{x}} + B\mathbf{x}_t + \sigma_t \mathbf{u}_{t+1}, \tag{3.16}$$

Here, I estimate the volatility shocks on the aggregate level separately to avoid the covariance complication from bottom-up approach; however, in the appendix, I provide evidence that this separated estimation generates consistent results. I estimate a first-order aggregate VAR as in Equation 3.16, where $\mathbf{x}_{t+1}$ is a 5×1 vector of state variables with the following order:

$$\mathbf{x}_{t+1} = [\quad EVol_{t+1} \quad VS_{t+1} \quad PE_{t+1} \quad MktRf_{t+1} \quad DEF_{t+1} \quad]', \tag{3.17}$$

and given this structure, news about the future expected volatility ($EVol$) can be written as

$$N_{EVol,t+1} = e_1' \rho B (I_K - \rho B)^{-1} \sigma_t \mathbf{u}_{t+1}, \tag{3.18}$$

where e_1' is a $K \times 1$ vector, whose first element is 1 and others 0, and I_K is a identity matrix. K is the number of state variables in the VAR system, and in this case, K equals 5.

The first variable is the expected volatility of equal-weighted anomaly return $EVol_{EAR,t+1}$. This variable is meant to capture the conditional *forward-looking* volatility of the E.A.R. in the following period, so the innovation to this variable naturally links to the volatility news term above, $News_V$. The estimation procedure follows Campbell et al. (2017) closely by using a two-stage VAR regression with quarterly data. To estimate the $EVol_{EAR,t+1}$, I run a regression of the realized volatility $RVol_{EAR,t+1}$ on $RVol_{EAR,t}$ as well as other state variables at time t, and then using the predicted value for $\widehat{RVol}_{EAR,t+1}$ as the $EVol_{EAR,t}$, which only depends on information available at time t.

Other aggregate variables include: small-stock value spread (VS), which is adapted from the literature (see Campbell and Vuolteenaho (2004) and Campbell et al. (2017))[5]; PE ratio, which is the cyclically adjusted price-to-earnings ratio, downloaded from Shiller's website; market excess return (MktRf); default spread (DEF), defined as the difference between the log yield on Moody's BAA and AAA bonds, downloaded from the Federal Reserve Bank of St. Louis, which is known to track the aggregate market returns and default probabilities, reflecting news about future volatility.

In the first stage of estimating the expected volatility, I deviate from the standard OLS in three ways: first, given that heteroskedasticity is modeled directly, I estimate this regression using weighted least square (WLS), where the weight of each observation pair is based on the realized volatility in the previous quarter; second, I make sure the predicted value (expected volatility) is positive by winsorizing the fitted values, which are negative or positive but close to zero; third, I shrink the weight towards the equal weight by choosing a shrinking ratio 0.9, which means 90% of the weight is based on past volatility. The last step is to make sure my results are not driven by observations in the low volatility environment. In the second stage, I use the inverse of expected volatility in time t to weight the regression with dependent variables in time t+1, as in Equation 3.17. The estimation results are reported in Table 3.4.

(Insert Table 3.4)

3.3.3 Bottom-Up VAR Approach for Rebalanced Portfolios

Cohen et al. (2003) argue that for rebalancing portfolios, the aggregate VAR approach will generate different interpretations from simple news decomposition. Lochstoer and Tetlock (2017) also point this out by providing a model to show that inferring cash flow and discount rate shocks directly from a VAR estimated using returns and cash flows of rebalanced anomaly portfolios (trading strategies) obfuscates the underlying sources of anomaly returns. In their Appendix, they provide extreme examples in which firms' expected cash flows are constant, but direct VAR estimation suggests that all return variation in the rebalanced anomaly portfolio comes from

[5] The value spread is also a strong predictor of the value premium, which can explain some part of the premia for many anomalies, since many of them are somehow 'value-ish'.

cash flow shocks. They suggest using the stock-level decomposition through a panel VAR following Vuolteenaho (2002), and then aggregating the stock-level cash flow and discount-rate news to the portfolio level. This consideration fits the anomaly case that I am studying. Therefore, I follow their procedure and incorporate techniques from other papers in this literature, including Vuolteenaho (2002), Cohen et al. (2003), Campbell et al. (2009), and Antón (2011).

Firm-Level VAR I first conduct the firm-level VAR, mainly following Lochstoer and Tetlock (2017)[6]. Vuolteenaho (2002) derives the present-value decomposition on stock level by assuming that clean-surplus accounting and the log-linearization both hold as Ohlson (1995). Therefore, I measure log clean-surplus return on equity $lnROE^{CS}$ as:

$$lnROE^{CS}_{i,t+1} \equiv r_{i,t+1} + \rho bm_{i,t+1} - bm_{i,t}. \tag{3.19}$$

Specification I assume the firm-level expected log returns are linear functions with observables (X, including firm-level observable vector $X_{i,t}$ and aggregate level observable vector $X_{A,t}$).

$$Z_{i,t} = \begin{bmatrix} r_{i,t} - \overline{r}_{i,t} \\ X_{i,t} - \overline{X}_{i,t} \\ X_{A,t} - \overline{X}_{A,t} \end{bmatrix}$$

The vector evolves according to a VAR(1):

$$Z_{i,t+1} = AZ_{i,t} + \epsilon_{i,t+1}, \tag{3.20}$$

where $\epsilon_{i,t+1}$ is a vector of shocks to each state variable. I assume all firms are following the same VAR(1) process and the firm-level characteristics help to capture the heterogeneity between different firms.

After I estimate the VAR system, I can back out the discount rate shocks as implied by the present-value relation (see Campbell and Shiller (1988) and Campbell

[6]The main difference from Lochstoer and Tetlock (2017) is that I conduct the analysis on quarterly basis as opposed to annual basis.

(1991)):

$$
\begin{aligned}
DR^{Shock}_{i,t+1} &= (\mathbb{E}_{t+1} - \mathbb{E}_t) \sum_{j=2}^{\infty} \rho^j r_{i,t+j} \\
&= e_1' \sum_{j=1}^{\infty} \rho^j A^j \epsilon_{i,t+1} \\
&= e_1' \rho A (I_K - \rho A)^{-1} \epsilon_{i,t+1}.
\end{aligned}
\tag{3.21}
$$

I can extract the cash-flow news by subtracting the minus discount-rate news from discount rate shocks following $r_{i,t+1} - \mathbb{E}_t r_{i,t+1} = CF^{Shock}_{i,t+1} - DR^{Shock}_{i,t+1}$:

$$
\begin{aligned}
CF^{Shock}_{i,t+1} &= r_{i,t+1} - \mathbb{E}_t r_{i,t+1} + DR^{Shock}_{i,t+1} \\
&= e_1' \sum_{j=1}^{\infty} \rho^j A^j \epsilon_{i,t+1} \\
&= e_1' \rho (I_K + A(I_K - \rho A)^{-1}) \epsilon_{i,t+1}.
\end{aligned}
\tag{3.22}
$$

Specification and Estimation I include the firm log returns, clean-surplus earnings, log book-to-market ratio, and two characteristics including momentum and iVol, as well as aggregate variable CoAnomaly that I am interested in. All firm-level variables apart from the first three are normalized to unconditional mean zero and standard deviation one. I do not allow feedback from firm-level shocks to aggregate levels, so I set the lower-left part of matrix A to zero.

To separate the different predictive effects for stocks in the anomaly long legs and short legs, I also include the interaction between CoAnomaly and anomaly dummy for long legs or short legs: the long leg dummy is 1 if the stock at that period is classified as a long leg stock by aggregating signals from 34 anomalies, and 0 otherwise; the short leg dummy is 1 if the stock at that period is classified as a short leg stock by aggregating signals from 34 anomalies[7].

To avoid possible complications with the use of the log transformation, I unlever the stock by 10 percent following the standard procedure in the literature, i.e., I define the stock return as a portfolio consisting of 90 percent of the firm's common stock and 10 percent investment in Treasury Bills. Following the same concern, I

[7] I use 4 out of 34 as threshold for long leg classification, and 8 out of 34 as threshold for short leg classification, as short leg stocks are more concentrated. Though these choices are somewhat ad hoc admittedly, my results are robust of variations in these thresholds.

define the log book-to-market ratio as $bm \equiv log(\frac{0.9\times BE+0.1\times ME}{ME})$.

After demeaning the data in each cross-section, I use WLS to estimate each row in the A matrix of the VAR. The weight is the product of two measures: the inverse of the number of stocks in each cross-section, which ensures that the estimates are not driven by the recent years with a growing number of traded firms, and the inverse of the expected volatility estimated from the volatility estimation step, as the heteroskedasticity is modeled directly. I impose zero intercepts on all state variables, and my results are robust to allowing intercepts.

(Insert Table 3.5)

Table 3.5 reports the estimation of the transition matrix A in Equation 3.20 and the news functions for discount-rate and cash-flow news. The t-stats are based on the double-clustered standard errors following Petersen (2009). The results are in line with the literature: stocks with a high book-to-market ratio, high past returns, and low idiosyncratic volatility have higher expected returns. Importantly, the predictability result shows up in the firm-level analysis as the interaction between CoAnomaly and the short leg dummy carries a negative coefficient, which means when CoAnomaly is high, stocks in the anomaly short legs will experience low returns.

Portfolio Level The discount-rate news (cash-flow news) on long-short portfolio level is defined as the difference in the weighted average discount-rate news (cash-flow news) between the long leg and short leg for each anomaly.

$$\begin{aligned} DR_{p,t+1} = & \equiv \sum_{i\in \text{long leg of Anomaly } p} w_i^p DR_{i,t+1} - \sum_{i\in \text{short leg of Anomaly } p} w_i^p DR_{i,t+1} \\ CF_{p,t+1} = & \equiv \sum_{i\in \text{long leg of Anomaly } p} w_i^p CF_{i,t+1} - \sum_{i\in \text{short leg of Anomaly } p} w_i^p CF_{i,t+1}. \end{aligned} \tag{3.23}$$

Then I define the discount-rate news (cash-flow news) of the E.A.R. as the equal-weighted average of the discount-rate news (cash-flow news) from the 34 anomalies. I also define the shock to the E.A.R. as the difference between the cash-flow news and the discount-rate news:

$$\begin{aligned} DR^{Shock}_{EAR,t+1} &\equiv \sum_{34 \text{ Anomalies}} DR^{Shock}_{p,t+1} \\ CF^{Shock}_{EAR,t+1} &\equiv \sum_{34 \text{ Anomalies}} CF^{Shock}_{p,t+1} \\ EAR_{t+1} - \mathbb{E}_t[EAR_{t+1}] &\equiv CF^{Shock}_{EAR,t+1} - DR^{Shock}_{EAR,t+1}. \end{aligned} \tag{3.24}$$

3.3.4 Estimation Result

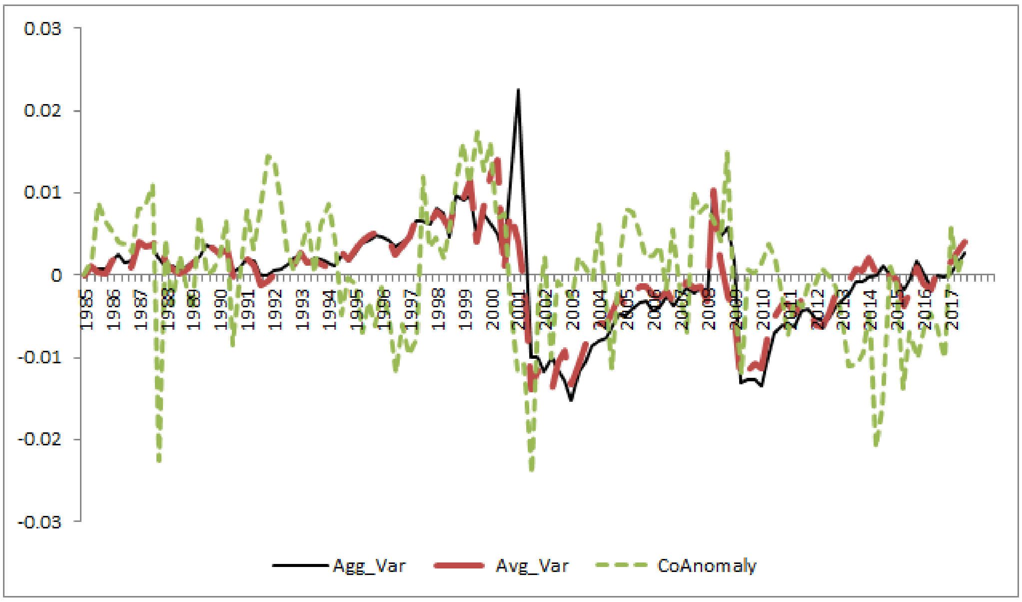
(Insert Table 3.6)

I extract the volatility news for three specifications as mentioned before: *Specification 1*, the aggregate variance of E.A.R.; *Specification 2*, the average of the 34 anomalies' variance, scaled by the unconditional mean of CoAnomaly; *Specification 3*, CoAnomaly, scaled by the unconditional average of the 34 anomalies' variance. Cash-flow news and discount-rate news of the E.A.R. are generated by summing up the cash-flow news and discount-rate news on the stock level as discussed before[8].

Panel A of Table 3.6 shows the correlations among different news. For the E.A.R., the unexpected return shock is mainly driven by cash flow news and there is a negative correlation between cash flow news and discount rate news. Both results are broadly consistent with the findings in Lochstoer and Tetlock (2017). Within the three specifications of volatility news, the average variance news is strongly correlated with the aggregate variance news, which could explain why the aggregate variance does not predict the future returns since the information it contains is contaminated by the average variance. On the other hand, CoAnomaly news is negatively correlated with two variance news. This pattern can also be confirmed in Figure 3.1, which draws the smoothed news terms for the three volatility specifications, and shows that the CoAnomaly news behaves differently from the two variance measures.

(Insert Figure 3.1)

[8] The cash flow and discount rate news reported are based on the estimation specification weighted according to aggregate volatility, and if I use other specifications, they have little effect on my results.



This figure plots the time-series of the smoothed volatility news from three different specifications: aggregate variance (solid black), average-variance-driven (thick dashed red), and CoAnomaly-driven (thin dashed green). The decay parameter is set to 0.08 per quarter, and the smoothed news series is generated as $MA_t(N) = 0.08N_t + (1 - 0.08)MA_{t-1}(N)$.

Figure 3.1: Smoothed Volatility News

Test Assets I use a total of 68 portfolios, which are the long legs and short legs of the 34 stock market anomalies, as the test assets to estimate the ICAPM model. Since I am studying the market-neutral investment universe, I removed the market component for each test asset by subtracting its in-sample market beta times the contemporaneous market return in each period.

Beta Estimation Similar to Campbell and Vuolteenaho (2004) and Campbell et al. (2017), I divide all three covariances by the variance of the E.A.R. shocks to compare to previous research:

$$\begin{aligned}
\beta_{i,CF_{EAR}} &\equiv \frac{Cov(r_{i,t}, CF^{Shock}_{EAR,t})}{Var(r_{EAR,t} - \mathbb{E}_{t-1}[r_{EAR,t}])} \\
\beta_{i,DR_{EAR}} &\equiv \frac{Cov(r_{i,t}, -DR^{Shock}_{EAR,t})}{Var(r_{EAR,t} - \mathbb{E}_{t-1}[r_{EAR,t}])} \\
\beta_{i,V_{EAR}} &\equiv \frac{Cov(r_{i,t}, N_{V_{EAR},t})}{Var(r_{EAR,t} - \mathbb{E}_{t-1}[r_{EAR,t}])}.
\end{aligned} \tag{3.25}$$

The estimated betas are reported in the appendix (see Panel B of Table 3.8). I find that the beta spreads between long legs and short legs are large and positive for both cash-flow news and discount rate news. I also find the volatility beta for the long-short anomalies is negative on average, which means these anomalies are betting on the CoAnomaly risk (as the CoAnomaly risk premium is negative as well). The empirical fact that beta spreads are large between long legs and short legs also partially alleviates the concern of the weak factor, as Bryzgalova (2014) points out.

Model Estimation

$$\overline{R}_i = g_1\widehat{\beta}_{i,CF_{EAR}} + g_2\widehat{\beta}_{i,DR_{EAR}} + g_3\widehat{\beta}_{i,V_{EAR}} + e_i \tag{3.26}$$

I estimate the above beta representation using GMM[9], with moment conditions listed in the appendix. This is equivalent to estimating the model parameter γ using the moment condition in Equation 3.13. I estimate the risk premia g as well as evaluate the pricing performance of the following asset pricing models: 1) the traditional 'CAPM' in anomaly universe by constraining the risk premia on the cash-flow news and discount-rate news to be the same; 2) the two-beta intertemporal CAPM by constraining the risk premium of discount-rate news to be the variance of E.A.R. shocks ($0.0008 \approx 0.028^2$); and 3) the three-beta intertemporal CAPM for 3 different volatility specifications by constraining the risk premium of discount-rate news to be the variance of E.A.R. and imposing the condition in Equation 3.11.

Panel B of Table 3.6 reports the estimation results. I find that the 'vanilla CAPM' works much better in the market-neutral investment universe than in the aggregate stock market, with a significantly positive price of risk which is close to the unconditional E.A.R.. This is not surprising considering the E.A.R. is the summation of these 34 anomalies as well as the sophisticated nature of the investors who are betting on them. Two-beta ICAPM only marginally improves the pricing power, as I find that the cash-flow component is dominating the premia from the anomaly returns.

Across three specifications of volatility news, I only find the volatility premium to be large and significant when CoAnomaly is used to proxy the risk, but not the aggregate variance or the average variance. This result is also consistent with the

[9]The model follows a linear factor structure (with the caveat that both raw returns and log returns show up), given the news terms.

predictive result in Guo (2018), where I show that only CoAnomaly remains a strong predictor of aggregate variance in the long-term. I also report the constrained and unrestricted risk premia in panel C of Table 3.6, which do not show many differences.

Meanwhile, I find that on average, these anomalies have negative loadings on the CoAnomaly-driven variance news (shown in the Table 3.8 in the appendix), which means these anomalies do gain higher returns by taking CoAnomaly risk premium. In other words, the CoAnomaly risk premium can be a potential explanation of why the 'CAPM-alphas' in these anomalies are not driven to zero by arbitrageurs, which is new evidence supporting limits of arbitrage. On average, around one eighth of the average anomaly return is explained by the premium on the CoAnomaly-driven variance news[10].

The implied risk aversion coefficient γ is large (around 20) compared to the one estimated in the aggregate stock market. This is not surprising since the type of investor that I study is famous for generating high returns with risk management[11]. It may also be that the estimation of the risk aversion coefficient is contaminated by the high leverage that arbitrageurs usually take[12]. Note that the overidentifying test (equivalent to testing the pricing errors equal to zero) is still rejected; however, it is much closer to no-rejection than testing ICAPM in the aggregate stock market.

3.3.5 Robustness: Alternative Specification of the Volatility Decomposition

To examine the robustness of my findings with the CoAnomaly-driven volatility specification from the simple decomposition approach, I consider another approach to extract the time-variation in aggregate volatility induced by CoAnomaly. I focus

[10] CoAnomaly risk premium in E.A.R. = (CoAnomaly-driven variance beta spread -0.04 - 0.07 = -0.11) * (Risk premium on CoAnomaly-driven variance news -0.0214) = 0.00235. The quarterly E.A.R. return premium can be calculated by either checking the 1-beta CAPM risk premia (0.0163) or by checking the Table 1.15 ($(1+0.0055)^3 - 1 = 0.0166$). Therefore, the proportion of CoAnomaly risk premium over the total E.A.R. return premium is $\frac{0.00235}{0.0166} = 0.142$.

[11] γ_t of a mean-variance agent who always holds the risky assets, as in Merton (1969), needs to satisfy $E_t[R_{t+1}] = \gamma_t \sigma_t^2$, which indicates $\gamma_t = \dfrac{E_t[R_{t+1}]}{\sigma_t^2}$.

[12] Rearrange the equation above, $1 = \dfrac{E_t[R_{t+1}]}{\sigma_t} \dfrac{1}{\sigma_t \gamma_t}$. The first term is the Sharpe ratio of the risky portfolio, which will not be affected by a single investor. If the investor is taking higher leverage, her portfolio volatility goes up and obfuscates the estimation of γ_t in the second term. Effectively, a high risk aversion coefficient might incorrectly capture the high leverage that arbitrageurs take but is not observable to us.

on the aggregate variance of E.A.R. and impose constraints in the VAR estimation to isolate the effect from CoAnomaly.

I follow the same two-stage procedure to estimate the volatility news. However, here I use only one volatility, the aggregate variance (Aggr.Var) of E.A.R., and in the VAR system, I also include CoAnomaly as a state variable. I estimate the following first-order VAR:

$$\mathbf{x}_{t+1} = \overline{\mathbf{x}} + B\mathbf{x}_t + \sigma_t \mathbf{u}_{t+1},$$

where the state variables include:

$$\mathbf{x}_{t+1} = [EVol(Aggr.Var)_{t+1} \quad VS_{t+1} \quad PE_{t+1} \quad MktRf_{t+1} \quad DEF_{t+1} \quad CoAnomaly_{t+1}]'.$$

I estimate the volatility news for three cases by imposing different restrictions on estimating the transition matrix B: (1) *no-constraint*, where I do not impose any constraint on estimating it; (2) *restricted*, where I do not allow the aggregate volatility feed back to itself and effectively I restrict the autoregressive coefficient on aggregate volatility to zero; (3) *no-feedback*, where I do not allow the aggregate volatility feedback to any state variable, which means I restrict the first column of B to be zero.

(Insert Table 3.7)

In the Panel A of Table 3.7, I report the news functions $e_1' \rho B(I_K - \rho B)^{-1}$ that map shocks to different state variables to the volatility news for above three cases. It shows that, in the *no-constraint* case, the volatility news is dominated by the negative shock to the expected aggregate variance due to the strong mean-reverting feature of aggregate variance; however, if I shut down the feedback loop within the aggregate variance itself, as shown in the *restricted* case, the effect becomes smaller, but still gets transitioned from other state variables; in the *no-feedback* case, I do not allow the aggregate variance feedback to any state variable, which yields a zero loading on the aggregate variance shock. In the last case, most of the change to future volatility comes from CoAnomaly shock, and this effectively carries the same logic when focusing on the only the CoAnomaly-driven volatility (*Specification 3* in the main result).

Panel B of Table 3.7 confirms the consistency between the two approaches: the time-series correlation between estimated volatility news of the CoAnomaly-driven specification and of the no-feedback case is close to 1. In nontabulated results, I also find the pricing effects are similar if I use the estimated volatility news from the *no-feedback* case. However, if the feedback from aggregate variance is allowed (fully or partially), the estimated volatility news behaves quite differently, even when I include CoAnomaly as a state variable. This is consistent with Pollet and Wilson (2010)'s argument that information about the risk from the aggregate variance is contaminated.

3.3.6 Interpretation of the Market-Neutral Investing ICAPM

In a nutshell, the special investor that I am studying 1) cares about the cash-flow news and CoAnomaly-driven volatility news and 2) does not pay a premium for assets that comove with the discount-rate news. These two facts also justify why CoAnomaly carries a negative price of risk given it positively predicts future anomaly returns: for the market-neutral investor, it is the volatility of the return (second moment of investment opportunity) that matters not simply the level (first moment of investment opportunity).

However, extrapolation of these results must be conducted with care. Campbell et al. (2017) interpret their results as a microeconomic description about the intertemporal considerations of conservative long-term equity investors (including institutions such as pension funds and endowments) but not as a description of a representative-agent model of general equilibrium in financial markets. This is because their model and calibration depend on the assumption that the long-term equity investor who is content to hold the aggregate stock market and her average consumption-wealth ratio is relatively stable. Unlike the aggregate stock market, my ICAPM setting in anomaly investing relies on a *stronger* assumption that assumes the arbitrageur studied here always holds the equal-weighted anomaly portfolio, with another implication following this assumption being that this arbitrageur does not manage her exposure to these portfolios according to stochastic volatility. This is a compromise given the fact that there is no ‘value-weighted’ portfolio of different anomalies, and researchers are agnostic about how much capital is allocated to each strategy even in the aggregate for all arbitrageurs as a whole.

Consequently, the result I show can only be interpreted as follows: if there is an arbitrageur chasing market neutrality by always holding these stock market anomalies with equal weight, how her SDF changes given the behavior of these anomaly assets, to what extent can the intertemporal considerations in the time-varying investment opportunity (level and volatility) explain the cross-section of anomaly returns, and most importantly, which part of volatility does she actually care about. Therefore, the implied risk-aversion coefficient from the estimated risk premia cannot be interpreted as the risk-aversion coefficient of a *representative* arbitrageur in a general equilibrium sense, given the limits to interpreting the risk-aversion coefficient of a representative agent in Campbell et al. (2017).

3.4 Conclusion

I first show that the naïve 1/N equal-weighted investing in 34 anomalies do serve as a robust benchmark. Five mean-variance-efficient-optimized portfolios, including three Sharpe-Ratio-based with weight constraints and two Dimensional-Reduction-based, generate higher Sharpe-ratio in the full sample but they also come with the cost of higher turnover ratio. Moreover, in the recent 18 years, this outperformance is not robust comparing with the equal-weighted anomaly return (EAR) portfolio. I also find that by taking an unconditional historical mean will provide better estimates of future anomaly returns than using a rolling-window approach. That being said, the starting point of these results is based on the anomalies already constructed, so potential gains can still be exploited from the bottom-up portfolio construction on stock level (see Brandt et al. (2009) for an example).

I next use the EAR portfolio as a conservative benchmark to study the intertemporal consideration of anomaly investors in an intertemporal CAPM framework with stochastic volatility. The model estimation shows the cash-flow news carries a large and positive risk premium and the premium on discount-rate news is negligible. As for the volatility news, I find evidence that only the CoAnomaly-induced part in the aggregate variance of the equal-weighted anomaly portfolio carries a negative and significant price of risk.

These results echo the findings in Guo (2018) and suggest that the comovement among these stock market anomalies is an important determinant of the marginal utility of arbitrageurs. Though the drivers of the time-varying comovement are not explored here, future research may study the links among the comovement, fundamental risks and the trading behavior of arbitrageurs.

3.5 Appendix: Estimation Procedure in the ICAPM

Here, I will discuss several issues in detail about the estimation of the intertemporal CAPM.

Average Variance and Correlation Decomposition I make two sets of strong assumptions to separate the effects from average variance and from the correlation. This is a compromise due to the difficulty in modeling the time-variation of both average variance and correlation at the same time in the ICAPM. So I choose to take a short-cut to study the relative importance of each component in the total volatility. As shown in Guo (2018), the changes in CoAnomaly and changes in the average variance of each anomaly can explain more than 80% of the changes in E.A.R. variance. So this simplification covers a large portion of the time-variation in volatility.

Specification 2 — I assume that the correlation structure between anomalies is constant through time and across asset pairs, and all assets have the same conditional variance σ^2_{t+j+1}. The second assumption can accommodate the case that assets have different conditional variance by splitting assets of high variance into smaller pieces.

$$
\begin{aligned}
\text{Avg.CoVariance}_{t+j+1} \quad & = \frac{\sum_{l\neq m} Cov_{t+j}[r_{l,t+j+1}, r_{m,t+j+1}]}{N(N-1)} \\
& = \frac{\sum_{l\neq m} \sigma_{l,t+j+1}\sigma_{m,t+j+1}\rho_{lm}}{N(N-1)} \\
\text{(Equal Conditional Variance)} \quad & = \frac{\sum_{l\neq m} \sigma_{t+j+1}\sigma_{t+j+1}\rho_{lm}}{N(N-1)} \\
& = \sigma^2_{t+j+1} \times \frac{\sum_{l\neq m} \rho_{lm}}{N(N-1)} \\
& = \sigma^2_{t+j+1} \times \overline{\text{Correlation}},
\end{aligned} \tag{3.27}
$$

where $\overline{\text{Correlation}}$ is the average pairwise correlation between assets. In the empirical part, I use the unconditional sample mean of CoAnomaly to proxy this.

Specification 3 — I assume the variance of each anomaly is identical and constant through time and across assets, and use the time-varying average correlation

(*CoAnomaly*) as the stochastic volatility measure.

$$\begin{aligned}
\text{Avg.CoVariance}_{t+j+1} \quad & = \frac{\sum_{l \neq m} Cov_{t+j}[r_{l,t+j+1}, r_{m,t+j+1}]}{N(N-1)} \\
& = \frac{\sum_{l \neq m} \sigma_{l,t+j+1}\sigma_{m,t+j+1}\rho_{lm}}{N(N-1)} \\
\text{(Constant Conditional Variance)} \quad & = \frac{\sum_{l \neq m} \sigma^2 \rho_{lm,t+j+1}}{N(N-1)} \\
& = \sigma^2 \times \frac{\sum_{l \neq m} \rho_{lm,t+j+1}}{N(N-1)} \\
& = \sigma^2 \times \text{Avg-Correlation}_{t+j+1},
\end{aligned} \tag{3.28}$$

where Avg-Correlation$_{t+j+1}$ is the average pairwise correlation between assets at each point of time, and in my setting, it is exactly the CoAnomaly measure.

Estimation of Model Parameters I do not follow Campbell et al. (2017) to estimate all three news terms in a single aggregate VAR. As argued before, unlike the aggregate stock market portfolio, the aggregate VAR approach on anomaly portfolios are facing much more serious problems than the market portfolio due to the changing weights and rebalancing of the constituents. However, if I conduct the bottom-up approach by extracting the discount-rate news and cash flow news on stock level and then summing them up to portfolio level, I face two challenges which do no show up in their setting: the estimation of model parameter ω and the log return decomposition.

Parameter omega ω in the model have to satisfy the following condition:

$$\omega\sigma_t^2 = (1-\gamma)^2 Var_t[N_{CF,t+1}] + \omega(1-\gamma)Cov_t[N_{CF,t+1}, N_{V,t+1}] + \frac{1}{4}\omega^2 Var_t[N_{V,t+1}]. \tag{3.29}$$

As shown in the online appendix of Campbell et al. (2017), the above equation can be directly mapped to the coefficients estimated the aggregate VAR. Since I do not estimate the aggregate VAR, I cannot follow this approach. Instead, I move the conditional variance term to the right-hand side and then use the sample moments to the scaled variances and covariances of different news to estimate ω.

$$\begin{aligned}\omega &= (1-\gamma)^2\frac{Var_t[N_{CF,t+1}]}{\sigma_t^2} + \omega(1-\gamma)\frac{Cov_t[N_{CF,t+1}, N_{V,t+1}]}{\sigma_t^2} + \frac{1}{4}\omega^2\frac{Var_t[N_{V,t+1}]}{\sigma_t^2} \\ &= (1-\gamma)^2 Var_t[\frac{N_{CF,t+1}}{\sigma_t}] + \omega(1-\gamma)Cov_t[\frac{N_{CF,t+1}}{\sigma_t}, \frac{N_{V,t+1}}{\sigma_t}] + \frac{1}{4}\omega^2 Var_t[\frac{N_{V,t+1}}{\sigma_t}].\end{aligned} \tag{3.30}$$

So the estimated $\hat{\omega}$ and estimated $\hat{\gamma}$ in my setting need to satisfy the sample moments of the above equation:

$$\hat{\omega} = (1-\hat{\gamma})^2\widehat{Var}[\frac{N_{CF,t+1}}{\sigma_t}] + \hat{\omega}(1-\hat{\gamma})\widehat{Cov}[\frac{N_{CF,t+1}}{\sigma_t}, \frac{N_{V,t+1}}{\sigma_t}] + \frac{1}{4}\hat{\omega}^2\widehat{Var}[\frac{N_{V,t+1}}{\sigma_t}], \tag{3.31}$$

which will have two solutions. Campbell et al. (2017) further show that $\hat{\omega}$ is the one with the negative sign on the radical[13]

$$\begin{aligned}&\hat{\omega} = \\ &\frac{[1-(1-\hat{\gamma})\widehat{Cov}[\frac{N_{CF,t+1}}{\sigma_t}, \frac{N_{V,t+1}}{\sigma_t}]]}{\frac{1}{2}\widehat{Var}[\frac{N_{V,t+1}}{\sigma_t}]} \\ &- \frac{\sqrt{[1-(1-\hat{\gamma})\widehat{Cov}[\frac{N_{CF,t+1}}{\sigma_t}, \frac{N_{V,t+1}}{\sigma_t}]]^2 - (1-\hat{\gamma})^2\widehat{Var}[\frac{N_{V,t+1}}{\sigma_t}]\widehat{Var}[\frac{N_{CF,t+1}}{\sigma_t}]}}{\frac{1}{2}\widehat{Var}[\frac{N_{V,t+1}}{\sigma_t}]}.\end{aligned}$$

Note that this equation requires an existence condition on γ, which can be simplified as:

$$1 - \frac{1}{(\rho_n+1)\sigma_{cf}\sigma_v} \le \gamma \le 1 - \frac{1}{(\rho_n-1)\sigma_{cf}\sigma_v} \tag{3.32}$$

where ρ_n is the correlation of the cash-flow and volatility news, $\sigma_{cf} = Var_t[\frac{N_{CF,t+1}}{\sigma_t}]$ and $\sigma_v = Var_t[\frac{N_{V,t+1}}{\sigma_t}]$. I use the sample moments to get the bound[14] for γ, which

[13] Inserting the sample moments for CoAnomaly-driven variance news: $\hat{\omega} = \frac{[1-(1-19.625)*0.0922 - \sqrt{[1-(1-19.625)*0.0922]^2 - (1-19.625)^2*0.011881*0.394384}}{\frac{1}{2}*0.011881} = 53.47$; however, for aggregate variance news and average variance news, the sample moment conditions are binding, so the GMM estimates are fitting the cash-flow news, which is the most prominent factor in the stock market anomaly space.

[14] Inserting the sample moments: $\gamma \le 1 - \frac{1}{(0.316-1)\times 0.628\times 0.109} = 22.35786$

requires it no larger than 22.36.

GMM Estimation Here are the moment conditions in estimating the ICAPM conditioning on estimated news terms:

$$g_{(4N+2)\times 1}(b) = \begin{bmatrix} \mathbb{E}\Big[\beta_{i,CF_{EAR}} - \dfrac{Cov(r_{i,t}, CF^{Shock}_{EAR,t})}{Var_{EAR}}\Big] \\ \mathbb{E}\Big[\beta_{i,DR_{EAR}} - \dfrac{Cov(r_{i,t}, -DR^{Shock}_{EAR,t})}{Var_{EAR}}\Big] \\ \mathbb{E}\Big[\beta_{i,V_{EAR}} - \dfrac{Cov(r_{i,t}, N_{V_{EAR},t})}{Var_{EAR}}\Big] \\ \mathbb{E}\Big[R_i - g_1\beta_{i,CF_{EAR}} - g_2\beta_{i,DR_{EAR}} - g_3\beta_{i,V_{EAR}}\Big] \\ \mathbb{E}\Big[g_2 - Var_{EAR}\Big] \\ \mathbb{E}\Big[\kappa(g_1, g_3)\Big] \end{bmatrix},$$

where the first three are defining three betas for N assets, next one is the pricing condition, and the last two are conditions from the model. $\kappa(g_1, g_3) = 0$ is equivalent to the equation between γ and ω since effectively $g_1 = \gamma V_{EAR}$ and $g_3 = -\frac{1}{2}\omega V_{EAR}$, analogous to Equation 3.12.

I then use a selection matrix $a_{(3N+5)\times(4N+2)}$ to define which linear combination of $g_{(4N+2)\times 1}(b)$ will be set to zero:

$$a_{(3N+5)\times(4N+2)} = \begin{bmatrix} I_{3N\times 3N} & 0_{3N\times N} & 0_{3N\times 2} \\ 0_{3\times 3N} & \beta'_{N\times 3} & 0_{3\times 2} \\ 0_{2\times 3N} & 0_{2\times N} & C_{2\times 2} \end{bmatrix},$$

where $\beta_{N\times 3} = [\beta_{CF,N\times 1} \quad \beta_{DR,N\times 1} \quad \beta_{V,N\times 1}]$ is the three-beta matrix, and $C_{2\times 2}$ is a choice matrix by selecting conditions when estimating the model. If I want to estimate the ICAPM, I choose $C_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, which effectively impose both conditions from the ICAPM; if I want to estimate the constrained model, I choose $C_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, which only imposes the premium on the discount rate news equals the variance of E.A.R.; if I want to estimate the unconstrained model, I choose $C_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, which imposes no constraints on the risk premia. In the first two cases, $\beta'_{N\times 3}$ in $a_{(3N+5)\times(4N+2)}$ will be modified accordingly to accommodate the change in numbers of moment conditions, for example, if I chose $C_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, the second row of $\beta'_{N\times 3}$ will be set to zero because I estimate g_2 from $\mathbb{E}\Big[g_2 - Var_{EAR}\Big]$ instead of $\mathbb{E}\Big[R_i - g_1\beta_{i,CF_{EAR}} - g_2\beta_{i,DR_{EAR}} - g_3\beta_{i,V_{EAR}}\Big]$.

The parameter estimates will be $\hat{b}$ such that

$$a_{(3N+5)\times(4N+2)} \times g_{(4N+2)\times 1}(\hat{b}) = 0_{(3N+5)\times 1}.$$

Note that the moments $g_{(4N+2)\times 1}(b)$ and the choice matrix C are just for the simplicity of illustrative purposes, and in the estimation, I delete irrelevant moments from the system, so the J-test is equivalent to testing the pricing errors to be zero for all test assets.

Log Returns I report the variance-covariance matrix for the scaled news in Table 3.8. I find that the scaled variance of E.A.R. shock happened to be close to 1, which is the implication of the model structure. However, I do not impose this restriction as I estimate the volatility news and cash-flow and discount-rate news in two VAR systems. So the result suggests that separating the news estimation does not generate modeling inconsistency.

(Insert Table 3.8)

The portfolio cash flow and discount rate shocks are not simple weighted averages of firms' cash flow and discount rate shocks because the firm-level variance decomposition applies to log returns[15]. To make sure my simplification is not driving my results, I also sum up the level shocks using a second-order Taylor expansion following Lochstoer and Tetlock (2017) and then take log again on the portfolio level, which does not change my findings in a qualitative way.

[15]See a simple case: $log(\frac{1}{2}R_1 + \frac{1}{2}R_1) \neq \frac{1}{2}log(R_1) + \frac{1}{2}log(R_2)$

Table 3.1: MVE Weights

This table reports the time-series average of the MVE-optimized portfolio weights on all 34 anomalies. For the maximizing Sharpe ratio problem, I consider three different box constraints for each weight: relative unconstrained $w_i \in [-1, 1]$ (SR_[-1,1]), non-negative weights $w_i \in [0, 1]$ (SR_[0,1]), and capped weights $w_i \in [0.01, 0.1]$ (SR_[0.01,0.1]). As for the dimension reduction with penalized regression of raw anomaly portfolios (DR_raw) or principal components (DR_pca), I do not tune the hyperparameters γ_1 and γ_2. I choose the $\kappa = 3$ as the prior belief about the maximum Sharpe ratio in the economy and a weight parameter to sparsity penalty of 0.05. All MVE-optimized portfolio weights are based on the look-back choice (2), using the past 10 years to estimate the covariance and using all past periods to estimate the mean.

MVE_methods	Stats	size	profit	value	acc	netissue_a	netissue_m	atgrowth	invest	piotroski	ato	gm	ohlson	roe	failprob	idiovol	mom12m	peadsue	peadcar3
SR_[-1,1]	mean	7.8%	4.7%	6.1%	3.6%	8.1%	0.4%	-2.1%	3.9%	4.0%	-0.6%	1.4%	3.9%	-3.9%	-6.6%	1.3%	4.9%	5.7%	10.7%
	std	3.0%	2.4%	1.4%	1.9%	5.1%	7.0%	1.4%	1.9%	1.4%	1.5%	2.4%	3.0%	2.5%	1.8%	2.3%	1.7%	1.7%	1.8%
	min	1.4%	-0.4%	3.0%	0.2%	1.5%	-11.6%	-6.7%	-0.4%	1.8%	-3.2%	-3.1%	-3.1%	-9.4%	-10.1%	-3.7%	1.5%	1.5%	6.9%
	max	11.8%	9.6%	9.6%	8.2%	18.5%	11.2%	0.9%	7.8%	6.7%	3.0%	6.3%	8.1%	0.1%	-3.6%	5.8%	11.1%	8.8%	14.3%
SR_[0,1]	mean	8.1%	3.2%	7.2%	3.2%	4.0%	2.5%	0.1%	3.3%	2.7%	0.4%	1.2%	2.9%	0.0%	0.0%	0.7%	0.9%	3.3%	9.4%
	std	2.2%	2.2%	1.3%	1.9%	2.4%	2.6%	0.1%	2.5%	1.1%	0.7%	1.4%	2.0%	0.0%	0.0%	0.9%	1.3%	1.8%	1.6%
	min	3.4%	0.0%	4.3%	0.0%	0.1%	0.0%	0.0%	0.0%	0.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	6.1%
	max	11.8%	6.3%	10.4%	7.4%	9.1%	7.7%	0.8%	7.9%	4.7%	2.7%	5.6%	6.7%	0.0%	0.0%	2.8%	6.4%	7.4%	12.9%
SR_[0.01,0.1]	mean	7.4%	2.6%	6.8%	3.0%	2.4%	2.6%	1.0%	3.0%	1.9%	1.1%	1.2%	2.2%	1.0%	1.0%	1.1%	1.2%	2.2%	8.5%
	std	1.8%	1.3%	1.9%	1.6%	1.4%	1.8%	0.0%	2.5%	0.9%	0.3%	0.4%	1.2%	0.0%	0.0%	0.2%	0.8%	1.6%	1.3%
	min	4.0%	1.0%	3.4%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	5.7%
	max	10.0%	5.2%	10.0%	7.1%	5.5%	6.5%	1.0%	8.1%	3.7%	2.4%	3.5%	5.1%	1.0%	1.0%	2.0%	5.7%	6.1%	10.0%
DR_raw	mean	2.1%	0.2%	3.3%	1.5%	3.2%	2.4%	1.8%	2.6%	1.7%	0.5%	-0.4%	0.5%	1.2%	-0.4%	1.3%	5.2%	3.4%	4.3%
	std	0.3%	0.2%	0.5%	0.4%	0.6%	0.3%	0.3%	0.3%	0.7%	0.3%	0.5%	0.4%	0.4%	0.7%	0.4%	0.6%	0.6%	0.3%
	min	1.4%	-0.2%	2.4%	0.7%	2.1%	1.6%	1.0%	1.7%	0.0%	0.0%	-1.7%	0.0%	0.1%	-2.5%	0.2%	4.1%	2.6%	3.6%
	max	3.0%	0.7%	4.5%	2.7%	4.1%	2.8%	2.8%	3.1%	2.9%	1.2%	0.0%	1.4%	1.9%	0.0%	2.2%	6.7%	5.0%	5.1%
DR_pca	mean	2.0%	0.5%	3.0%	1.7%	2.7%	2.6%	2.3%	2.5%	1.7%	1.3%	-0.7%	0.8%	1.6%	-0.2%	1.5%	4.8%	3.2%	3.5%
	std	0.4%	0.6%	0.6%	0.4%	0.3%	0.6%	0.5%	0.4%	0.6%	0.6%	0.5%	0.3%	0.6%	0.9%	0.5%	0.5%	0.6%	0.6%
	min	1.1%	-0.6%	1.7%	0.6%	1.9%	1.5%	1.4%	1.7%	-0.1%	0.2%	-1.7%	0.1%	0.3%	-2.5%	0.2%	3.6%	1.4%	2.1%
	max	2.7%	2.0%	4.5%	2.6%	3.4%	4.0%	3.6%	3.6%	3.6%	2.7%	0.3%	1.8%	2.9%	1.4%	2.4%	6.0%	4.8%	4.9%

MVE_methods	Stats	rev60m	rome	roa	beta	indmom1m	relrev1m	rev1m	seasonal	relrev1mlow	valprof	valmomprof	valmom	hfcombo1	hfcombo2	ceissue	netoa	Average
SR_[-1,1]	mean	1.9%	9.3%	-1.1%	2.0%	0.3%	3.1%	1.0%	4.7%	8.8%	0.8%	1.7%	-1.3%	10.9%	3.5%	1.3%	-0.3%	
	std	1.7%	2.1%	1.9%	2.2%	0.7%	3.4%	2.1%	1.9%	2.5%	1.8%	3.0%	2.0%	2.9%	1.3%	3.3%	2.9%	2.5%
	min	-1.3%	5.8%	-6.9%	-2.0%	-1.5%	-2.6%	-3.3%	0.6%	4.7%	-3.4%	-2.5%	-4.9%	6.6%	1.1%	-5.1%	-7.0%	-1.9%
	max	6.2%	14.4%	3.5%	6.2%	2.7%	8.6%	4.4%	8.0%	13.9%	4.6%	6.6%	3.0%	17.2%	6.3%	7.8%	5.7%	7.7%
SR_[0,1]	mean	1.5%	6.0%	0.1%	1.3%	0.2%	4.0%	2.7%	5.1%	8.6%	0.8%	1.0%	0.2%	10.5%	3.8%	0.3%	0.6%	
	std	1.2%	1.5%	0.4%	1.7%	0.4%	2.8%	1.7%	2.5%	2.5%	1.0%	1.1%	0.4%	2.3%	1.8%	0.6%	0.9%	1.4%
	min	0.0%	2.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.9%	4.7%	0.0%	0.0%	0.0%	6.6%	1.0%	0.0%	0.0%	0.8%
	max	4.5%	8.7%	2.5%	5.4%	1.9%	7.6%	5.9%	9.4%	13.6%	3.2%	3.4%	1.8%	15.8%	7.2%	2.4%	2.9%	6.1%
SR_[0.01,0.1]	mean	1.4%	4.3%	1.0%	1.6%	1.0%	5.2%	3.1%	4.9%	8.3%	1.3%	1.1%	1.0%	9.1%	4.4%	1.0%	1.0%	
	std	0.6%	2.2%	0.0%	1.1%	0.1%	2.3%	1.4%	2.9%	1.8%	0.5%	0.2%	0.0%	1.1%	2.2%	0.1%	0.1%	1.1%
	min	1.0%	1.1%	1.0%	1.0%	1.0%	1.0%	1.0%	1.0%	4.7%	1.0%	1.0%	1.0%	5.9%	1.0%	1.0%	1.0%	1.6%
	max	3.5%	7.9%	1.0%	4.9%	2.2%	8.7%	6.6%	9.9%	10.0%	2.8%	2.6%	1.0%	10.0%	8.2%	1.8%	1.9%	5.2%
DR_raw	mean	1.7%	5.8%	1.5%	2.5%	3.5%	5.6%	2.1%	3.5%	7.8%	3.5%	4.9%	4.5%	9.2%	7.8%	2.2%	-0.3%	
	std	0.5%	0.5%	0.4%	0.7%	0.6%	0.8%	0.7%	1.1%	0.7%	0.6%	0.3%	0.5%	0.7%	0.3%	0.2%	0.4%	0.4%
	min	0.8%	4.4%	0.4%	0.6%	2.1%	4.2%	0.9%	1.0%	6.3%	2.4%	4.3%	3.7%	8.4%	7.1%	1.2%	-1.4%	0.9%
	max	3.5%	6.8%	2.2%	3.7%	4.8%	7.1%	3.5%	4.9%	9.1%	4.9%	5.5%	6.0%	10.9%	8.6%	2.6%	0.0%	2.8%
DR_pca	mean	1.9%	5.0%	1.9%	2.3%	4.0%	5.7%	2.5%	3.4%	7.1%	3.5%	4.7%	4.7%	8.7%	7.7%	2.4%	-0.3%	
	std	0.6%	0.7%	0.6%	0.7%	0.5%	0.8%	0.8%	1.0%	0.9%	0.6%	0.5%	0.5%	0.8%	0.4%	0.4%	0.6%	0.5%
	min	0.5%	3.6%	0.6%	0.3%	2.5%	4.3%	1.0%	1.0%	5.2%	2.3%	3.5%	3.5%	7.0%	6.5%	1.2%	-1.4%	0.7%
	max	3.7%	6.3%	3.5%	3.7%	5.1%	7.2%	3.6%	4.7%	8.8%	5.0%	6.3%	6.0%	10.1%	8.4%	3.1%	1.3%	3.2%

Table 3.2: Out-of-Sample Performance: Returns, Sharpe Ratio and Turnover

This table reports the out-of-sample annualized return and Sharpe ratio of MVE-optimized portfolios as well as their monthly turnover. I normalize the weights so that each portfolio has the summation of the weights on 34 anomalies up to 1. I consider three look-back choices for estimating the weights: (1) using the past 10 years to estimate the mean and covariance of anomaly portfolios, (2) using the past 10 years to estimate the covariance and using all past periods to estimate the mean, and (3) using all past periods to estimate the mean and covariance. Monthly turnover on the anomaly level is reported in the last column.

Panel A: Out-of-Sample Annualized Return and Sharpe Ratio - 1983-2017									
Look-back		(1)			(2)			(3)	
MVE_methods	Return	Sharpe.R	Turnover (M)	Return	Sharpe.R	Turnover (M)	Return	Sharpe.R	Turnover (M)
SR_[-1,1]	7.06%	0.94	36.2%	6.08%	1.71	7.5%	6.43%	1.78	5.0%
SR_[0,1]	5.79%	1.50	7.4%	5.61%	1.65	4.1%	6.03%	1.76	2.9%
SR_[0.01,0.1]	5.68%	1.49	6.0%	5.43%	1.57	3.5%	5.86%	1.67	2.7%
DR_raw	6.36%	0.86	8.4%	6.44%	1.23	2.4%	6.50%	1.42	2.4%
DR_pca	6.27%	0.85	9.1%	6.40%	1.24	3.7%	6.47%	1.42	2.6%
EAR	5.19%	0.99	3.0%						

Panel B: Out-of-Sample Annualized Return and Sharpe Ratio - 2000-2017									
Look-back		(1)			(2)			(3)	
MVE_methods	Return	Sharpe.R	Turnover (M)	Return	Sharpe.R	Turnover (M)	Return	Sharpe.R	Turnover (M)
SR_[-1,1]	6.46%	0.64	60.0%	3.83%	0.93	7.7%	3.83%	0.91	5.1%
SR_[0,1]	4.17%	0.88	8.4%	3.35%	0.84	4.1%	3.61%	0.90	2.9%
SR_[0.01,0.1]	3.96%	0.84	6.8%	3.20%	0.77	3.7%	3.46%	0.82	2.8%
DR_raw	3.34%	0.34	12.0%	3.40%	0.52	2.4%	3.49%	0.63	2.7%
DR_pca	3.24%	0.33	13.1%	3.39%	0.52	4.4%	3.48%	0.63	2.8%
EAR	3.68%	0.54	3.3%						

Table 3.3: Comparing MVE-Optimized Portfolios with EAR

$$R_t = \alpha + \beta_m MktRf_t + \beta_e EAR_t + \epsilon_t,$$

This table reports the regression coefficients $(\alpha, \beta_m, \beta_e)$ of regressing MVE-optimized portfolios on market excess returns and EAR. The MVE-optimization is based on look-back choice (2), using the past 10 years to estimate the covariance and using all past periods to estimate the mean. The regression is based on daily returns and the intercepts are reported as annualized returns. The standard errors with Newey and West (1987) correction for 10 lags (two weeks) are reported in the parentheses.

	Full Sample: 1983-2017			Second Half-Sample: 2000-2017		
	Intercept	MktRf	EAR	Intercept	MktRf	EAR
SR_[-1,1]	3.75%	0.018	0.418	1.86%	0.055	0.444
	(0.55%)	(0.007)	(0.027)	(0.81%)	(0.007)	(0.032)
SR_[0,1]	3.12%	0.018	0.451	1.29%	0.053	0.478
	(0.48%)	(0.006)	(0.030)	(0.68%)	(0.006)	(0.035)
SR_[0.01,0.1]	2.67%	0.016	0.505	0.96%	0.050	0.531
	(0.45%)	(0.006)	(0.029)	(0.66%)	(0.006)	(0.034)
DR_raw	1.49%	0.019	0.920	-0.33%	0.042	0.949
	(0.43%)	(0.005)	(0.021)	(0.60%)	(0.006)	(0.025)
DR_pca	1.49%	0.016	0.921	-0.30%	0.037	0.945
	(0.40%)	(0.005)	(0.020)	(0.55%)	(0.005)	(0.023)

Table 3.4: VAR Estimates for Volatility

This table reports the two-stage VAR results for volatility. Different specifications start with different realized variance measures: *Specification 1* uses the time-varying aggregate variance of E.A.R.; *Specification 2* uses the time-varying average variance; and *Specification 3* uses the time-varying CoAnomaly. All raw measures of volatility are scaled with the same unconditional mean before the first stage. Under each specification, the first row reports the regression results of the first stage, and the second panel reports the VAR coefficients at the second stage, and the last row reports the news function $e_1'\rho B(I_K - \rho B)^{-1}$, which maps the shocks to different state variables to the volatility news. VS is the small-stock value spread; PE ratio is the cyclically adjusted price-to-earnings ratio downloaded from Shiller's website; MktRf is the market excess return from French's website; DEF is the default spread, defined as the difference between the log yield on Moody's BAA and AAA bonds and downloaded from the Federal Reserve Bank of St. Louis. T-stats, calculated with bootstrap standard errors, are reported in parentheses.

	Specification 1: Aggregate Variance						
	Forecasting Quarterly Realized Variance of E.A.R.						
	Constant	R.Variance t	VS t	PE t	MktRf t	DEF t	Adj. R-Squared
R.Variance t+1	0.09	0.85	-0.01	0.07	-0.07	0.01	43.7%
	(2.31)	(10.25)	(-0.23)	(2.13)	(-1.91)	(0.23)	
	VAR Estimates						
	Constant	E.Variance t	VS t	PE t	MktRf t	DEF t	Adj. R-Squared
E.Variance t+1	0.10	0.83	-0.01	0.07	-0.01	0.00	47.7%
	(2.42)	(10.22)	(-0.22)	(2.22)	(-0.33)	(0.10)	
VS t+1	-0.10	0.17	0.87	0.04	0.03	-0.01	82.3%
	(-2.15)	(1.88)	(24.66)	(1.08)	(0.96)	(-0.25)	
PE t+1	0.03	-0.04	0.00	1.01	0.00	0.02	97.7%
	(1.99)	(-1.21)	(-0.29)	(78.70)	(0.34)	(1.16)	
MktRf t+1	0.09	-0.14	-0.08	0.05	0.00	0.23	0.3%
	(0.84)	(-0.65)	(-0.87)	(0.53)	(-0.01)	(2.17)	
DEF t+1	-0.01	0.03	0.01	0.03	-0.13	0.88	75.1%
	(-0.28)	(0.37)	(0.22)	(0.82)	(-3.92)	(21.20)	
	News Function						
Shocks		E.Variance t	VS t	PE t	MktRf t	DEF t	
		-2.29	-0.42	33.95	-0.37	3.92	

	Specification 2: Variance driven by Average Variance						
	Forecasting Quarterly Realized Variance of E.A.R.						
	Constant	R.Variance t	VS t	PE t	MktRf t	DEF t	Adj. R-Squared
R.Variance t+1	0.11	0.81	0.00	0.09	-0.02	-0.01	41.1%
	(2.91)	(8.88)	(-0.07)	(2.30)	(-0.44)	(-0.29)	
	VAR Estimates						
	Constant	E.Variance t	VS t	PE t	MktRf t	DEF t	Adj. R-Squared
E.Variance t+1	0.14	0.78	0.00	0.09	-0.01	-0.01	47.0%
	(3.10)	(9.11)	(-0.15)	(2.61)	(-0.15)	(-0.32)	
VS t+1	-0.10	0.17	0.88	0.02	0.03	0.02	84.0%
	(-2.15)	(1.89)	(25.48)	(0.45)	(0.77)	(0.44)	
PE t+1	0.03	-0.03	0.00	1.01	0.01	0.01	97.6%
	(1.63)	(-0.91)	(-0.01)	(72.11)	(0.64)	(0.84)	
MktRf t+1	0.04	-0.05	-0.07	0.02	0.02	0.18	-0.6%
	(0.35)	(-0.24)	(-0.76)	(0.26)	(0.21)	(1.80)	
DEF t+1	0.00	0.01	0.02	0.04	-0.13	0.87	75.6%
	(0.02)	(0.11)	(0.52)	(1.08)	(-3.78)	(21.34)	
	News Function						
Shocks		E.Variance t	VS t	PE t	MktRf t	DEF t	
		-1.61	0.57	42.77	-0.14	3.80	

	Specification 3: Variance driven by CoAnomaly						
	Forecasting Quarterly Realized Variance of E.A.R.						
	Constant	R.Variance t	VS t	PE t	MktRf t	DEF t	Adj. R-Squared
R.Variance t+1	0.29	0.49	-0.03	0.09	-0.08	0.07	25.9%
	(7.13)	(6.30)	(-0.78)	(4.09)	(-2.95)	(1.79)	
	VAR Estimates						
	Constant	E.Variance t	VS t	PE t	MktRf t	DEF t	Adj. R-Squared
E.Variance t+1	0.36	0.38	0.00	0.06	-0.01	0.07	39.1%
	(8.10)	(4.85)	(-0.20)	(3.43)	(-0.69)	(3.17)	
VS t+1	-0.14	0.24	0.88	0.05	0.03	-0.01	80.8%
	(-1.33)	(1.32)	(24.14)	(1.28)	(0.89)	(-0.25)	
PE t+1	0.00	0.02	-0.02	1.00	0.02	0.01	97.3%
	(-0.09)	(0.32)	(-1.34)	(67.22)	(1.17)	(0.69)	
MktRf t+1	0.01	-0.01	-0.11	-0.01	0.06	0.19	1.0%
	(0.05)	(-0.01)	(-1.24)	(-0.15)	(0.70)	(1.71)	
DEF t+1	-0.12	0.21	0.01	0.00	-0.13	0.85	73.9%
	(-1.10)	(1.13)	(0.27)	(0.01)	(-3.47)	(17.42)	
	News Function						
Shocks		E.Variance t	VS t	PE t	MktRf t	DEF t	
		0.93	-0.69	5.84	-0.10	1.13	

Table 3.5: Firm-level VAR Estimation

This table reports the WLS parameter estimates of the firm-level VAR model by regressing the state variables on a one-quarter lagged value of these state variables. Log(Ret), log(BM), and log(ROE) satisfy the log clean-surplus condition as $lnROE^{CS}_{i,t+1} \equiv r_{i,t+1} + \rho bm_{i,t+1} - bm_{i,t}$. Momentum is the cumulative return in the last 12 months ignoring the most recent month, and iVol is the idiosyncratic volatility measured by daily volatility orthogonal to the market component in a given quarter. These two measures are normalized to mean zero and standard deviation one. CoAnomaly is the average pairwise partial correlation among 34 stock market long-short anomalies. The dummy L is 1 if the stock is in the long legs for more than 4 anomalies, and dummy S is 1 if the stock is in the short legs for more than 8 anomalies, and zero otherwise. T-stats, reported in parentheses, are calculated with standard errors clustered by time and firm. The sample period for the dependent variables is 1973Q2—2017Q4.

Panel A: Transition Matrix of Firm-level VAR

	log(Ret)	log(BM)	log(ROE)	Mom	iVol	CoAnomaly	CoAnomaly * L	CoAnomaly * S
log(Ret)	0.0049	0.0206	0.0012	0.0083	-0.0128	0.0022	-0.0019	-0.0068
	(0.22)	(5.36)	(0.08)	(2.21)	(-2.02)	(1.24)	(-0.74)	(-2.30)
log(BM)	-0.0709	0.9496	-0.1490	0.0189	-0.0181	-0.0072	0.0038	0.0031
	(-3.41)	(74.04)	(-10.64)	(5.84)	(-3.41)	(-1.11)	(0.97)	(0.31)
log(ROE)	-0.0659	-0.0396	-0.1432	0.0193	-0.0304	-0.0049	0.0007	-0.0037
	(-4.78)	(-10.29)	(-2.46)	(5.08)	(-7.25)	(-0.81)	(0.19)	(-0.22)
Mom	1.0868	-0.0320	-0.1061	0.6108	0.0209	0.0026	0.0121	0.0032
	(14.57)	(-2.41)	(-3.58)	(14.40)	(1.67)	(0.14)	(0.80)	(0.10)
iVol	-0.5482	0.0228	0.1212	-0.0398	0.7640	-0.0112	-0.0115	0.0105
	(-10.82)	(4.66)	(4.56)	(-6.94)	(30.41)	(-0.93)	(-1.24)	(0.70)
CoAnomaly	0	0	0	0	0	0.6673	0.0022	0.0327
	.	.	.	.	.	(55.03)	(0.24)	(2.17)
CoAnomaly * L	0	0	0	0	0	0.0427	0.3020	0.0038
	.	.	.	.	.	(3.52)	(32.66)	(0.26)
CoAnomaly * S	0	0	0	0	0	0.0090	0.0026	0.3354
	.	.	.	.	.	(0.74)	(0.28)	(22.26)

Panel B: News Function of Firm-level VAR

Discount-Rate News	1.071	0.595	-0.095	0.050	-0.114	-0.001	-0.005	0.009
Cash-Flow News	0.071	0.595	-0.095	0.050	-0.114	-0.001	-0.005	0.009

Table 3.6: Market-Neutral Asset Pricing Tests

$$\overline{R}_i = g_1\widehat{\beta}_{i,CF_{EAR}} + g_2\widehat{\beta}_{i,DR_{EAR}} + g_3\widehat{\beta}_{i,V_{EAR}} + e_i$$

Panel A reports the summary statistics of the estimated news and the time-series correlations among them. Panel B reports the risk premium estimates for different factors in three different asset pricing models. 1-beta CAPM constrains the cash-flow news and discount-rate news having the same price of risk. 2-beta ICAPM and 3-beta ICAPM constrain the risk premium of discount-rate news to be the variance of the equal-weighted anomaly return. Moreover, 3-beta ICAPM also imposes the constraint on risk premia between cash flow news and volatility news. Panel C reports the risk premia estimates for another two cases: constraining the risk premium on discount rates to be the unconditional variance and no restrictions at all. Test assets are the long legs and short legs of the 34 stock market anomalies. T-stats, reported in parentheses, are calculated with GMM and conditioned on the estimated news.

Panel A: Cash-flow, Discount-rate and Variance News for the Equal-weighted Anomaly Portfolio

	E.A.R. Shock	CF News	DR News	Var News (Aggr)	Var News (Avg)	Var News (CoAnomaly)
Mean	0.000	0.001	0.001	0.001	0.001	0.000
Standard Deviation	0.028	0.021	0.013	0.005	0.005	0.006
				Correlation		
Shock E.A.R.	1					
CF News	0.90	1				
DR News	-0.68	-0.28	1			
Var News (Aggr)	-0.31	-0.17	0.38	1		
Var News (Avg)	-0.52	-0.39	0.48	0.88	1	
Var News (CoAnomaly)	0.32	0.32	-0.12	-0.30	-0.28	1

Panel B: Asset Pricing Test with Model Restrictions

	1-beta CAPM	*2-beta ICAPM*	*3-beta ICAPM*		
Volatility Proxy			***Agg. Var***	***Avg. Var***	***CoAnomaly***
Cash-Flow Premium	0.0163	0.0183	0.0164	0.0172	0.0157
	(5.35)	(6.04)	(4.57)	(5.14)	(4.47)
Discount-Rate Premium	0.0163	0.0008	0.0008	0.0008	0.0008
	(5.35)	-	-	-	-
Volatility Premium			-0.0035	-0.0007	-0.0214
			(-0.27)	(-0.06)	(-3.94)
Implied gamma	20.4	22.9	20.5	21.5	19.6
Implied omega	-	-	8.8	1.8	53.5
Overidentifying p-value	0.009	0.013	0.013	0.014	0.018
Cross-sectional Adj. R-squared	48.1%	48.3%	49.4%	49.2%	58.3%

Panel C: Asset Pricing Test without Model Restrictions

	Constrained			*Unrestricted*		
Volatility Proxy	***Agg. Var***	***Avg. Var***	***CoAnomaly***	***Agg. Var***	***Avg. Var***	***CoAnomaly***
Cash-Flow Premium	0.0170	0.0172	0.0138	0.0167	0.0168	0.0166
	(3.64)	(3.79)	(3.15)	(3.06)	(3.26)	(3.42)
Discount-Rate Premium	0.0008	0.0008	0.0008	0.0038	-0.0066	-0.0078
	-	-	-	(0.21)	(-0.57)	(-0.13)
Volatility Premium	-0.0013	-0.0023	-0.0353	-0.0015	-0.0025	-0.0320
	(-1.12)	(-1.37)	(-3.44)	(-0.34)	(-0.92)	(-3.21)
Overidentifying p-value	0.020	0.019	0.022	0.023	0.021	0.032
Cross-sectional Adj. R-squared	61.2%	60.9%	64.2%	62.7%	62.4%	65.3%

Table 3.7: Robustness: Volatility Specification

Panel A reports the news functions $e_1'\rho B(I_K - \rho B)^{-1}$ that map shocks to different state variables to the volatility news for three cases estimated: *no-constraint*, *restricted*, and *no-feedback*. Panel B reports the time-series correlation between the estimated volatility news between CoAnomaly-driven (*Specification 3* in the main result) and three cases above.

Panel A: News Funtion from Shocks to State Variables						
	E.Aggr.Var t	VS t	PE t	MktRf t	DEF t	CoAnomaly t
no-constraint	-2.71	-0.62	29.36	-0.40	2.80	2.05
restricted	-1.88	-1.28	39.73	-0.40	3.37	2.34
no-feedback	0	-1.40	17.55	-0.68	2.89	1.18

Panel B: Time-Series Correlation between Volatility News				
	CoAnomly-driven	no-constraint	restricted	no-feedback
CoAnomly-driven	1			
no-constraint	0.06	1		
restricted	0.56	0.45	1	
no-feedback	0.98	0.11	0.52	1

Table 3.8: Scaled News Terms and Anomaly Betas

Panel A reports the variance-covariance matrix for the in-sample scaled news terms. All raw news is divided by the conditional variance $EVAR$ estimated from the last periods. Panel B reports the estimated betas on cash-flow, discount-rate and CoAnomaly-driven variance news for both long legs and short legs of all 34 anomalies.

Panel A: Variance-Covariance Matrix for Different Scaled News

	E.A.R.	CF	DR	Vol (Aggr)	Vol (Avg)	Vol (CoAnomay)
E.A.R.	0.996					
CF	0.702	0.394				
DR	-0.294	-0.065	0.229			
Vol (Aggr)	0.062	0.038	-0.024	0.032		
Vol (Avg)	-0.087	-0.045	0.041	-0.017	0.019	
Vol (CoAnomay)	-0.056	0.092	0.033	-0.023	0.031	0.012
Standard Deviation	0.998	0.628	0.478	0.178	0.137	0.109

Panel B: Betas of 68 Portfolios

	Long leg			Short leg		
	CF-beta	DR-beta	Vol-beta (CoAnomaly-driven)	CF-beta	DR-beta	Vol-beta (CoAnomaly-driven)
acc	-0.73	-0.07	-0.04	-0.22	-0.20	0.02
atgrowth	0.30	-0.02	-0.09	-0.64	0.00	0.12
ato	0.78	-0.07	-0.15	-0.92	-0.16	0.18
beta	0.48	0.32	-0.56	-0.59	-0.11	0.23
ceissue	0.63	-0.05	-0.13	-1.17	0.33	0.18
failprob	0.39	-0.01	-0.27	-1.26	-0.49	0.22
gm	-0.20	-0.05	0.03	-0.04	-0.08	-0.08
hfcombo1	0.50	-0.12	0.03	-0.62	0.01	-0.06
hfcombo2	0.63	-0.02	0.03	-0.60	0.01	-0.06
idiovol	0.37	0.13	-0.01	-1.14	-0.22	0.04
indmom1m	0.64	-0.06	-0.13	-0.58	-0.02	0.21
invest	0.39	0.06	-0.10	0.13	-0.12	0.08
mom12m	0.01	-0.03	0.06	-1.17	-0.39	0.29
netissue_a	0.49	0.22	-0.06	-0.36	-0.08	0.06
netissue_m	0.39	0.18	-0.09	-0.66	-0.12	0.14
netoa	0.67	-0.01	0.02	-0.51	0.01	-0.08
ohlson	-0.08	-0.04	0.01	-0.65	-0.25	0.00
peadcar3	-0.26	-0.06	-0.05	-0.25	-0.08	0.18
peadsue	0.51	0.06	0.02	-0.64	-0.10	0.01
piotroski	0.10	-0.02	0.03	-1.58	-0.34	0.08
profit	0.37	-0.05	-0.04	-0.60	-0.17	-0.12
relrev1m	-0.85	-0.06	0.29	-0.54	-0.07	-0.06
relrev1mlow	0.43	-0.11	0.10	0.46	0.33	0.06
rev1m	-0.96	-0.11	0.30	-0.10	-0.02	-0.01
rev60m	0.41	-0.03	-0.14	-0.24	-0.06	0.02
roa	0.49	0.00	0.02	-1.38	-0.18	0.10
roe	0.48	0.06	-0.01	-1.44	-0.19	0.17
rome	0.69	0.01	-0.01	-1.36	-0.17	0.14
seasonal	-0.02	0.08	0.11	-0.09	-0.03	-0.15
size	-0.06	-0.03	-0.13	0.00	0.03	-0.02
valmom	0.67	0.06	-0.17	0.13	-0.19	0.04
valmomprof	0.50	0.06	-0.14	-0.51	-0.20	0.29
valprof	0.73	-0.04	-0.06	-0.50	-0.08	0.21
value	0.82	0.11	-0.03	-0.09	-0.09	0.06
E.A.R.	0.29	0.01	-0.04	-0.58	-0.10	0.07

Bibliography

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *American Economic Review*, 106(7):1705–41.

Adrian, T., Etula, E., and Muir, T. (2014). Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, 69(6):2557–2596.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Antón, M. (2011). Cash-flow driven covariation.

Asness, C. S., Frazzini, A., and Pedersen, L. H. (2014). Low-risk investing without industry bets. *Financial Analysts Journal*, 70(4):24–41.

Asness, C. S., Moskowitz, T. J., and Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3):929–985.

Baker, M., Bradley, B., and Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, 67(1):40–54.

Baker, M., Taliaferro, R., and Wurgler, J. (2006). Predicting returns with managerial decision variables: Is there a small-sample bias? *The Journal of Finance*, 61(4):1711–1730.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680.

Barroso, P., Edelen, R. M., and Karehnke, P. (2017). Institutional crowding and the moments of momentum.

Barroso, P. and Maio, P. F. (2018). The risk-return tradeoff among equity factors.

Barroso, P. and Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120.

Berger, D., Dew-Becker, I., and Giglio, S. (2017). Uncertainty shocks as second-moment news shocks.

Bergmeir, C. and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213.

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, 45(3):444–455.

Blitz, D. C. and Van Vliet, P. (2007). The volatility effect. *The Journal of Portfolio Management*, 34(1):102–113.

Brandt, M. W., Kishore, R., Santa-Clara, P., and Venkatachalam, M. (2008). Earnings announcements are full of surprises. *SSRN eLibrary*.

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447.

Bryzgalova, S. (2014). Spurious factors in linear asset pricing models.

Buraschi, A., Kosowski, R., and Trojani, F. (2013). When there is no place to hide: Correlation risk and the cross-section of hedge fund returns. *The Review of Financial Studies*, 27(2):581–616.

Buraschi, A., Porchia, P., and Trojani, F. (2010). Correlation risk and optimal portfolio choice. *The Journal of Finance*, 65(1):393–420.

Campbell, J. (1991). A variance decomposition for stock returns. *Economic Journal*, 101(405):157–79.

Campbell, J. Y., Giglio, S., Polk, C., and Turley, R. (2017). An intertemporal capm with stochastic volatility. *Journal of Financial Economics*.

Campbell, J. Y., Polk, C., and Vuolteenaho, T. (2009). Growth or glamour? fundamentals and systematic risk in stock returns. *The Review of Financial Studies*, 23(1):305–344.

Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies*, 1(3):195–228.

Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Vuolteenaho, T. (2004). Bad beta, good beta. *The American Economic Review*, 94(5):1249–1275.

Cao, J., Hsu, J. C., Xiao, Z., and Zhan, X. (2018). Smart beta,'smarter'flows.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Chen, L., Kadan, O., and Kose, E. (2009). Fresh momentum.

Chinco, A., Clark-Joseph, A. D., and Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1):449–492.

Cho, T. (2017). Turning alphas into betas: Arbitrage and endogenous risk.

Chordia, T., Roll, R., and Subrahmanyam, A. (2011). Recent trends in trading activity and market quality. *Journal of Financial Economics*, 101(2):243–263.

Clarke, R., Silva, H. d., and Thorley, S. (2013). Risk parity, maximum diversification, and minimum variance: An analytic perspective. *The Journal of Portfolio Management*, 39(3):39–53.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Cohen, R. B., Polk, C., and Vuolteenaho, T. (2003). The value spread. *The Journal of Finance*, 58(2):609–642.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3):373–394.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of Financial Economics*, 21(2):255–289.

Cooper, M. J., Gutierrez, R. C., and Hameed, A. (2004). Market states and momentum. *The Journal of Finance*, 59(3):1345–1365.

Daniel, K., Klos, A., and Rottke, S. (2017). Overpriced winners.

Daniel, K. and Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2):221–247.

DeMiguel, V., Garlappi, L., and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

Dew-Becker, I., Giglio, S., Le, A., and Rodriguez, M. (2017). The price of variance risk. *Journal of Financial Economics*, 123(2):225–250.

Djogbenou, A., Gonçalves, S., and Perron, B. (2015). Bootstrap inference in regressions with estimated factors and serial correlation. *Journal of Time Series Analysis*, 36(3):481–502.

Driessen, J., Maenhout, P. J., and Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance*, 64(3):1377–1406.

Epstein, L. G. and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of political Economy*, 99(2):263–286.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84.

Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Feng, G., Giglio, S., and Xiu, D. (2017). Taming the factor zoo.

Ferson, W. E., Sarkissian, S., and Simin, T. T. (2003). Spurious regressions in financial economics? *The Journal of Finance*, 58(4):1393–1413.

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1–25.

Gargano, A. and Rossi, A. G. (2018). Does it pay to pay attention? *The Review of Financial Studies*, 31(12):4595–4649.

Garleanu, N., Panageas, S., and Yu, J. (2012). Technological growth and asset pricing. *The Journal of Finance*, 67(4):1265–1292.

Ghosh, A., Julliard, C., and Taylor, A. P. (2016). An information-theoretic asset pricing model.

Green, J., Hand, J. R., and Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, page hhx019.

Greenwood, R. and Hanson, S. G. (2012). Share issuance and factor timing. *The Journal of Finance*, 67(2):761–798.

Grundy, B. D. and Martin, J. S. M. (2001). Understanding the nature of the risks and the source of the rewards to momentum investing. *The Review of Financial Studies*, 14(1):29–78.

Gu, S., Kelly, B., and Xiu, D. (2018). Empirical asset pricing via machine learning.

Guo, J. T. (2018). Coanomaly: Correlation risk in stock market anomalies.

Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of political economy*, 99(2):225–262.

Hanson, S. G. and Sunderam, A. (2013). The growth and limits of arbitrage: Evidence from short interest. *The Review of Financial Studies*, 27(4):1238–1286.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). . . . and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

He, Z., Kelly, B., and Manela, A. (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of Finance*, 54(6):2143–2184.

Hou, K., Xue, C., and Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705.

Hou, K., Xue, C., and Zhang, L. (2017). Replicating anomalies.

Hu, G. X., Pan, J., and Wang, J. (2013). Noise as information for illiquidity. *The Journal of Finance*, 68(6):2341–2382.

Huang, S. (2015). The momentum gap and return predictability. In *WFA 2015 Seattle Meetings Paper*.

Huang, S., Lou, D., and Polk, C. (2016). The booms and busts of beta arbitrage.

Jacobs, B. I. and Levy, K. N. (2014). Smart beta versus smart alpha. *Journal of Portfolio Management*, 40(4):4–7.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48(1):65–91.

Jegadeesh, N. and Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance*, 56(2):699–720.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Khandani, A. and Lo, A. (2007). What happened to the quants in august 2007?(digest summary). *Journal of investment management*, 5(4):29–78.

Khandani, A. E. and Lo, A. W. (2011). What happened to the quants in august 2007? evidence from factors and transactions data. *Journal of Financial Markets*, 14(1):1–46.

Koijen, R. S. and Yogo, M. (2015). An equilibrium model of institutional demand and asset prices.

Kondor, P. and Vayanos, D. (2019). Liquidity risk and the dynamics of arbitrage capital. *The Journal of Finance*, 74(3):1139–1173.

Kothari, S. P. and Shanken, J. (1992). Stock return variation and expected dividends: A time-series and cross-sectional analysis. *Journal of Financial Economics*, 31(2):177–210.

Kozak, S., Nagel, S., and Santosh, S. (2017). Shrinking the cross section.

Kozak, S., Nagel, S., and Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, 73(3):1183–1223.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849.

Lettau, M. and Wachter, J. A. (2007). Why is long-horizon equity less risky? a duration-based explanation of the value premium. *The Journal of Finance*, 62(1):55–92.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics*, 96(2):175–194.

Lochstoer, L. A. and Tetlock, P. C. (2017). What drives anomaly returns?

Lou, D. and Polk, C. (2013). Comomentum: Inferring arbitrage activity from return correlations.

Maio, P. and Santa-Clara, P. (2012). Multifactor models and their consistency with the icapm. *Journal of Financial Economics*, 106(3):586–613.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.

Martin Utrera, A., DeMiguel, V., Uppal, R., and Nogales, F. J. (2017). A transaction-cost perspective on the multitude of firm characteristics.

McCracken, M. W. (2007). Asymptotics for out of sample tests of granger causality. *Journal of econometrics*, 140(2):719–752.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, pages 247–257.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica*, 55(3):703–708.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics*, 103(3):429–453.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics*, 112(2):137–146.

Novy-Marx, R. (2015). Fundamentally, momentum is fundamental momentum.

Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.

Ohlson, J. A. (1995). Earnings, book values, and dividends in equity valuation. *Contemporary accounting research*, 11(2):661–687.

Pástor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, 55(1):179–223.

Pástor, L. and Stambaugh, R. F. (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics*, 56(3):335–381.

Pastor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3).

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1):435–480.

Pollet, J. M. and Wilson, M. (2010). Average correlation and stock market returns. *Journal of Financial Economics*, 96(3):364–380.

Roll, R. (1977). A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. *Journal of financial economics*, 4(2):129–176.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1:939–974.

Shanken, J. (1992). On the estimation of beta-pricing models. *The review of financial studies*, 5(1):1–33.

Shumway, T. and Warther, V. A. (1999). The delisting bias in crsp's nasdaq data and its implications for the size effect. *The Journal of Finance*, 54(6):2361–2379.

Sotes-Paladino, J. M. (2017). Optimal trading in correlated anomalies.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.

Stambaugh, R. F. and Yuan, Y. (2016). Mispricing factors. *The Review of Financial Studies*, 30(4):1270–1315.

Stein, J. C. (2009). Presidential address: Sophisticated investors and market efficiency. *The Journal of Finance*, 64(4):1517–1548.

Stivers, C. and Sun, L. (2010). Cross-sectional return dispersion and time variation in value and momentum premiums. *Journal of Financial and Quantitative Analysis*, 45(4):987–1014.

Stutzer, M. (1995). A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics*, 68(2):367–397.

Vayanos, D. and Woolley, P. (2013). An institutional theory of momentum and reversal. *Review of Financial Studies*, 26(5):1087–1145.

Vuolteenaho, T. (2002). What drives firm-level stock returns? *The Journal of Finance*, 57(1):233–264.

Wang, K. Q. and Xu, J. (2015). Market volatility and momentum. *Journal of Empirical Finance*, 30:79–91.

Welch, I. and Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Zhang, L. (2005). The value premium. *The Journal of Finance*, 60(1):67–103.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.