Essays on Asset Pricing and Portfolio Optimization

D I S S E R T A T I O N
of the University of St.Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Finance

submitted by

Christian Koeppel

from

Germany

Approved on the application of

Prof. Dr. Roland Füss

and

Prof. Dr. Karl Frauendorfer

Dissertation no. 5120

Difo-Druck GmbH, Untersiemau 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, May 25, 2021

The President:

Prof. Dr. Bernhard Ehrenzeller

Acknowledgement

Writing a doctoral thesis is a time-consuming and demanding challenge, requiring persistence, discipline, and compromises over an extended period of time. While this obviously applies to the author, the sacrifices of his social environment are too often overlooked. The completion of this work, however, would have been impossible without the motivation and support of many people whom I would like to acknowledge here.

I express my sincere gratitude to my supervisor Roland Füss. His initial skepticism whether the completion of a part-time PhD is at all feasible, was additional incentive for me to prove him otherwise. I am deeply indebted to him for still taking me on-board and giving me the trust and opportunity to pursue a PhD under his supervision alongside my work. I would like to express my deep gratitude for his patience, encouragement, and useful critiques of my research work as well as for being approachable at any time over the last five years. I know that this cannot be taken for granted.

Furthermore, I would like to thank my supervisors at work, Heiko Sundermann and Niklas Loer, for the opportunity to pursue both my professional and academic career. Without their flexibility and support this endeavor would have been impossible.

My special thanks are extended to my co-authors Massimo Guidolin and Felix Miebs for many motivational discussions and their timely and in-depth review of my work. Your passion for your research and aim for perfection were truly inspiring.

I would also like to express my gratitude to my family. Their love provided me with the necessary strength and motivation. Especially, my sister with her own pursue of a PhD was an appreciated sparring partner and reviewer of my academic work.

Special thanks goes also to my girlfriend, Megan Morrow, on whose support I could always count. In particular, in the last months she kept me on track, kept my back free, always put a smile on my face, made sure I kept hydrated and fed with cake. I cannot imagine how the last years would have been without her.

Dietikon, 22.06.2021 Christian Koeppel

Table of Contents

List of Figures **5**

List of Tables **6**

Summary **8**

Zusammenfassung **9**

Essay I: Sentiment risk premia in the cross-section of global equity **10**

1 Introduction . . . 12
2 Is sentiment a priced risk factor? . . . 18
2.1 Related literature . . . 18
2.2 Theoretical foundations . . . 21
3 Data . . . 24
3.1 Sentiment indicators . . . 24
3.2 Comparison of sentiment indicators . . . 26
3.3 Global equity markets data . . . 30
4 Sentiment in the cross-section of portfolio sorts . . . 31
5 Sentiment-augmented asset pricing models . . . 33
5.1 Linear factor models including sentiment . . . 35
5.2 Benchmarking MarketPsych against other sentiment proxies . . . 46
5.3 Global versus local sentiment . . . 48
6 Sentiment risk premia . . . 49
6.1 Risk premia for global sentiment . . . 51
6.2 Long-term sentiment premia for developed and emerging markets 53
7 Robustness checks . . . 56
7.1 State-dependent Fama-French five-factor model . . . 56
7.2 Tests on an extended set of global equity indices . . . 58
7.3 Sub-samples analysis . . . 58
7.4 Separating the effects of news from social media channels . . . 59
7.5 Fama-MacBeth without orthogonalization of the factors . . . 60
8 Conclusion . . . 60
A Supplementary tables . . . 64
B Details on Refinitiv-MarketPsych's sentiment construction . . . 66
C Orthogonalization approach . . . 77
D Market efficiency test . . . 78
E Additional tables . . . 79
F Sub-sample analysis using NBER classification . . . 90

Essay II: Does social media sentiment matter in the pricing of U.S. stocks? 93

1 Introduction . . . 95
2 U.S. equity stocks and sentiment indicators . . . 98

3 Factor-mimicking portfolios . . . 103
4 Risk factor construction and summary statistics . . . 110
5 Sentiment-augmented asset pricing models . . . 114
6 Relation between fundamental factors and sentiment . . . 120
7 Regression details of hard-to-value stock portfolios . . . 123
7.1 Extreme growth stocks . . . 124
7.2 Extreme investment stocks . . . 124
7.3 Small stocks that invest heavily despite weak profitability . . . 126
7.4 Sentiment-mimicking portfolios . . . 129
8 News-based sentiment . . . 131
9 Conclusion . . . 132
A Details on Refinitiv-MarketPsych's sentiment construction . . . 135
B News-based sentiment . . . 142
C Sentiment, momentum, and liquidity . . . 147

Essay III: Diversifying estimation errors: An efficient averaging rule for portfolio optimization 150

1 Introduction . . . 152
2 Portfolio optimization and estimation errors . . . 155
2.1 Minimum-variance portfolio optimization . . . 155
2.2 Alleviating estimation errors: Existing approaches . . . 155
2.3 Estimation error diversification: An averaging approach . . . 158
3 Simulation study . . . 161
3.1 Simulation procedure . . . 161
3.2 Simulation results . . . 164
4 Empirical study . . . 173
4.1 Data . . . 173
4.2 Performance evaluation . . . 174
4.3 Out-of-sample results . . . 176
5 Conclusion . . . 187
A Description of estimators . . . 189
B Additional simulation results . . . 194
C Results on monthly data frequency . . . 215
D Portfolio strategies and averaging with short-sale constraint . . . 222

References 226

Curriculum Vitae 235

List of Figures

Essay I: Sentiment risk premia in the cross-section of global equity 10

1 Schematic difference between Refinitiv-MarketPsych and Baker-Wurgler . 27

2 Refinitiv-MarketPsych and Baker-Wurgler indicators for the U.S. 30

3 Cumulative returns of global sentiment portfolio sorts 41

4 Comparing the R^2 of the CAPM to the sentiment-augmented CAPM . . 46

B.1 Sentiment classification system: valence and arousal 69

B.2 Example of MarketPsych's human language processing system 70

B.3 Equity markets and their local sentiment 73

Essay II: Does social media sentiment matter in the pricing of U.S. stocks? 93

1 Cumulative returns of the Fama-French and sentiment risk factors 115

A.1 Sentiment classification system: valence and arousal 138

A.2 Example of MarketPsych's human language processing system 139

Essay III: Diversifying estimation errors: An efficient averaging rule for portfolio optimization 150

1 Average model weights in the simulated data sets 170

2 Average strategy weights of AV-Wgt for portfolios and stocks 179

B.3 Kernel densities of the out-of-sample variances and Sharpe ratios of AV-Wgt 197

B.4 Kernel densities of the out-of-sample variances and Sharpe ratios of AV-Inv 199

B.5 Kernel densities of the out-of-sample variances and Sharpe ratios of AV-Cov 201

B.6 Kernel densities of the out-of-sample variances of AV-Wgt compared to established strategies and benchmarks 203

B.7 Kernel densities of the out-of-sample variances of AV-Inv compared to established strategies and benchmarks 205

B.8 Kernel densities of the out-of-sample variances of AV-Cov compared to established strategies and benchmarks 207

B.9 Kernel densities of the out-of-sample Sharpe ratio of AV-Wgt compared to established strategies and benchmarks 209

B.10 Kernel densities of the out-of-sample Sharpe ratio of AV-Inv compared to established strategies and benchmarks 211

B.11 Kernel densities of the out-of-sample Sharpe ratio of AV-Cov compared to established strategies and benchmarks 213

List of Tables

Essay I: Sentiment risk premia in the cross-section of global equity 10

1 Portfolios sorted by sentiment . . . 34
2 Comparing the CAPM to alternative sentiment-augmented CAPM models 40
3 Comparing Fama-French factors and momentum with sentiment-augmented multi-factor models . . . 43
4 Cross-sectional pricing comparison of the CAPM, FF5, and sentiment-augmented models . . . 54
5 Cross-sectional pricing comparison between developed and emerging markets 57
A.1 Global equity indices list . . . 64
A.2 Pearson correlation analysis between predictor Variables . . . 65
B.1 Construction of the Refinitiv-MarketPsych sentiment indicator . . . 68
B.2 Description of Refinitiv-MarketPsych indicators . . . 71
D.1 Market efficiency tests . . . 78
E.1 Comparing the Refinitv-MarketPysch sentiment indicators with alternative sentiment measures . . . 80
E.2 Comparing the CAPM with local and global sentiment . . . 82
E.3 Factor loadings and statistics for principal component analysis (PCA) . . 84
E.4 A Fama-French five-factor model with sentiment interactions . . . 85
E.5 Comparing the CAPM to different sentiment-augmented CAPM for additional indices . . . 86
E.6 Sub-samples comparisons . . . 87
E.7 Comparison between news- and social media-based sentiment indicators . 88
E.8 Cross-sectional pricing comparisons of the CAPM and the SNT-CAPM without orthogonalization . . . 89
F.1 Sub-samples Comparisons - NBER Recession Dates . . . 92

Essay II: Does social media sentiment matter in the pricing of U.S. stocks? 93

2 List of economic indicators . . . 101
3 Regressions of social media-based sentiment against economic indicators . 102
4 Descriptive statistics of social media-based sentiment by groups . . . 104
5 5x5 portfolios using social media-based sentiment . . . 107
6 2x2x3x3 portfolios using social media-based sentiment . . . 109
7 Factor construction . . . 111
8 Summary statistics for factor returns using social media-based sentiment 113
9 Summary statistics for regressions using social media-based sentiment . . 119
10 Regressions on factors using social media-based sentiment . . . 122
11 Regressions on Size-B/M portfolios using social media-based sentiment . 125
12 Regressions on Size-Inv portfolios using social media-based sentiment . . 127
13 Regressions on Size-B/M-OP-Inv portfolios using social media-based sentiment 128

14 Regressions on Size-Snt portfolios using social media-based sentiment . . 130
A.1 Construction of the Refinitiv-MarketPsych sentiment indicator 137
A.2 Description of Refinitiv-MarketPsych indicators 140
B.1 5x5 portfolios using news-based sentiment 142
B.2 2x2x3x3 portfolios using news-based sentiment 143
B.3 Summary statistics for factor returns using news-based sentiment 144
B.4 Summary statistics for regressions using news-based sentiment 145
B.5 Summary statistics for regressions using orthogonalized social media-based sentiment 146
C.1 Summary statistics for factor returns including momentum using social media-based sentiment 148
C.2 Summary statistics for factor returns including liquidity using social media-based sentiment 149

Essay III: Diversifying estimation errors: An efficient averaging rule for portfolio optimization 150

3 List of averaging rule, single strategies, and benchmarks 162
4 Summary statistics of simulation parameters 163
5 Descriptive statistics of estimation errors across estimation windows . . . 166
6 Simulation-based average out-of-sample variances for a varying number of strategies 168
7 Simulation-based average out-of-sample variances for the averaging rule, strategies, and benchmarks 171
8 Simulation-based average out-of-sample Sharpe ratios for the averaging rule, strategies, and benchmarks 172
9 List of the empirical data sets for daily data 175
10 Daily out-of-sample variances in empirical data sets 178
11 Daily out-of-sample Sharpe ratios in empirical data sets 181
12 Average monthly turnover in empirical data sets based on daily observations 183
13 Average daily short interest in empirical data sets 184
14 Daily out-of-sample Sharpe ratios adjusted for transaction costs in empirical data sets 186
B.15 Sharpe ratio comparison of averaging strategies in terms of averaged strategies 196
C.16 List of the empirical data sets for monthly data 217
C.17 Monthly out-of-sample variances in empirical data sets 218
C.18 Monthly out-of-sample Sharpe ratios in empirical data sets 219
C.19 Average monthly turnover in empirical data sets based on monthly observations 220
C.20 Average monthly short interest in empirical data sets 221
D.21 Daily out-of-sample variances with short-sale constraints 223
D.22 Daily out-of-sample Sharpe ratios with short-sale constraints 224
D.23 Average monthly turnover with short-sale constraints 225

Summary

This doctoral thesis focuses on the effects of investor sentiment on asset pricing and the challenges of portfolio optimization under parameter uncertainty.

The first essay "Sentiment risk premia in the cross-section of global equity" applies a recently developed sentiment proxy to the construction of a new risk factor and provides a comprehensive understanding of its role in sentiment-augmented asset pricing models for international equity indices. We empirically demonstrate the existence of a statistically significant and economically relevant sentiment premium. Differentiating between developed and emerging markets we reveal different patterns of return reversals / persistence. Our results contribute to the explanation of global cross-sectional average excess returns, demonstrating superiority in terms of predictive power when compared to competing definitions of sentiment.

The second essay "Does social media sentiment matter in the pricing of U.S. stocks?" finds that the inclusion of micro-grounded, social media-based sentiment significantly improves the performance of the five-factor model from Fama and French (2015, 2017). This holds for different industry and style portfolios such as size, value, profitability, and investment. Applying a robust GMM estimator, the sentiment risk premium provides the missing component in the behavioral asset pricing theory of Shefrin and Belotti (2008) and (partially) resolves the pricing puzzles of small extreme growth, small extreme investment stocks and small stocks that invest heavily despite low profitability.

The third essay "Diversifying estimation errors: An efficient averaging rule for portfolio optimization" proposes a combination of established minimum-variance strategies to minimize the expected out-of-sample variance. The proposed averaging rule overcomes the strategy selection problem and diversifies estimation errors of the strategies included in our rule. Extensive simulations show that the contributions of estimation errors to the out-of-sample variances are uncorrelated between the considered strategies. We therefore conclude that averaging over multiple strategies offers sizable diversification benefits.

Zusammenfassung

In der vorliegenden Arbeit wird der Frage nachgegangen, wie sich die Anlegerstimmung auf die Preisgestaltung von Aktien auswirkt. Zudem werden Herausforderungen der Portfoliooptimierung unter Parameterunsicherheit eingehend untersucht.

Das erste Papier "Sentiment risk premia in the cross-section of global equity" stellt ein neuartiges Stimmungsbarometer für Aktienbewertungsmodelle vor und untersucht deren Rolle auf internationale Aktienindizes. Die empirischen Befunde legen nahe, dass AnlegerInnen eine statistisch signifikante und wirtschaftlich relevante Sentiment-Risikoprämie fordern. Im Vergleich zwischen Industrie- und Schwellenländern zeigen sich langfristig unterschiedliche Muster von Trendwenden und Nachhaltigkeit der Erträge. Die empirischen Befunde liefern einen wertvollen Beitrag zur Erklärung durchschnittlicher Überschussrenditen und erweisen sich als aussagekräftiger als bisherige Stimmungsbarometer.

Das zweite Papier "Does social media sentiment matter in the pricing of U.S. stocks?" zeigt, dass die Einbeziehung eines mikro-fundierten, auf sozialen Medien basierenden Stimmungsbarometers die Aussagekraft des Fünf-Faktoren-Modell von Fama and French (2015, 2017) für verschiedene Branchen- und Stilportfolios, die sich hinsichtlich ihrer Größe, Buch-Kursverhältnisses, Rentabilität und Investitionsbereitschaft unterscheiden, erheblich verbessert. Unter Verwendung eines robusten GMM-Schätzers stellt die entsprechende Risikoprämie die in der Verhaltensportfoliotheorie von Shefrin and Belotti (2008) bislang noch fehlende Komponente dar, mit der die bisher nur unzureichend erfasste Preisbildung schwer zu bewertender Aktien besser erklärt werden kann.

Aus dem dritten Papier "Diversifying estimation errors: An efficient averaging rule for portfolio optimization" geht eine neue Mittelungsregel hervor, die die Portfoliovarianz mittels einer Kombination etablierter Strategien zur Reduktion von Schätzfehlern in der Kovarianzmatrix verringert. Durch Anwendung dieser Regel können die genannten Schätzfehler diversifiziert werden. Umfangreiche Simulationen zeigen, dass keine Korrelationen zwischen den Schätzfehlern der berücksichtigten Strategien bestehen, weshalb die Mittelungsregel erhebliche Diversifizierungsvorteile bietet.

Essay I: Sentiment risk premia in the cross-section of global equity

Sentiment risk premia in the cross-section of global equity*

Roland Füss† Massimo Guidolin‡ Christian Koeppel§

Abstract

This paper applies a recently developed sentiment proxy to the construction of a new risk factor and provides a comprehensive understanding of its role in sentiment-augmented asset pricing models. We find that news and social media search-based indicators are significantly related to excess returns of international equity indices. Adding sentiment factors to both classical and more recent linear factor pricing models leads to a significant increase in their performance. When it is estimated using the Fama-MacBeth procedure, our sentiment-adjusted pricing model implies positive (negative) estimates of the risk premium for positive (negative) sentiment factors. We further differentiate between developed and emerging markets and uncover different patterns of return reversals / persistence in the long-term. Our results contribute to the explanation of global cross-sectional average excess returns and are robust to augmenting the model with fundamental factors, momentum, idiosyncratic volatility, skewness, kurtosis, and the returns on international currencies. When compared to competing definitions of sentiment factors popular in the literature, our novel sentiment risk variable turns out to be superior in terms of predictive power.

JEL Classification Codes: $C53$, $G12$, $G41$

Keywords: *Asset pricing*; *behavioral finance*; *financial markets*; *investor sentiment*; *sentiment risk premium.*

*We are grateful to Richard Peterson and Elijah DePalma for granting access to the MarketPsych indicators as well as to Claudius Ulmer for providing the financial market data. We also thank Francesco Audrino, Martin Brown, Bertraind Calderon, Giacomo Nocera, Steffen Meyer, Nancy Mo, Zakaria Moussa, as well as the seminar and session participants to the 19th Cologne Colloquium on Financial Markets, the Financial Management Association Annual Meetings, the 33rd Australasian Finance and Banking Conference, the 2020 Virtual World Finance & Banking Symposium, the 6th PFMC Conference in Paris, the 6th LOD conference in Siena, and the Finance Research Seminars at University of St.Gallen, Université Catholique de Louvain, Université de Nante Audencia Business School, and the University of Liechtenstein for their helpful comments and suggestions.

†Swiss Institute of Banking and Finance (s/bf), University of St.Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland, Research Associate at the Centre for European Economic Research (ZEW), Mannheim, Germany and a Visiting Professor and Fellow at the Center for Real Estate and Environmental Economics at NTNU Business School, Trondheim, Norway; Phone: +41 (0)71 224-7055, Fax: +41 (0)71 224-7088; Email: roland.fuess@unisg.ch

‡Bocconi University, Finance dept., via Roetgen 1, 20136 Milan, Italy; Phone: +39 (0)258 36 35 56; Email: massimo.guidolin@unibocconi.it

§University St.Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland; Phone: +41 (0)76 38 30 757, Email: christian.koeppel@live.com

1 Introduction

Classical finance theories rest upon the assumption that investors are rational and form their return expectations based on fundamental values and hard-fact news. These models work well in "normal" times but fail to capture deviations in pricing from their intrinsic values, at least in volatile and extreme-sentiment market phases (see, e.g., Yu & Yuan, 2011). The observed over- and under-reaction of asset prices to news announcements (see, e.g., Abarbanell & Bernard, 1992; Frazzini, 2006; Sinha, 2016; Veronesi, 1999), often followed by sudden plunges in states of financial turmoil, cannot be entirely explained by rational behavior based on fundamental factors. Instead, these abnormal price dynamics have been traced back to irrational investor sentiment waves, like fear and panic, greed and overconfidence, all of which are known to influence human decision-making (see, e.g., Barberis, Shleifer, & Vishny, 1998; Ben-Rephael, Da, & Israelsen, 2017; Ottaviani & Sørensen, 2015). Hence, these investors' moods which generally fail to be related to objective, fundamental features of traded securities, appear to drive asset prices via supply and demand forces (see, e.g., Bushee & Friedman, 2016).

In this paper, we test this key behavioral hypothesis by estimating sentiment risk premia in international equity markets leveraging on recently developed, direct search-based investor sentiment indicators. These measures are derived from human language processing and distilled from newly available bases of data that collect asset-specific information as it circulates through public news as well as social media channels (see, e.g., Chen, De, Hu, & Hwang, 2014). Following Shefrin and Belotti (2008), we hypothesize that sentiment may be priced differently depending on the type of events it reflects, whether positive or negative. Given that sentiment covers a spectrum of moods rather just an aggregated level, sentiment risk commands separately signed and estimable risk premia for positive and negative valences. By adding these mood factors to classical asset pricing models, we find that positive and negative deviations of sentiment from its long-term mean, as commonly observed in extreme market states and in the presence of tail events, add significant explanatory power to otherwise standard linear pricing models.

Using newly created sentiment measures based on Refinitiv (formerly, Thomson Reuters) MarketPsych's human-language processing engine applied to news and social media feeds to construct a novel sentiment risk factor, this paper establishes the existence of a strong predictive relationship between sentiment, several well-known systematic risk factors (e.g., market, size, value, profitability, investment, momentum) and the excess returns of a number of international equity indices. To circumvent any concerns regarding endogeneity or reverse causality, we use lagged sentiment in line with Nofsinger (2005) who finds that business activity tends to follow, rather than lead social moods. We show that our international sentiment factor(s) contain(s) additional explanatory power over standard factors and contribute(s) to the partial resolution of the long-standing puzzle of the dispersion in cross-sectional equity premia in the international dimension. Exploiting the fact that global equity indices each represent a (locally) diversified portfolio, we show that their excess returns can be better explained by adding sentiment factors to asset pricing models. These contain otherwise well-known factors that relate to the largely understood mean-variance, risk-return trade-off logic, or to profitability, investment, and value. We also provide an explanation to the occurrence of persistent mispricing during bubbles and financial crashes by separating the effect of sentiment into positive and negative measures that reflect the structure of its differential time-variation relative to economic and financial market cycles.

The sentiment indicators used in our empirical analyses are based on a measure of micro-grounded, bottom-up news and social media sentiment. The findings derived from portfolio sorts and linear factor model estimation support the hypothesis that sentiment represents an aggregate measure of investor beliefs (whether rational or not) on the outlook of cash flows and future asset values. We uncover a significant relationship between abnormal sentiment shifts and realized returns. In particular, a standard sorting of international equity indices into portfolios provides empirical evidence that high (low) abnormal sentiment scores are followed by large, positive (negative) realized average excess returns. We show that positive (negative) deviations of sentiment from its long-term mean, i.e., positive (negative) *abnormal sentiment*, forecasts positive (negative) excess

equity index returns, while *average sentiment* has no significant effect. Interestingly, this result differs from the empirical finding that positive (negative) returns follow negative (positive) sentiment as reported by Baker and Wurgler (2006, 2007). Based on a lead-lag correlation analysis, we argue that our sentiment indicator captures a different life-cycle of the emotional processes of investors and therefore may serve as a leading indicator that precedes a composite index of fundamental variables of the type advocated by Baker and Wurgler (2006, 2007). Furthermore, under the hypothesis that sentiment is a direct measure of investor mood, it is likely to affect individual regional markets differently and reflects their heterogeneous exposures to the unavoidable swings in the *relative* moods for different assets and across geographies. A sentiment breakdown based on the underlying indices proves that there are sentiment-sensitive markets and others, which retain instead a prevailing correlation to classical risk factors. We relate this heterogeneity to their overall level of market efficiency. Following Griffin, Hirschey, and Kelly (2011), we associate sentiment with reduced informational efficiency, making investors in inefficient markets more subject to behavioral biases, sentiment, and herding. This is consistent with hard-fact news being generally more useful to forecast stock returns in developed markets than in emerging markets due to the better informational environment, more effective mechanisms of news dissemination, the lower level of insider trading, and possibly technologically higher developed trading platforms. [1].

Our empirical results open a further avenue of inquiry to explain a number of asset pricing anomalies, based on the introduction of a novel, global sentiment risk factor. However, due to the complexity of the dynamics of sentiment as a risk factor, we also find evidence that a simple "positive-minus-negative sentiment" (PMNSNT) factor in the style of Fama and French (1993) cannot fully capture the priced contribution of sentiment to asset pricing relationships. The international coverage and sophistication of MarketPsych's indicators allow us to split the global factor into positive and negative sentiment portfolios, respectively, to cover the multi-dimensional plane of human emotions

[1] Additionally, by separately estimating the sentiment risk premia in developed versus emerging markets, we uncover different patterns as the forecasting horizons lengthens, which allows us to discriminate between sentiment and information theories in Tetlock (2007)

along the arousal-valence framework in a holistic fashion. Using the excess returns of the negative versus positive sentiment sorted portfolios as factor-mimicking representations, our estimates show that positive and negative sentiment factors are differently priced. Negative sentiment leads to statistically significant and sizable underperformance compared to standard models, while positive sentiment bears a significant positive risk premium. These findings provide the empirical link to the theoretical model by Shefrin and Belotti (2008) who argue that sentiment is best understood as a distribution rather than as a scalar. Simply netting excessively bullish and bearish emotions in a single market sentiment factor may result in an oversimplified characterization, which is exactly what our empirical findings show.

We use these insights to benchmark different sentiment-augmented linear factor models against the standard Capital Asset Pricing Model (henceforth CAPM) and more recent asset pricing models. In this process, we resort to both the estimation of simple time series models and to the more sophisticated two-stage approach introduced by Fama and MacBeth (1973). The Fama-MacBeth method (henceforth FMB) provides estimates for the price of sentiment risk in our cross-section of international equity index returns. Because we want to rule out that sentiment may simply proxy for other, fundamental risk factors we perform systematic comparisons with the Fama-French five-factor model (henceforth FF5). Based on the FMB specification, we show that a sentiment-augmented pricing model outperforms the CAPM and the FF5. To reduce any omitted variable bias, we further control for the returns on international currencies, for lagged statistical stock return features like idiosyncratic volatility, skewness or kurtosis, as well as for momentum. We argue that our sentiment indicators and, in particular, the positive and negative deviations from its long-term mean factors, capture a new notion of pure investor sentiment that usefully separates fear-driven and bullish mood dynamics. This is in contrast to the existing literature which primarily focuses on negative sentiment (see, e.g., Tetlock, 2007) affecting the left tail of the objective, backward-looking return distribution, and hence may mostly capture (presumably rational) aversion to left-skewness, i.e., losses from the extreme left tail. We therefore contribute to the literature by adding another

key piece to the mosaic explaining the cross-section of international excess equity returns recently investigated by a number of papers (see, e.g., Hou, Xue, & Zhang, 2020; Lettau, Maggiori, & Weber, 2014).

In a range of robustness checks, we also control for the effects of economic cycles, we differentiate sentiment factors extracted from news and social media, and benchmark our sentiment measure against alternative fundamental-, market-, and search-based sentiment proxies now popular in the empirical finance literature. Our main results still hold when using sub-samples of our data to explore sentiment during alternative bull and bear market phases. We use both a naïve classification of bull and bear markets and the categorization of recessions published by the National Bureau of Economic Research (NBER). We show that pure social media-based factors are better sentiment predictors, presumably because they are more strongly related to the investment behavior of retail investors. We compare our sentiment indices with a number of well-known sentiment proxies from the literature (such as Baker and Wurgler's (2006) sentiment measures and Da et al.'s (2015) FEARS index) as well as with directly competing sentiment indicators that share a similar search-based bottom-up construction (such as RavenPack's event sentiment score). We demonstrate the superiority of our sentiment indicators in terms of predictive power for most of (in some cases, all) the country equity indices under investigation. We also test an alternative hypothesis that treats sentiment not as an explicit risk factor but rather as a conditioning state variable that changes the pricing of existing, fundamental factors from the classical FF5 model. The results, however, do not provide sufficient evidence to reject our original hypothesis according to which sentiment is a direct risk driver in an international asset pricing context. We further show that in line with Baker, Wurgler, and Yuan (2012), equity markets are driven by global rather than local sentiment.

Two related papers are Sun, Najand, and Shen (2016) and Jiao, Veiga, and Walther (2020) who have also used Thomson Reuters (now Refinitiv) MarketPsych Indices (RMI). Sun et al. (2016) focus on the predictive relation between high-frequency investor sentiment and S&P 500 returns. They report evidence that index returns are predictable using

lagged half-hour sentiment even when intraday momentum is controlled for. Jiao et al. (2020) use RMI daily data on approximately 3,000 individual US stocks to find that coverage in social and news media are associated with markedly different patterns of subsequent return volatility and trading volume: high social media buzz around a given stock predicts a statistically significant increase in idiosyncratic return volatility and trading activity over the following month; high news media buzz predicts a significant decrease in volatility and trading activity. Therefore, it appears that news media may be a leading indicator for social media, which is consistent with the view that social media contents are generated by repeating and discussing existing news (what they call an "echo chamber" effect). We contribute to this literature by performing an analysis of the predictability of the cross-section of international stock index returns using a low frequency that is more typical of the empirical asset pricing literature.

The paper is organized as follows. Section 2 briefly reviews the literature on behavioral asset pricing and, based on this survey, derives a sentiment-augmented asset pricing framework. Section 3 describes the data. In particular, it provides a detailed description of the Refinitiv-MarketPsych's sentiment indicators and showcases their features compared to existing sentiment proxies. In Section 4, we bridge our framework to asset pricing models and show that sentiment, as captured by our novel indicators, leads to sharply different excess returns by means of portfolio sorting. Section 5 benchmarks different sentiment-augmented linear factor models against the standard CAPM and more recent linear models using time-series regression. We differentiate between global, local, and total sentiment, as well as alternative approaches for sentiment factor construction. Section 6 applies the two-stage FMB method to estimate global sentiment risk premia from the cross-section of global equity indices. We also apply FMB methods over a longer time horizon to discriminate between sentiment and information theories of sentiment in developed and emerging markets as in Tetlock (2007). Section 7 reports the results of a variety of robustness checks. Section 8 concludes.

2 Is sentiment a priced risk factor?

2.1 Related literature

The Capital Asset Pricing Model (CAPM) proposed by Sharpe (1964), Lintner (1965) and Mossin (1966) relates the expected returns of an asset to its sensitivity (beta) to the market risk premium. The single-factor CAPM was subsequently extended by including additional systematic risk factors, represented by shocks to state variables correlated with the marginal utility of investors' wealth.[2] However, it is now well-known that (I)CAPM-style models tend to break down in abnormal times (i.e., during financial crises as well as in periods of massive overvaluation, often imputed to alleged bubbles) when asset prices significantly deviate from their intrinsic values (see, e.g., Daniel & Titman, 1997; Finter, Niessen-Ruenzi, & Ruenzi, 2012; Lakonishok, Shleifer, & Vishny, 1994). Keynes (1936) and Livermore (1940) had already emphasized that fluctuations in asset prices might also be due to the influence of investors' "animal spirits" like greed, fear, ignorance, and hope. These non-fundamental, arguably not completely rational factors, could move asset prices by massive amounts away from their fundamental, intrinsic valuations. Such a recurring pattern of irrationality has led to the detection of widely debated and hence investigated phenomena such as fads, bubbles, and panics. For instance, the prospect theory proposed by Kahneman and Tversky (1979) may provide a more accurate description of decision making compared to standard expected utility theory based on rational preferences (see, e.g., De Bondt, 1998). Because people do not purely base their decisions on rational expectations about security payoffs, but rather use heuristics to evaluate potential risks and losses, decision making cannot be disconnected from human sentiment (see, e.g., Dolan, 2002; Nofsinger, 2005; Peterson, 2007). For instance, seminal theory work by De Long, Shleifer, Summers, and Waldmann (1990) and more recently by Shu (2010) finds that sentiment affects equilibrium asset prices, and is thus able to magnify market fluctuations and cause excess volatility. Accordingly, there is now an increasing consensus

[2]See, e.g., Basu (1977), Banz (1981), Jaffe, Keim, and Westerfield (1989), Fama and French (1993, 2015), Jegadeesh and Titman (1993), and Carhart (1997).

that sentiment should be considered as an integral part of asset pricing theory (see, e.g., Brown & Cliff, 2005; Da, Engelberg, & Gao, 2015; Shefrin & Belotti, 2008).

Because of the growing awareness of the importance of irrational trading and pricing motives, there is an expanding literature exploring the role of news, social media, and sentiment in asset pricing. Tetlock (2007) suggests that the frequency of negative words in a famed Wall Street Journal column may be a proxy of the journalist's mood and that this has predictive power for stock returns. Ahern and Sosyura (2015) study the stock market impact of the accuracy of rumor articles concerning mergers and report that it has a significant impact on returns, even though investors overestimate the accuracy of the average rumor. Da, Engelberg, and Gao (2011) show that individual investors prefer stocks covered by attention-grabbing news and this would obviously be reflected by a risk premium in the equity cross-section. In fact, evidence in Engelberg and Parsons (2011) shows that investors trade stocks primarily based on narratives in newspaper articles, despite easy access to firms' press releases and analysts' reports. Da et al. (2015) use daily Google internet searches by households to construct an aggregate indicator of sentiment, the Financial and Economic Attitudes Revealed by Search (FEARS) indicator. They find that this measure predicts short-term return reversals, temporary volatility spikes, and mutual fund flows out of equity and into bond funds. They argue that search-based methods bear advantages compared to survey-based techniques: online news and social media data are available in real time and reveal rather than just inquire about attitudes whilst the incentive to answer surveys or questionnaires honestly and truthfully is often questionable. Moreover, survey methods are typically characterized by small sample sizes and limited participation rates.[3] Furthermore, Da et al. (2015) find that an increased occurence of the terms "recession" and "bankruptcy" in a search volume indicator (SVI) made public by GoogleTrends leads decline in the University of Michigan Consumer Sentiment Index (MCSI) one month later. In an earlier study, Da et al. (2011) associated the SVI changes with trading by less sophisticated individual investors, a finding that has

[3]For example, the Michigan Consumer Sentiment survey employed in Lemmon and Portniaguina (2006) is sent to only 500 households, and the Consumer Confidence Index to 5000 households. Schmeling (2009) uses consumer confidence levels within 18 countries as a measure of sentiment.

been confirmed by other researchers (see, e.g., Joseph, Babajide Wintoki, & Zhang, 2011). Ben-Rephael et al. (2017) propose a related (yet distinct) measure of institutional investor attention using the news searching and reading activity at Bloomberg terminals. They report that announcements accompanied by abnormal institutional attention experience larger returns (in absolute terms) and modest, subsequent price drift. When institutional investors fail to pay sufficient attention, prices initially underreact to information, resulting in a subsequent drift.

Besides search- and survey-based methods, indicators that are based on fundamentals have also often been treated as proxies for sentiment. For instance, Baker and Wurgler (2006) build a composite indicator (henceforth BW) using principal component analysis applied to a vector of fundamental variables.[4] Empirically, they show that depending on the value of this indicator at the beginning of a period, the subsequent returns of hard-to-value shares like small, young, or unprofitable stocks are high (low) in low- (high-) sentiment states. This finding is further supported by the theoretical model in Baker and Wurgler (2007) using a top-down approach which focuses on aggregate sentiment and traces its effects on market and individual stock returns back to two central forces emphasized by modern behavioral finance: sentiment and limits to arbitrage. The concept is further developed and validated in international markets outside the U.S. by Baker et al. (2012). Extending this approach, Huang, Jiang, Tu, and Zhou (2015) have used partial least squares to eliminate the noise component in common sentiment proxies of BW as a methodology improvement. Brealey, Cooper, and Kaplanis (2014) show that the sentiment measured by the BW indicator and meant to proxy for the trading activity of arbitrageurs, predicts the reversion of share prices to their fundamental value, while retail sentiment, expressed by a naïve trend-following metric, has some short-term explanatory power for return momentum.[5]

[4]To isolate the common component of sentiment proxies from fundamental variables, the BW index is based on the common variation in six underlying proxies for sentiment: the closed-end fund discount, the NYSE share turnover, the number of and the average first-day returns on IPOs, the equity share out of new security issuance activity, and the dividend premium.

[5]Laborda and Olmo (2014) aggregate similar indicators to those in Baker and Wurgler (2006) to build a single market sentiment factor in order to predict the risk premium on U.S. sovereign bonds. The forecasting performance of such a sentiment index turns out to be time-varying and is generally stronger during recessions.

In an effort similar to ours, Siganos, Vagenas-Nanos, and Verwijmeren (2014) examined the relation between daily sentiment, as captured by Facebook's Gross National Happiness Index and returns on 20 international stock markets finding a contemporaneous, positive relationship; however, the fact that sentiment on Sundays affects stock returns on Mondays, suggests causality. We fully explore this pattern in our paper by estimating predictive linear pricing models that include sentiment factors. These results hold for different regions, languages, and religions. Moreover, these results are not solely driven by particular days on which Facebook's measure of sentiment reaches extremely high or low levels.

2.2 Theoretical foundations

Although the main contribution of this paper is empirical, we embed our analysis in a simple theoretical framework to rigorously pin down the relationship between investor sentiment and asset pricing and derive testable hypotheses. In this sub-section, following the arguments in Shefrin and Belotti (2008), we formulate the theoretical foundations for (i) why sentiment, if decomposed into positive and negative tones, should be treated as a risk factor in asset pricing models, and (ii) why such a factor is likely to be applicable to international equity index returns.

Shefrin and Belotti (2008) formally define sentiment as a proxy for distorted probabilities stemming from (i) deviations in aggregated investors' / market's beliefs from objective beliefs (see Johnson & Tversky, 1983; Loewenstein, 2000), and (ii) deviations of investors' time discount factor from an objective discount factor that would apply under the condition that all investors held correct beliefs (see Becker & Mulligan, 1997). Moreover, experimental psychological research has linked good (bad) moods with decreased (increased) risk aversion, as in the theoretical model by Shu (2010). This relation has been motivated by (at least) two behavioral explanations: (i) good mood raising people's "reference point" and bad mood lowering it, with loss aversion characterizing their preferences, and (ii) good mood making people behave cautiously, and bad mood

hastily ("mood maintenance theory").[6] As a result, sentiment is time-varying because market participants are not systematically excessively bullish or bearish but subject to a great variety of potential trajectories of human emotions. The emotions of individual investors aggregate into market sentiment as a collage of different investors' beliefs and time preferences. As a result, the risk premium for any security is the sum of a fundamental premium and a sentiment premium. When the sentiment premium is large relative to the fundamental component, risk premia will reflect both mispricing and compensations for bearing sentiment-based risk. However, if sentiment is zero, the risk premium is fully determined by the fundamental premium. Given that a single sentiment factor is not able to fully capture the oscillation of the sentiment function in the presence of heterogeneity in beliefs and time discount factors, we further empirically break-down the sentiment risk premium into negative and positive. By doing so, we aim at capturing the errors in probability estimation by different investors in the case of positive versus negative tail events.

The shape of the market sentiment function according to Shefrin and Belotti (2008) is affected by the wealth-weighted aggregated mixture of individual investors' sentiment indicators. We argue that such an aggregation best manifests in the sentiment for international equity indices and hence will affect global equity risk premia. In most other applications, sentiment may be diversified away and any mispricing easily eliminated by arbitrageurs who invest across different national stock markets. As stated by Kozak, Nagel, and Santosh (2018), arbitrageurs are able to neutralize components of sentiment-driven asset demand that are orthogonal to common factor covariances as long as they do not expose themselves to factor risk. Therefore, only the case of aggregate international stock markets maximizes the chances for sentiment-driven demand to have a substantial impact on expected returns. In the model by Kozak et al. (2018), they impose a "near-arbitrage" opportunity restriction and exclude high levels of leverage and unbounded short sales as implausible assumptions. Sentiment-driven investors may still construct strong tilts in their portfolios but the restrictions imposed prevent the most extreme cases. They

[6]For instance, Kliger and Levy (2003) use options data to recover risk preferences to show a negative correlation between mood (proxied by a weather variable, total cloud cover) and risk aversion.

argue that, because deviations from rational pricing must be caused by sentiment being either orthogonal to existing factor exposures or correlated with them, trading by the arbitrageurs largely eliminates the effects of the orthogonal components of sentiment-driven asset demand, while the effects that are correlated with common factor exposures may survive because arbitrageurs are not willing to accommodate these demands without a compensation for risk exposure. Kozak et al. (2018) do not extend their model into this direction but if heterogeneity in beliefs were to persist in the aggregate as maintained by Shefrin and Belotti (2008), arbitrageurs would be reluctant to trade and expose themselves to sentiment risk. We therefore postulate in the following that (orthogonal) sentiment should be part of asset pricing models and command a risk premium of its own. In order to test Kozak's model in the aggregate, we also validate a model in which sentiment affects the pricing of existing fundamental factors.

Despite our hypothesis on the breakdown of sentiment risk into two different sentiment premia (negative versus positive), we are aware that this approach still represents at best an approximation of the time-varying sentiment premium driven by fluctuating sentiment. Yet, this framework represents a more sophisticated specification than in the empirical behavioral finance literature because, thus far, empirical models have been limited to account for either high or low sentiment, mostly without incorporating any notion of valence, either positive or negative. In Section 3, we therefore describe how MarketPsych's sentiment data may capture these complex behavioral phenomena and how it can be used to support estimation and testing of sentiment-extended empirical asset pricing models. Furthermore, by explicitly dealing with the effects of valence, we are able to perform a direct comparison of risk factors based on MarketPsych's novel sentiment indicators with earlier, well-known sentiment measures used in the literature. Yet, while many sentiment indexes are typically only available for the United States (like the University of Michigan Consumer Sentiment Index) or for a small number of developed markets (the UBS/Gallup Index of Investor Optimism offers monthly sentiment levels for France, Germany, Italy, Spain, and the United Kingdom), in this paper we exploit MarketPsych's global coverage – similarly to Siganos et al. (2014) with reference to Facebook

happiness scores – to analyze the cross-section of international stock returns. Moreover, MarketPsych's proprietary sentiment is based on a comprehensive commercial-strength textual analysis of sources from news wires, internet news sources, and social media while prior studies that focus on textual analysis of media contents almost exclusively relied on a single source, e.g., the columns from the Wall Street Journal and the New York Times from Tetlock (2007) and García (2013), respectively, or the analysis of articles published in seekingalpha.com by Chen et al. (2014).[7]

3 Data

In this section, we introduce MarketPsych's sentiment indicators and discuss their differences with respect to the classical proxies like Baker and Wurgler (2006, 2007) and Da et al. (2015) or measures like RavenPack. We highlight that their international coverage of news and social media channels allows us to study the sentiment effect globally across developed and emerging equity markets. We provide insights into how the superior human language processing system supports a more sophisticated analysis to associate our empirical findings to the theoretical framework compared to a naïvely self-constructed sentiment proxy. This step is important as understanding the sources of departures from the earlier literature is crucial in interpreting our empirical findings.

3.1 Sentiment indicators

Our search-based sentiment index is the Refinitiv-MarketPsych sentiment indicator (RMI). The automatic language processing system from MarketPsych uses a human-made lexicon, which associates words and word groups to different kinds of indicators related to the performance of financial assets. Words and word groups in a message are annotated with so-called "Psych Words" (e.g., volatility, conflict, safety), defining a novel conceptual space.

[7] As argued by Sun et al. (2016) another clear advantage of using MarketPsych's TRMI sentiment buzz indicators is that these are available at a 1-minute frequency and therefore lend themselves to high-frequency predictability analysis. We shall not pursue this avenue of research applied to international stock returns, due to their relative illiquidity compared to Sun et al.'s application to U.S. traded assets and indices.

To define groups of words and create relationships, the lexical distance is assessed by applying weights on a scale from 0.0 to 1.0 to account for proximity in the text, while punctuation and additional structures are also taken into account. This process results in tuples, which are then recorded and aggregated as sentiment indicators. The scores are again divided by the total of the scores for all psych categories, called the $Buzz$, i.e., the weight of all messages and phrases of interest over a certain period. This ratio gives an indication of how important (or commonly discussed) a subject was over a given time interval and we use it to differentiate between low and abnormally high sentiment. This normalization allows equally weighted comparisons among numerous topics and nouns. Compared to other sentiment providers like RavenPack (see, e.g., Audrino, Sigrist, & Ballinari, 2020; Y. Shi, Ho, & Liu, 2016), MarketPsych indicators are not pre-calibrated to fit financial market prices and events using a training sample. Hence, we are able to use back-fitted time-series without any concern for the existence of hindsight biases. Because of this construction method, MarketPsych's approach goes far beyond the often used bag-of-words or other similar techniques applied in previous studies (see, e.g., Jiang, Lee, Martin, & Zhou, 2019; Tetlock, 2007).[8]

In our empirical analysis, we use aggregated RMI investor sentiment for international equity indices from both developed and emerging markets for the period 1998-2017. The MarketPsych sentiment indicator captures the net negative versus positive references in traditional and social media concerning a given equity market. It can be interpreted as an overall market sentiment proxy, void of any insights on the fundamental reasons for why references to a security may be positive or negative. Weighted by $Buzz$, we refer to abnormally high-positive or high-negative sentiment. Since the level of sentiment shows trend behavior and is non-stationary, we compute the change to the long-term mean since inception ("sentiment changes"). Such an average is estimated and updated on a recursive basis to avoid any hindsight biases, i.e., the mean is computed since inception of the sentiment index until t, not the entire period T. As a result, the change in sentiment is

[8]For a detailed description and examples of the RMI sentiment indicators see Appendix B.

comparable across indices as we account for local differences of sentiment perception.[9] This approach is advantageous over a weekly change since it allows us to consider time-varying perception of sentiment and thus better reflects the time-oscillation of sentiment as postulated in Section 2.2. The interested reader is referred to Section 3.2 and Appendix B for further details about the construction of the RMI sentiment indicators.

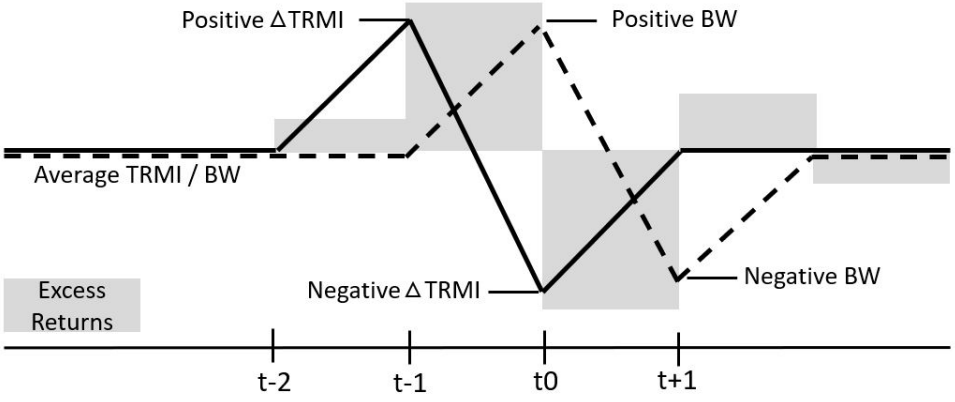
3.2 Comparison of sentiment indicators

The current academic standard on the topic of sentiment indicators, Baker and Wurgler's (2006) index (henceforth BW), extracts sentiment from fundamental variables that reflect trading volumes, stock issuance activity, and hence, directly or indirectly also asset prices. It therefore only captures a specific type of sentiment, namely the one that emerges after market participants have taken trading or investment actions as reflected by prices and trades. For instance, Baker and Stein (2004) suggest that turnover and liquidity are proxies for investor sentiment. In a market subject to short-sales constraints, retail investors participate only when they are optimistic, and thereby add liquidity to the market. Hence, high liquidity can also be seen as an indicator of overvalued stock prices. The BW indicator captures such an above-average liquidity, maps it to the overvaluation of stocks, and refers to it as a contribution to positive sentiment. Different from the traditional view, DeVault et al. (2019) find that these commonly used measures of investor sentiment capture the demand shocks of institutional rather than individual investors. Nofsinger (2005) confirms that emotions and moods have a severe impact on financial decision-making. Due to its nature as an emotional barometer, the stock market itself can be interpreted as an indicator of social mood. This idea of sentiment being driven by negative market fluctuations is also captured by the downside risk model from Lettau et al. (2014), which is motivated by Ang, Chen, and Xing (2006), who argue that investors who place higher weight on downside risk, demand additional compensation for holding stocks

[9]This approach is backed by the literature and relates changes in sentiment to demand shocks. DeVault, Sias, and Starks (2019) identify whether trades explained by sentiment metrics are, in the aggregate, initiated by individual or institutional investors. Such a classification exploits the fact that changes in sentiment will be positively related to changes in sentiment traders' demand (i.e., demand shocks) in the case of speculative stocks and inversely related to demand shocks in the case of "safe stocks".

Figure 1: Schematic difference between Refinitiv-MarketPsych and Baker-Wurgler

This figure visualizes the differences between the Refinitiv MarketPsych (RMI) and the Baker and Wurgler (BW) indices, showing that RMI leads BW by construction. To ease the comparison, we initialize both the RMI (solid line) and the BW (dashed line) indices at zero at $t-2$. A sentiment signal for RMI is measured by its deviation from the long-term mean, while we use the level of the BW as in Baker and Wurgler (2007). The grey bars show excess returns on an index at a point in time t.



with high sensitivities to downward market price shifts. Downside risk is defined here as the market return being one standard deviation below the mean. However, Nofsinger (2005) states that business activity, including trading and investment, tends to follow, rather than lead, social mood.

On the other side, the index we use in this paper extracts sentiment directly from news and social media posts, which are expected to at least partially anticipate the actions by investors. Kahneman and Tversky (1979), Damasio (1994), and Dolan (2002) investigate how emotions affect parts of the human brain and influence the decision-making process. Research by Peterson (2007) exploits advanced neuroimaging techniques to obtain information about psychological processes in the human brain and their connection to financial decisions. Peterson's work offers the foundations to MarketPsych's indicators. The academic literature, however, just stands at the beginning of exploring this novel data with very few papers currently published or in progress at the time of writing (see, e.g., Audrino & Tetereva, 2019; Changsheng & Yongfeng, 2012; Crone & Koeppel, 2014; Daszynska-Zygadlo, Szpulak, & Szyszka, 2014; Jiao et al., 2020; Michaelides, Milidonis, Nishiotis, & Papakyriakou, 2015; Sun et al., 2016).

In the stylized model in Figure 1, we assume that BW (dashed line) and RMI (solid line) are simply initialized at zero in $t-2$. There is no sentiment-driven signal in the market, and consequently, also no significant excess returns (measured by the grey bars) in the subsequent period, $t-1$. However, in $t-1$ a positive shock affects RMI, while BW remains unaffected. For instance, imagine the case in which investors become optimistic about the general economic outlook and enter the market according to the theoretical foundations discussed in Baker and Stein (2004). Their actions drive excess equity returns up in t. We argue that our sentiment index is able to detect such a surge in positive sentiment in the period before retail investors enter the market and increase liquidity as well as share prices. In contrast to the BW indicator that reflects the sentiment shock only later, a positive sentiment shock to RMI is associated with an immediate price increase and a prediction of positive equity market returns. At time t fundamental factors such as liquidity and volume, reflect trading activities in the previous period and are interpreted by BW as positive sentiment. However, we argue that on many occasions, investors may perceive these very dynamics in liquidity, volume, and prices just as a manifestation of ongoing overvaluation of the market at the time when this is disclosed in the news and social media. It may therefore even be recorded as a negative RMI mood shock, referring to a more pessimistic outlook. RMI would capture this turn of events as negative sentiment and we would observe a decline in prices derived from negative excess returns in period $t+1$. In this case, BW would record the dynamics in observable trading activities with a delay and characterize these as a downturn in sentiment. Therefore, true but unobserved sentiment carries a positive relationship with equity market returns as the measurement gap between true sentiment dynamics and realized return increases.

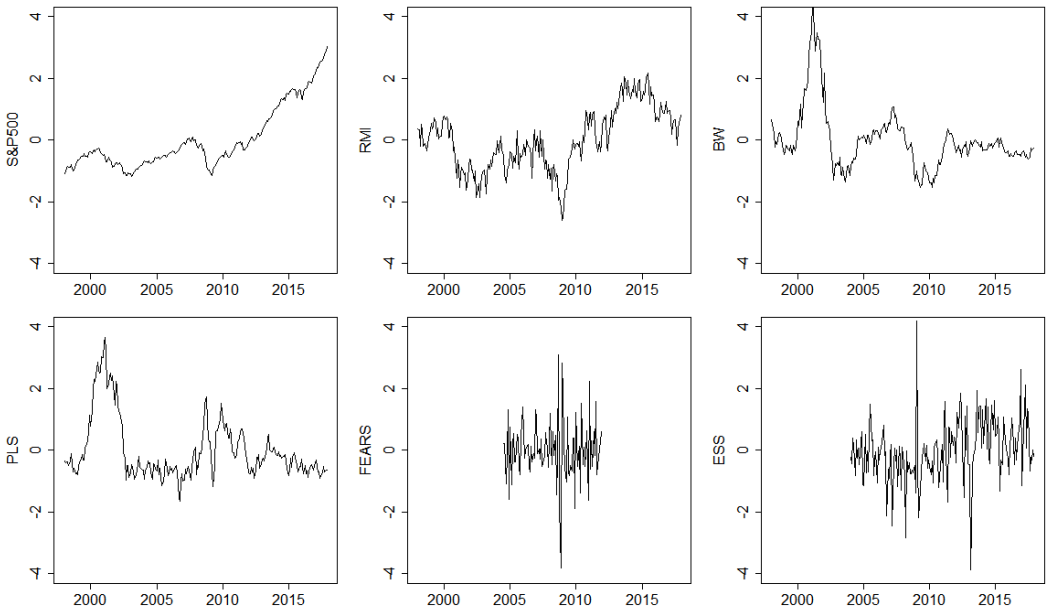
To provide support to these conjectures on the different dynamics of search-based versus market outcome-based sentiment measures, we compute the lead-lag correlation between BW and RMI when the latter is sampled at a monthly frequency. Of course, we are aware of the potential loss of information that this causes to a higher-frequency indicator such as RMI. We find that the BW and RMI U.S. equity market sentiment indices carry a significantly negative correlation of -0.13. However, when we lag the

variables according to the conceptual framework in Figure 1, the correlation switches to positive +0.11. When we increase the lag between the two series, the correlation climbs even higher until the lead-lag difference is increased up to six months. It then remains stable at a highly significant +0.22 and starts declining towards zero when the lead-lag differential exceeds ten months. Such a dynamic cross-serial relationship varies over time and we can identify three distinct phases. The first phase spans the sub-sample from 1998 to 2002 and is characterized by a high and significant negative contemporaneous correlation between sentiment indicators equal to -0.37. However, when the RMI index is lagged, the correlation is +0.20 and consistent with the full sample statistic reported earlier. In the second phase, between 2003 and 2011, the estimated correlations turn positive, at +0.15 and +0.24, respectively, for the contemporaneous and lagged series. The strength of the linear association declines in the third phase, between 2012 and 2017, when the correlations are +0.08 and +0.10 only, without and with lags.

Aside from the BW index, we benchmark our RMI index against a wide range of sentiment proxies. Figure 2 contrasts various sentiment indicators for the U.S. market and the S&P 500. Another fundamental-based index is the improvement of the BW index by Huang et al. (2015) using partial least squares (PLS) to eliminate the common noise component of multiple sentiment proxies and smooth the BW index. Both indicators show extremely high sentiment during the dot-com crisis in the early 2000s but deviate significantly in the aftermath of the financial crisis 2007-08. We use the downside risk model from Lettau et al. (2014) as a sentiment barometer driven by negative market fluctuations. In contrast to those fundamental or market-based sentiment proxies, Da et al. (2015) use the daily internet search volume from millions of households in the U.S. related to common sources of concern (e.g. recession, unemployment, bankruptcy) to construct a bottom-up sentiment measure called the Financial and Economic Attitudes Revealed by Search (FEARS) index. The data availability is very limited in time (2004 to 2011) and coverage (U.S. only), showing high volatility. FEARS was one of the first search-based sentiment measures in the literature and uses a conceptually similar, but less complex approach to MarketPsych's. The closest benchmark to RMI is the country-based

Figure 2: Refinitiv-MarketPsych and Baker-Wurgler indicators for the U.S.

This figure plots the monthly time series of the S&P 500 index, the Refinitiv MarketPsych (RMI), the Baker and Wurgler (BW), the BW improvement by Huang (PLS), the FEARS and the RavenPack ESS sentiment indices for the U.S. stock market. The data are re-scaled to the same base for better visualization.



event sentiment score from RavenPack News Analytics[10], also because it is based on a proprietary human language processing system and available for international markets. The time-series plot suggests a trend behavior but with much higher variability than RMI.

3.3 Global equity markets data

We analyze weekly excess returns for global equity markets over a sample January 1998 - December 2017.[11] We use a Wednesday to Wednesday time window in order to avoid any day-of-the-week effect. We select the one-month U.S. Treasury Bill yield as the risk-free rate from the publicly available data repository maintained by Fama and French (1993, 2015). This choice is appropriate because all indices are expressed in U.S. dollars. As a

[10]We selected the country indices due to our international context with topics coverage for business, the economy, the environment, politics, and society as they might impact the countries' stock market indices at large. Following the research guidelines by RavenPack, we filter the scores by applying a minimum relevance score of 75 for being significant. To avoid stale news we also limit our analysis to data on the event novelty scores (ENS).

[11]See Table A.1 in the Appendix for the list of equity indices.

proxy for the unobservable market portfolio return, we use the excess return on the MSCI World performance index from the Fama-French data repository.[12]

4 Sentiment in the cross-section of portfolio sorts

It is well known that empirical tests of standard asset pricing models based on traditional risk factors generally fail to explain price deviations from the intrinsic value of assets (see, e.g., Ferson & Korajczyk, 2002). One source of these mispricings may be traced back to the existence of irrational components in investors' beliefs. If our sentiment indicator represents an aggregate measure of investors' beliefs on an asset, we expect that the fit of otherwise traditional asset pricing models may improve when a new sentiment factor is added to the empirical framework. In this section, we show that the RMI sentiment indicator is priced in the cross-section of international equity portfolio sorts. We proceed to form portfolios based on the exposure of the equity indices to deviations of sentiment from its long-term mean. This is a common approach (see, e.g., Borochin & Zhao, 2015; Fama & French, 2015; Fama & MacBeth, 1973; Lettau et al., 2014) to show that the portfolios created in this way yield significantly positive/negative average excess returns. We differentiate between developed and emerging markets, as well as a global sort that comprises both categories.

More precisely, we sort our assets into sentiment-based portfolios: on a weekly basis, we rank the international equity indices on the basis of the corresponding, aggregate market sentiment measure. Psychological and cultural traits suggest that sentiment has no common definition and as such it may not be comparable in terms of its level across markets even when measured on a common scale of values. Inconsistencies may also occur

[12] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The MSCI World index is a capital-weighted total return index that includes the largest companies from all developed markets. The constituents list overlaps greatly with the indices for which RMI sentiment indicators are available, i.e., Australia, Canada, France, Germany, Hong Kong, Japan, Russia, Spain, Switzerland, the UK, and the US. The following countries are instead represented in the MSCI World index but are simply consolidated as a Eurozone overall index in RMI indices: Austria, Belgium, Denmark, Finland, Ireland, the Netherlands, Norway, Portugal, and Sweden. RMI indicators are also available for Brazil, China, India, and Russia but not for Israel and New Zealand, two further constituents of the MSCI World index. In light of such a considerable overlap, we consider the MSCI World index an appropriate proxy for the market portfolio in our application.

because MarketPsych is only able to evaluate English-written content. Thus, we sort the indices at every time t according to the deviation of sentiment from its historical, expanding window mean from inception up to time t. We apply three alternative sorts based on positive, negative, and neutral sentiment. For each week t, we determine whether an equity index belongs to the lower $s(-)$, middle $s(0)$, or upper $s(+)$ quantile of equity indices sorted based on previous week's sentiment. The difference between the positive $s(+)$ minus negative $s(-)$ sentiment-mimicking portfolio is our Fama-French like single sentiment factor, denoted as PMNSNT. This is either constructed as a global, emerging or developed market factor depending on the index universe. Alternatively, we use the positive $s(+)$ and negative $s(-)$ sentiment-mimicking portfolio directly as two disjoint sentiment factors for a better representation of the human trajectory of emotions.

Following DeVault et al. (2019), our primary objective is to examine correlation patterns rather than testing for causal effects. However, we acknowledge the existence of dynamic relationships between financial returns and sentiment by using one-week lagged sentiment to avoid a potential reverse causality problem in the sense that it might be returns leading to sentiment shifts and not the reverse. In Panel A of Table 1, we present the results of Welch's two-sample t-test of *equality of average weekly excess percentage returns* across these sorts using both developed and emerging markets to form a global perspective (Welch, 1947, see). The first column gives the average weekly percentage excess returns sorted in accordance to the average weekly sentiment reported in the third column. Sort $s(-)$ implies a small, negative mean excess return of -0.01%, while $s(0)$ leads to a moderate positive weekly mean excess return of 0.10%, or 5.36% per year. As such, this portfolio has a similar return as the MSCI World market index with 6.57%. Portfolio $s(+)$ implies a significantly positive, large mean excess return of 0.31%, or 15.90% per year. The sentiment level of sort $s(-)$ is deeply negative with -8.11%, practically zero for $s(0)$ and strongly positive at 7.03% in the case of $s(+)$. The standard deviation of sentiment across sorts, however, does not differ much and lies between 3.64% ($s(0)$) and 5.01% ($s(+)$). The weekly standard deviation of excess equity returns is the highest with 2.30% for $s(-)$, compared to 1.97% and 2.07% for $s(0)$ and $s(+)$, respectively. As a result,

positive deviations of sentiment from its long-term mean are associated with high, positive mean excess returns. The standard deviation of excess returns allows us to conclude, however, that returns are highly volatile within each sort. Columns 5 and 6 report the p-values associated with Welch t tests, indicating that the differences in mean excess return across sorts are significantly different from zero. Conceptually, negative sentiment is associated with negative excess returns. Even though this makes intuitive sense, this is at odds with earlier findings by Baker and Wurgler (2006, 2007) based on the BW index, by which high sentiment predicts low returns in the cross-section. However, as we have argued in Section 3.1, BW's proxy measures sentiment when this has already been reflected in equilibrium stock prices and trades by conscious decisions, whereas our RMI indicator captures emotions that investors consciously or unconsciously express through their news and social media activity and that are likely to precede their trading decisions. In addition, the BW index is a pure representation of the sentiment for the U.S. while we construct global sentiment-mimicking portfolios. In comparison to Panels B and C, we see that the global sorts are driven by the behavior of the (dominating) developed countries. For those countries negative sentiment is associated with even more negative returns of -0.0548% compared to a positive return of 0.2659% for emerging markets only. Positive sentiment for the latter returns 0.5354% on a weekly average. Thus, we report significantly different behavior of sentiment between developed and emerging countries, which has important implications for the long-term impacts discussed in Section 6.2.

5 Sentiment-augmented asset pricing models

First, we use the global sentiment-based mimicking portfolios to enrich standard asset pricing models like the CAPM, to demonstrate that sentiment significantly increases the explained variation in excess returns and may represent a priced risk factor. In addition, we account for the Fama and French's five-factor model and a momentum factor to empirically estimate the additional contribution of our new sentiment factor. Second, we benchmark our results against both well-established and novel sentiment indicators.

Table 1: Portfolios sorted by sentiment

This table shows the descriptive statistics of the portfolio sorts and the results of the Welch two-sample t-test of equality of the average weekly excess percentage returns across three sorts. For each time interval $[t, t-1]$, the asset is either in the lower $s(-)$, middle $s(0)$, or upper $s(+)$ quartile of stock indices sorted based on the previous week's sentiment change. Panel A shows global sentiment sorts, Panel B developed makets sorts and Panel C emerging countries sorts. The table reports the average weekly excess percentage return (Avg.) and its standard deviation (Std.) for each sort along with the corresponding average sentiment level and the related standard deviation. Columns 5 and 6 report the results of a Welch's two-sample t-test of the null hypothesis in which the difference in means between two sorts is zero. ***, **, and * indicate significance at the 0.1%, 1%, and 5% level, respectively. The sample period is from January 1998 to December 2017 with weekly portfolio rebalancing.

	Return		Sentiment		$s(0)$	$s(+)$
	Avg.	Std.	Avg.	Std.		
Panel A: Developed and emerging equity markets						
$s(-)$	-0.0112	2.2954	-8.1142	3.9440	0.2228	0.0010***
$s(0)$	0.1030	1.9673	-0.7092	3.6356		0.0221**
$s(+)$	0.3058	2.0741	7.0270	5.0066		
Panel B: Developed equity markets						
$s(-)$	-0.0548	2.1464	-7.2563	4.2384	0.0966*	0.0019***
$s(0)$	0.0948	1.9575	-0.8499	3.9155		0.1229
$s(+)$	0.2290	2.0115	6.7306	5.0299		
Panel C: Emerging equity markets						
$s(-)$	0.2659	3.5239	-10.8669	6.3472	0.5595	0.1032
$s(0)$	0.1702	2.7580	-0.4284	4.4126		0.0115**
$s(+)$	0.5354	2.8112	8.6372	6.3659		

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Third, we separate local and global sentiment and provide alternative specifications for the global sentiment variables using principal component analysis (PCA).

5.1 Linear factor models including sentiment

We investigate whether the global sentiment-based mimicking portfolios can be used to enhance traditional asset pricing models including the capital asset pricing model (CAPM), a Fama-French like sentiment-augmented CAPM model that includes a positive-minus-negative sentiment portfolio (PMNSNT-CAPM), and a CAPM-like model for positive and negative sentiment deviations (SNT-CAPM).[13] We estimate linear models for international equity index returns, and compare their coefficients and their goodness-of-fit to the standard CAPM and Fama-French five-factor model (FF5). In the following, all standard errors of the estimates are adjusted to account for time-series correlation and heteroscedasticity by using Newey-West corrections. Given the multiple testing set-up, we also correct the p-values for multiple testing bias and deal with the problem of multiple comparisons across markets by applying the Holm-Bonferroni correction proposed by Sture Holm (1979).

The first model, the traditional CAPM, projects the excess returns of each index on the excess market return of the MSCI World:

$$r^e_{i,t} = \alpha_i + \beta_{CAPM,i} r^e_{m,t} + \epsilon_{i,t}, \text{for all } t \in T, \tag{1}$$

where $\beta_{CAPM,i}$ is the standard CAPM beta for index i and $r^e_{m,t}$ is the market excess return.

As for the second model, we construct a single global sentiment risk factor as the excess return on a portfolio of long-positive/short-negative sentiment-sensitive indices from both developed and emerging markets. We test whether sentiment represents an

[13] In Section 7, we also control for the global five factors proposed in Fama and French (2015) and include a momentum factor to address concerns that our sentiment indicators may capture news that are already incorporated in traditional risk factors or may just reflect momentum.

additional, priced risk factor. The estimated model is

$$r^e_{i,t} = \beta_{CAPM,i} r^e_{m,t} + \beta_{PMNSNT,i} r^e_{PMNSNT,t} + \epsilon_{i,t}, \tag{2}$$

where the benchmark CAPM is nested under the restriction $\beta_{PMNSNT,i} = 0$. $\beta_{PMNSNT,i}$ is the beta on the excess return r^e_{PMNSNT} of a long-positive/short-negative sentiment portfolio formed by the difference between the first and third sentiment-ranked portfolios defined above. To avoid multicollinearity with the market risk premium and to contrast our findings from Kozak et al. (2018) in which sentiment only remains if correlated to fundamental factor exposure, we orthogonalize this factor in the same manner described in the following for the extended SNT-CAPM.

For the third model, we split the long-positive/short-negative portfolio and use the excess returns of the global sorts directly. We estimate the sensitivity of assets to the portfolio returns mimicking the reaction of international equity markets to positive and negative changes in investor sentiment. According to our hypothesis in Section 2.2, we allow for a more complex (composite) hypothesis by assuming that different "signs" of sentiment fluctuations may be priced differently, so that there is no single global sentiment risk factor. The model is specified as

$$r^e_{i,t} = \beta_{CAPM,i} r^e_{m,t} + \beta_{s(+),i} r^e_{s(+),t} + \beta_{s(-),i} r^e_{s(-),t} + \epsilon_{i,t}, \tag{3}$$

where $\beta_{CAPM,i}$ is the standard CAPM beta for asset i, $r^e_{m,t}$ is the market excess return, $\beta_{s(-),i}$ and $\beta_{s(+),i}$ are the betas on the excess returns of the positive and negative sentiment portfolios, respectively.

Because our portfolio sorts use the weekly returns of the constituents, one may pose concerns regarding multicollinearity and argue that we simply construct an alternatively weighted market index that is highly correlated tp the MSCI World index. The correlation analysis in Panel A of Table A.2 in the Appendix indeed shows the existence of highly significant correlations between the sentiment and fundamental factors. Interesting to note is the high positive correlation of 0.73 between positive and negative sentiment. This

indicates that the returns of both portfolios tend to move in the same direction, while the positive sentiment-mimicking portfolio outperforms the negative. The correlation with market returns is highly significant for both positive and negative sentiment, yielding estimates in excess of 0.5, while other factors significantly correlate between -0.35 (CMA) and 0.12 (SMB) with positive, and between -0.34 (CMA) and -0.07 (HML) with negative sentiment. In order to address such a potential multicollinearity, we orthogonalize the variables following a stepwise regression approach and use the residuals as the final portfolio return is estimated. Given the results in Section 4, positive abnormal sentiment is likely to contain most information, followed by negative abnormal sentiment. This finding dictates the sequence in our orthogonalization approach which leads to insignificant, (near-) zero relationships between the sentiment and fundamental variables.[14] Details about the orthogonalization approach can be found in Appendix C.

While this approach yields unbiased coefficient estimates, it often complicates their economic interpretation. For models containing sentiment risk factors, we also provide information on the relative importance of each coefficient by presenting the estimated relative importance index (RI). RI allows the estimation of the relative contribution of each regressor to the total explained variation and proceeds to decompose the overall coefficient of determination to estimate the contribution of each risk factor. We follow the methodology in Sen, Lindeman, Merenda, and Gold (1981) and report the relative contribution to the R^2 to represent such a characterization.

Panel A of Table 2 shows that for all equity indices, the estimated CAPM coefficients are highly significant at the 1% level. The intercepts α are small and mostly insignificant. If an asset pricing model is able to completely capture the variation in expected excess returns, the intercept should be close to zero. Given that we use well-diversified equity indices this makes intuitive sense. One can debate whether these results justify adding another risk factor; however, as the explained variation by the CAPM is not very high and we are interested in the relative and not absolute performance of sentiment-augmented

[14] We acknowledge that such orthogonalization steps may remain somewhat arbitrary. In unreported results, we try various permutations of the orthogonalization sequence. The main results hold as long as the resulting variables are uncorrelated to each other.

models compared to benchmarks, we still perceive a clear justification for adding sentiment to our model.[15] The R^2 ranges from a maximum of 53.3% for the S&P 500 (US500), the main contributor to the MSCI World, to a moderate 36.5% for the German DAX (DE30), to a minimum of 2.0% for the Chinese index (CN300).

In Panel B, we extend the CAPM to include a single global sentiment risk factor based on a portfolio of long-positive/short-negative sentiment indices (what we call the PMNSNT-CAPM). The results show that sentiment exposure is occasionally estimated with precision; when it is, the contribution of sentiment to the overall explained variation hardly exceeds 5%. The exception is China's CN300 index, which shows a high relative contribution of 33.7% despite the resulting R^2 is a negligible 3.0%. We conjecture that this first global sentiment indicator may lack the power to capture domestic sentiment in aggregated equity indices. The absolute values of the estimated intercepts are equal to those estimated in the CAPM model, i.e., our sentiment-augmented model with a single risk factor based on a long-positive/short-negative sentiment-sorting of equity indices fails to capture any additional variation in excess returns.

In Panel C, we estimate the time series regressions using the two distinct uncorrelated sentiment factors represented by global mimicking portfolios of equities with positive and negative sentiment (SNT-CAPM). The two sentiment factors are mostly significant at the 1% level (excluding negative sentiment for the Chinese index CN300). The absolute alphas are smaller or equal versus the CAPM, indicating that the sentiment-augmented model is able to capture variation in excess returns left unexplained by the traditional model. The intercepts are all insignificant (except for the Indian IN50 and the Brazilian BR50 index) and indistinguishable from zero as required by well-specified asset-pricing models (see, e.g., Fama & French, 1993; Merton, 1973). The R^2 increases substantially for all indices reaching 73.9% in the case of the U.S. market. We plot the total explained variation benchmarked against the CAPM in Figure 4 and visualize the improved model fit when the two distinct sentiment factors are added. The estimated betas of all sentiment

[15]The low absolute and insignificant intercepts make more sophisticated methods like GRS tests proposed in Gibbons, Ross, and Shanken (1989) redundant and we shall limit our discussion to the estimates of absolute α.

variables are generally positive. These coefficients can be interpreted as the sensitivity of the equity indices to uncorrelated portfolios of positive and negative sentiment assets conditioning out the effects of the market risk factor according to our orthogonalization procedure. The relative importance scores emphasize that positive sentiment provides the highest relative contribution, after market risk.[16]

Table 2 also reveals that sentiment seems more relevant in terms of relative importance in emerging than in developed markets, which are known to be (more) efficient in overall terms (see, e.g., Griffin, Kelly, & Nardari, 2010). In order to support this claim, we apply two tests: a simple comparison of the CAPM-implied R^2 coefficients and the variance ratio test proposed by Lo and MacKinlay (1988). The R^2 of the market model is often seen as a naïve metric for stock price informational efficiency.[17] If we compute the correlation between the relative importance of sentiment over the market risk premium and the R^2 of the CAPM model, we obtain a strongly negative and highly significant estimate of -0.72: the higher the R^2 from the CAPM, i.e., the more efficient the market, the less important the sentiment factor-mimicking portfolios.[18] The second, more sophisticated approach to the measurement of market efficiency employs the variance ratio test under the null hypothesis of a random walk with homoskedastic ($M1$) or heteroskedastic increments ($M2$).[19] A high value of the variance ratio statistic leads to a rejection of the null of market informational efficiency. We apply variance ratio tests at lags $k = 2, 5, 10$ as suggested in Morck et al. (2000) and Bramante et al. (2013a, 2013b). Next, we compute the correlations between the aggregated, relative importance of all sentiment variables on top of the market risk premium against all sample values of the variance ratio metrics $M1(k)$ and $M2(k)$ at different lags. All correlations are positive and peak at a value of 0.87 for both statistics at lag 10 ($M1(10)$ / $M2(10)$). This correlation is significant at the 1% level. The results support our conjecture that sentiment risk matters more for

[16]This is of course partially due to the sequential orthogonalization with positive abnormal sentiment being the primary factor.

[17]See Bramante, Petrella, and Zappa (2013a, 2013b); Morck, Yeung, and Yu (2000) for comprehensive studies of the use of R^2 as a price efficiency measure.

[18]Due to the non-normality of our data we report Spearman's rank correlation coefficients.

[19]Under the null hypothesis, the associated test statistic has an asymptotic standard normal distribution with finite variance for all the time series.

Table 2: Comparing the CAPM to alternative sentiment-augmented CAPM models

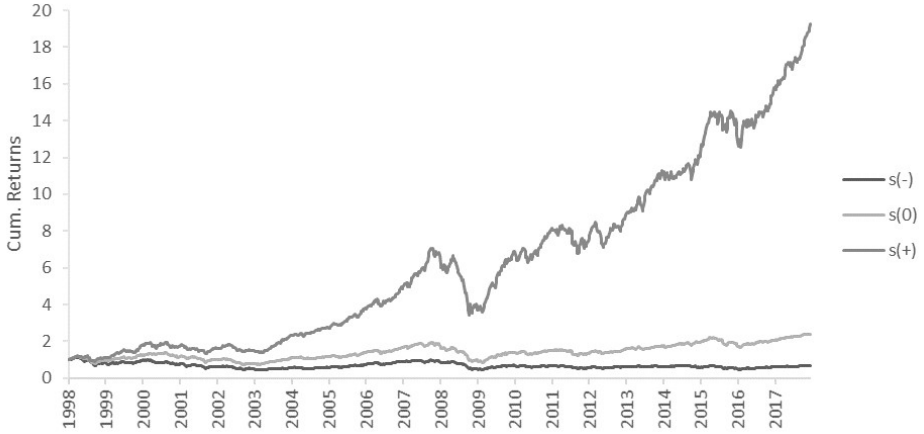
This table compares the standard CAPM to a number of sentiment-augmented linear factor models. Panel A concerns results for the standard CAPM based on the market risk premium (MRP) as its single priced factor. Panel B is the extension of the CAPM to include the return of a long-positive/short-negative sentiment portfolio as a single sentiment risk factor (PMNSNT-CAPM). Panel C uses two separate sentiment risk factors based on the returns of portfolios mimicking negative and positive sentiment (SNT-CAPM). The table reports the coefficient estimates, the Newey-West corrected standard errors to adjust for autocorrelation and heteroscedasticity in returns in parentheses and the adjusted R^2. In the case of models including sentiment, we also report the relative importance indicator for each coefficient, which estimates its relative contribution to the total explained variation. The regressions are estimated for each equity index individually on weekly returns from 1998 to 2017.

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel A: CAPM															
Intercept	0.0002	0.0015*	0.0005	0.0001	0.0023*	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017**	0.0002	0.0009	0.0000	0.0004
	(0.0004)	(0.0009)	(0.0005)	(0.0005)	(0.0019)	(0.0006)	(0.0006)	(0.0005)	(0.0003)	(0.0008)	(0.0009)	(0.0008)	(0.0019)	(0.0007)	(0.0003)
MRP	0.3206***	0.7534***	0.5059***	0.5074***	0.2060***	0.6743***	0.6076***	0.6369***	0.5159***	0.5197***	0.4208***	0.4664***	0.8088***	0.4012***	0.5996***
	(0.0231)	(0.0509)	(0.0307)	(0.0430)	(0.0688)	(0.0352)	(0.0408)	(0.0303)	(0.0200)	(0.0416)	(0.0415)	(0.0394)	(0.0970)	(0.0385)	(0.0196)
Adj. R^2	20.23	33.43	36.44	32.12	2.03	36.52	30.58	39.41	39.87	18.81	13.06	16.71	14.60	18.45	53.25
Panel B: PMNSNT-CAPM															
Intercept	0.0002	0.0015*	0.0005	0.0001	0.0023*	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017*	0.0002	0.0009	0.0000	0.0004
	(0.0004)	(0.0009)	(0.0004)	(0.0005)	(0.0018)	(0.0006)	(0.0006)	(0.0005)	(0.0003)	(0.0008)	(0.0009)	(0.0008)	(0.0019)	(0.0007)	(0.0003)
MRP	0.3206***	0.7534***	0.5059***	0.5074***	0.2060***	0.6743***	0.6076***	0.6369***	0.5159***	0.5197***	0.4208***	0.4664***	0.8088***	0.4012***	0.5996***
	(0.0232)	(0.0521)	(0.0297)	(0.0369)	(0.0700)	(0.0339)	(0.0389)	(0.0296)	(0.0199)	(0.0428)	(0.0435)	(0.0387)	(0.0955)	(0.0383)	(0.0189)
PMNSNT	0.0026	-0.1815**	0.0039	-0.1685***	0.2763**	-0.0942**	-0.1380***	-0.1048***	-0.0787**	0.0159	-0.0624	0.0126	-0.2646***	0.0069	-0.0224
	(0.0338)	(0.0990)	(0.0346)	(0.0414)	(0.1505)	(0.0581)	(0.0430)	(0.0445)	(0.0352)	(0.0745)	(0.0527)	(0.0560)	(0.1803)	(0.0566)	(0.0311)
Adj. R^2	20.16	33.96	36.38	33.79	2.99	36.81	31.28	39.87	40.27	18.74	13.12	16.64	15.29	18.36	53.25
RI-MRP	100.00	98.20	100.00	94.91	66.30	99.06	97.55	98.70	98.88	99.95	98.94	99.96	95.04	99.99	99.93
RI-PMNSNT	0.00	1.80	0.00	5.09	33.70	0.94	2.45	1.30	1.12	0.05	1.06	0.04	4.96	0.01	0.07
Panel C: SNT-CAPM															
Intercept	0.0002	0.0015**	0.0005	0.0001	0.0023	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017***	0.0002	0.0009	0.0000	0.0004
	(0.0003)	(0.0007)	(0.0004)	(0.0005)	(0.0016)	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0006)	(0.0007)	(0.0007)	(0.0015)	(0.0005)	(0.0003)
MRP	0.3206***	0.7534***	0.5059***	0.5074***	0.2060***	0.6743***	0.6076***	0.6369***	0.5159***	0.5197***	0.4208***	0.4664***	0.8088***	0.4012***	0.5996***
	(0.0189)	(0.0332)	(0.0195)	(0.0404)	(0.0683)	(0.0348)	(0.0345)	(0.0288)	(0.0167)	(0.0263)	(0.0349)	(0.0293)	(0.0637)	(0.0247)	(0.0175)
$s(+)$	0.5322***	0.9017***	0.5587***	0.4584***	0.6656***	0.7517***	0.7065***	0.6692***	0.5316***	0.9600***	0.7517***	0.7669***	1.4420***	0.7718***	0.4577***
	(0.0301)	(0.0412)	(0.0247)	(0.0304)	(0.1133)	(0.0378)	(0.0317)	(0.0346)	(0.0245)	(0.0360)	(0.0486)	(0.0465)	(0.0805)	(0.0350)	(0.0321)
$s(-)$	0.2443***	0.6864***	0.2552***	0.4091***	0.0488	0.4591***	0.4890***	0.4331***	0.3387***	0.4277***	0.4220***	0.3418***	0.9744***	0.4377***	0.2388***
	(0.0476)	(0.0788)	(0.0300)	(0.0526)	(0.1536)	(0.0702)	(0.0559)	(0.0583)	(0.0380)	(0.0710)	(0.0692)	(0.0710)	(0.1878)	(0.0528)	(0.0352)
Adj. R^2	55.98	65.02	64.86	55.30	13.26	68.67	61.70	71.06	70.44	59.69	41.65	45.45	49.31	60.77	73.90
RI-MRP	36.19	51.43	56.20	58.07	15.94	53.20	49.58	55.48	56.62	31.59	31.44	36.81	29.69	30.43	72.05
RI-$s(+)$	55.01	36.88	37.82	26.14	83.84	36.48	36.98	33.79	33.17	59.46	55.33	54.91	52.24	58.87	23.16
RI-$s(-)$	8.80	11.69	5.98	15.80	0.22	10.32	13.44	10.74	10.21	8.95	13.23	8.28	18.06	10.69	4.78

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Figure 3: Cumulative returns of global sentiment portfolio sorts

This figure displays the cumulative returns of the global sentiment portfolio sorts. For each time interval $[t, t-1]$, the asset is either in the lower $(s(-))$, middle $(s(0))$, or upper $(s(+))$ quartile of stock indices sorted based on the previous week's sentiment change. The index universe covers both developed and emerging markets. The sample period is from January 1998 to December 2017.



less efficient markets.[20] These findings are in line with Griffin et al. (2011) who report that the information environment, the news dissemination mechanism, the level of insider trading, and technological development are key criteria making hard-fact news generally more useful for explaining stock returns in developed markets than in emerging markets. In informationally efficient markets, stock prices preferentially reflect fundamentals rather than sentiment shocks. Because sentiment is associated with lower informational efficiency, the investors in inefficient markets are more vulnerable to behavioral biases, sentiment, and herding.

To address any concerns that our RMI indicators may fail to measure sentiment and instead just reflect market information contained in news that may be captured by more traditional variables, we also employ a Fama-French five-factor model (FF5) as in Fama and French (2015, 2017), augmented with our sentiment indicators. These concerns are grounded in the way sentiment is constructed as it includes references to fundamentals-related subjects like accounting results, earning expectations, and the general economic outlook. Another extension of our empirical model includes a momentum factor to address

[20] See Table D.1 in the Appendix for a robustness analysis of these findings based on correlation coefficients applied to alternative metrics and different lags.

a concern that our sentiment proxies may simply capture price trends, also in the light of the empirical evidence on the relationship between news coverage and momentum (see, e.g., Da, Engelberg, & Gao, 2012; Huynh & Smith, 2017).

We use the Fama-French factors for developed countries for our international context.[21] Note that the applicability of the FF5 model to aggregate international equity indices is debatable (see, e.g., Cakici, 2015) given that these assets are already generously diversified portfolios. On the one hand, any sensitivity to traditional company-specific attributes should have been already averaged out and removed at the aggregation stage. We do not expect that such a rich factor structure may over-turn our findings nor that the FF factors may offer a significant contribution to explaining cross-sectional returns. On the other hand, the equity indices themselves by construction reflect a strong selection bias because only the largest and most successful companies are constituents of the country factors distilled by Fama and French. Similarly to the methodology followed above, we orthogonalize all variables to remove any multicollinearity.[22] Finally, to address a concern that sentiment may just be a proxy for momentum, we augment the five-factor model by including the developed markets Fama-French momentum factor. We add this additional factor to estimate the ultimate SNT-MOM-FF5 model:

$$\begin{aligned} r_i^e = \beta_{CAPM,i} r_m^e + \beta_{SMB,i}\hat{\epsilon}_{SMB} + \beta_{HML,i}\hat{\epsilon}_{HML} \\ +\beta_{RMW,i}\hat{\epsilon}_{RMW} + \beta_{CMA,i}\hat{\epsilon}_{CMA} + \beta_{MOM,i}\hat{\epsilon}_{MOM} \\ +\beta_{s(+),i}\hat{\epsilon}_{s(+)} + \beta_{s(-),i}\hat{\epsilon}_{s(-)} + \epsilon_i, \end{aligned} \tag{4}$$

where $\beta_{CAPM,i}$ measures the sensitivity of the returns on asset i to the market factor, $\beta_{SMB,i}$ to the size factor, $\beta_{HML,i}$ to the value premium, $\beta_{RMW,i}$ to the profitability factor, $\beta_{CMA,i}$ to the investment factor, and $\beta_{MOM,i}$ to the momentum factor. The remaining terms refer to the orthogonalized sentiment indicators as previously defined.

[21] In August 2019, emerging markets portfolios were added to the research returns of Kenneth's homepage. The global portfolios and factors used in a previous version of this paper have been renamed to developed. We are limited to the developed versions due to missing daily data for emerging markets.

[22] We emphasize that we first orthogonalize the FF factors and then proceed with the sentiment factors to be able to disentangle any sentiment effects after any other (by now) classical FF5 factors have been considered. Table A.2 in the Appendix includes a correlation analysis between sentiment and the Fama and French factors before and after orthogonalization.

Table 3: Comparing Fama-French factors and momentum with sentiment-augmented multi-factor models

This table compares the Fama-French five-factor (FF5) model with different sentiment-augmented linear factor models. Panel A shows results for the traditional FF5 model. Panel B reports the FF5 model with an extension by the long-positive/short-negative sentiment portfolio (PMNSNT-FF5). Panel C extends the FF5 model with two sentiment risk factors based on the returns of portfolios mimicking negative and positive sentiment (SNT-FF5). Panel D incorporates the Fama-French momentum factor (SNT-MOM-FF5). The table shows the coefficient estimates, Newey-West corrected standard errors to adjust for autocorrelation and heteroscedasticity in returns in parentheses as well as the adjusted R^2. We also report the relative importance indicator for each coefficient, which estimates its relative contribution to the total explained variation. The regressions are estimated for each equity index individually on weekly returns from 1998 to 2017. The table is shortened to the important variables for better accessibility.

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel A: FF5															
Intercept	-0.0002	0.0012	0.0006	-0.0002	0.0025	0.0004	0.0003	0.0003	-0.0001	0.0006	0.0014	0.0001	0.0006	-0.0003	0.0004
	(0.0004)	(0.0009)	(0.0004)	(0.0006)	(0.0018)	(0.0006)	(0.0006)	(0.0005)	(0.0003)	(0.0008)	(0.0009)	(0.0008)	(0.0017)	(0.0007)	(0.0004)
MRP	0.3937***	0.7326***	0.4978***	0.5483***	0.1640	0.6419***	0.5764***	0.6355***	0.5602***	0.5023***	0.4948***	0.5002***	0.8221***	0.4264***	0.6165***
	(0.0274)	(0.0607)	(0.0352)	(0.0418)	(0.0833)	(0.0443)	(0.0411)	(0.0373)	(0.0261)	(0.0489)	(0.0473)	(0.0438)	(0.0822)	(0.0445)	(0.0327)
SMB	0.5537***	0.6409***	0.5780***	0.1308	0.3399	0.3622***	0.3714***	0.3098***	0.2447***	0.6950***	0.7773***	0.7393***	1.1904***	0.6294***	0.2193***
	(0.0684)	(0.1414)	(0.0626)	(0.0744)	(0.1722)	(0.0870)	(0.0895)	(0.0753)	(0.0646)	(0.1154)	(0.1127)	(0.1299)	(0.2094)	(0.1085)	(0.0708)
HML	0.1141	0.2755	0.0983	0.0812	0.0411	0.1278	0.2830***	0.1559	0.0552	0.2673**	-0.0437	0.1774	0.5389***	0.1979**	-0.0604
	(0.0550)	(0.1524)	(0.0707)	(0.0938)	(0.1996)	(0.1078)	(0.1142)	(0.0980)	(0.0742)	(0.1097)	(0.0939)	(0.1193)	(0.2416)	(0.0933)	(0.0673)
RMW	0.2033**	0.1756	-0.0199	0.0123	-0.2049	-0.2548**	-0.3205***	-0.0729	0.1733*	-0.1535	0.0382	-0.1733	0.0216	0.1080	-0.0412
	(0.0670)	(0.2319)	(0.0909)	(0.0831)	(0.3106)	(0.1061)	(0.1171)	(0.0871)	(0.0656)	(0.1286)	(0.1083)	(0.1135)	(0.2239)	(0.1061)	(0.0676)
CMA	-0.1585	-0.4575**	-0.4611***	0.1004	-0.2613	-0.2914**	-0.2955**	-0.2210*	-0.0640	-0.5549***	-0.1388	-0.2588	-0.9420***	-0.4178***	-0.0084
	(0.0800)	(0.2589)	(0.0965)	(0.1695)	(0.3243)	(0.1629)	(0.1782)	(0.1556)	(0.1129)	(0.1588)	(0.1265)	(0.1528)	(0.3295)	(0.1194)	(0.0998)
Adj. R^2	30.06	36.83	47.13	32.69	2.34	39.34	34.28	41.33	41.27	26.38	20.13	24.44	21.23	27.07	54.37
RI-MRP	61.74	69.94	58.18	80.37	47.04	68.19	66.25	72.99	78.94	53.50	55.66	55.53	53.84	56.21	75.84
RI-SMB	19.13	4.78	9.24	2.51	16.12	2.44	2.95	2.25	2.33	12.89	22.09	17.57	14.97	17.15	2.13
RI-HML	2.24	3.46	2.78	1.70	2.44	1.90	3.06	1.98	1.54	3.07	1.83	2.27	3.80	2.87	1.57
RI-RMW	4.40	5.46	8.27	7.90	11.72	12.79	15.00	9.55	5.66	10.80	7.16	11.94	7.79	5.71	8.86
RI-CMA	12.49	16.36	21.54	7.52	22.67	14.68	12.74	13.23	11.53	19.74	13.26	12.70	19.59	18.05	11.60
Panel B: PMNSNT-FF5															
Intercept	0.0002	0.0015	0.0005	0.0001	0.0023	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017	0.0002	0.0009	0.0000	0.0004
	(0.0004)	(0.0008)	(0.0004)	(0.0005)	(0.0017)	(0.0006)	(0.0006)	(0.0005)	(0.0003)	(0.0008)	(0.0008)	(0.0007)	(0.0016)	(0.0006)	(0.0003)
PMNSNT	-0.0137	-0.2096***	-0.0174	-0.1722***	0.2685*	-0.1100**	-0.1553***	-0.1172***	-0.0855**	-0.0112	-0.0861	-0.0132	-0.3054***	-0.0430	-0.0291
	(0.0302)	(0.0879)	(0.0316)	(0.0386)	(0.1476)	(0.0597)	(0.0419)	(0.0434)	(0.0381)	(0.0744)	(0.0512)	(0.0553)	(0.1763)	(0.0557)	(0.0327)
Adj. R^2	30.01	37.56	47.10	34.43	3.23	39.76	35.19	41.92	41.75	26.32	20.32	24.38	22.17	27.07	54.39
RI-PMNSNT	0.06	2.13	0.04	5.17	25.02	1.18	2.73	1.54	1.27	0.02	1.29	0.03	4.49	0.27	0.11

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table continues on the next page

Table 3 continues here:

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel C: SNT-FF5															
Intercept	0.0002	0.0015**	0.0005	0.0001	0.0023	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017**	0.0002	0.0009	0.0000	0.0004
	(0.0003)	(0.0007)	(0.0003)	(0.0005)	(0.0016)	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0006)	(0.0007)	(0.0007)	(0.0014)	(0.0005)	(0.0003)
$s(+)$	0.4947***	0.8787***	0.4937***	0.5110***	0.6755***	0.7673***	0.7075***	0.6935***	0.5678***	0.9225***	0.6983***	0.7059***	1.3555***	0.7339***	0.4841***
	(0.0302)	(0.0439)	(0.0261)	(0.0306)	(0.1139)	(0.0402)	(0.0348)	(0.0397)	(0.0265)	(0.0361)	(0.0541)	(0.0481)	(0.0890)	(0.0346)	(0.0278)
$s(-)$	0.2352***	0.6868***	0.2390***	0.4281***	0.0543	0.4689***	0.4954***	0.4446***	0.3517***	0.4218***	0.4103***	0.3282***	0.9588***	0.4361***	0.2485***
	(0.0472)	(0.0783)	(0.0304)	(0.0527)	(0.1514)	(0.0687)	(0.0573)	(0.0580)	(0.0352)	(0.0708)	(0.0650)	(0.0708)	(0.1811)	(0.0519)	(0.0349)
Adj. R^2	57.50	65.22	66.98	57.66	12.86	69.16	62.44	71.52	71.93	59.77	42.42	46.08	49.40	61.26	74.64
RI-$s(+)$	39.57	31.67	24.47	26.64	77.34	32.30	31.35	30.86	31.73	46.89	39.99	39.17	39.37	44.97	21.96
RI-$s(-)$	7.84	11.51	5.02	16.38	0.26	10.56	13.47	11.11	10.66	8.59	12.09	7.42	17.20	10.41	5.07
Panel D: SNT-MOM-FF5															
Intercept	0.0002	0.0013	0.0006	0.0001	0.0024	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017**	0.0002	0.0010	-0.0001	0.0004
	(0.0003)	(0.0007)	(0.0004)	(0.0005)	(0.0016)	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0006)	(0.0007)	(0.0007)	(0.0014)	(0.0005)	(0.0003)
MOM	-0.0455**	0.0158	-0.0182	-0.0605**	0.0785	-0.0770***	-0.0271	-0.0288	-0.0333*	-0.0650**	-0.0719*	-0.0094	-0.1448**	-0.0994***	-0.0758***
	(0.0147)	(0.0498)	(0.0194)	(0.0214)	(0.0776)	(0.0264)	(0.0285)	(0.0257)	(0.0179)	(0.0242)	(0.0343)	(0.0319)	(0.0565)	(0.0226)	(0.0168)
$s(+)$	0.4958***	0.8280***	0.4961***	0.5136***	0.6741***	0.7699***	0.7127***	0.6980***	0.5709***	0.9254***	0.7004***	0.7100***	1.3598***	0.7149***	0.4838***
	(0.0295)	(0.0435)	(0.0269)	(0.0312)	(0.1095)	(0.0399)	(0.0346)	(0.0389)	(0.0274)	(0.0373)	(0.0548)	(0.0484)	(0.0877)	(0.0352)	(0.0308)
$s(-)$	0.2340***	0.6877***	0.2390***	0.4265***	0.0518	0.4668***	0.4955***	0.4445***	0.3513***	0.4204***	0.4083***	0.3289***	0.9553***	0.4329***	0.2459***
	(0.0453)	(0.0755)	(0.0310)	(0.0519)	(0.1537)	(0.0672)	(0.0567)	(0.0564)	(0.0365)	(0.0663)	(0.0654)	(0.0692)	(0.1747)	(0.0499)	(0.0350)
Adj. R^2	57.58	65.29	66.94	57.77	13.03	69.27	62.40	71.49	71.92	59.79	42.49	46.05	49.50	61.69	75.03
RI-MOM	0.58	0.02	0.06	0.63	1.17	0.56	0.08	0.09	0.18	0.41	0.76	0.01	0.78	1.51	0.92
RI-$s(+)$	39.28	31.22	24.43	26.40	76.42	32.04	31.39	30.85	31.63	46.71	39.70	39.26	39.14	42.70	21.51
RI-$s(-)$	7.61	12.79	5.00	16.08	1.32	10.36	13.36	11.01	10.52	8.39	11.77	7.33	16.98	9.71	4.91

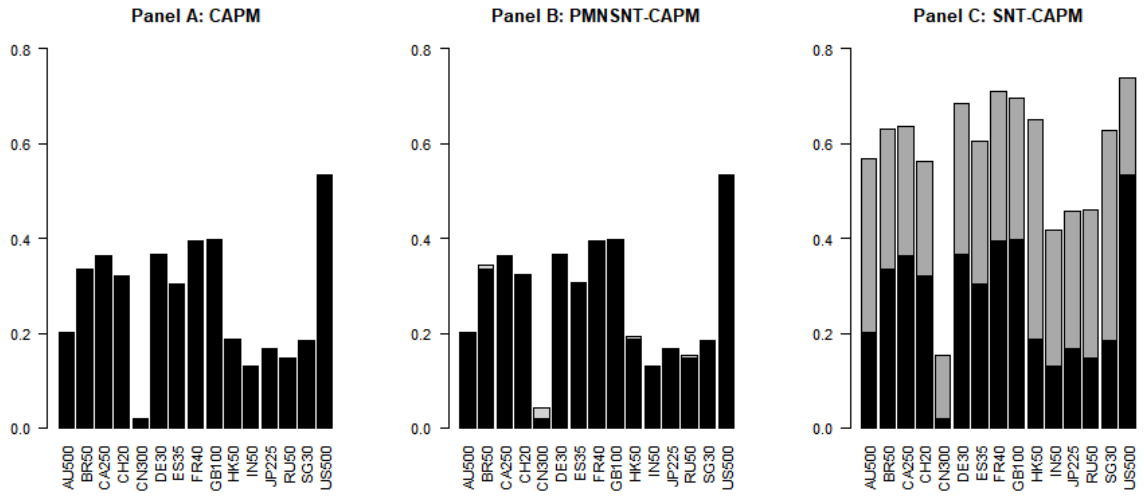
* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table 3 reports the estimation results for the Fama-French five-factor models with and without sentiment according to Equation (4) in Panels A and C, respectively. The coefficients associated with the sentiment mimicking returns remain mostly unchanged in terms of sign, estimated coefficient size, and relative significance when the additional variables are added to the CAPM. Although the Fama-French factors, in particular SMB and CMA, matter in various equity markets, no clear pattern emerges. SMB is often precisely estimated, which can be explained by the selection bias in favor of large companies, which represent the constituents of the major equity indices. In these cases, when SMB is significant, the adjusted R^2 increases sensibly, although the market remains the most relevant factor. Here, the RMI indicators provide the largest contributions to the explained variation after the CAPM component, which appears to be a robust finding.

We also estimate a model for the single sentiment factor in Panel B. The PMNSNT-CAPM model confirms that a single sentiment factor is unable to capture the variability in market sentiment and hence remains mostly unchanged to the CAPM specification when the Fama-French factors are added. If we also include a momentum factor, which is (despite considerable debate) missing from the Fama-French five-factor model, a negligible increase in the model's fit is recorded. Panel D of Table 3 shows that the momentum factor is always significant at the 10% level but its contribution to the explained variation is significantly lower than sentiment, hardly exceeding 1%. We conclude that a single sentiment factor measured as the return of a long-positive/short-negative sentiment portfolio loses further significance if traditional variables based on the Fama-French five-factor model are included in the empirical model. However, the separate sentiment mimicking portfolios, which distinguish between positive and negative sentiment shocks, are accurately estimated and economically significant in our linear specifications. Moreover, none of the sentiment sorts simply captures momentum, as they all remain significant with estimated coefficients mostly unchanged. Momentum does not appear to represent a meaningful addition to explain the excess returns on international equity indices.

Figure 4: Comparing the R^2 of the CAPM to the sentiment-augmented CAPM

This figure displays the adjusted R^2 for various sentiment-augmented linear factor models (light grey), benchmarked against the standard CAPM (black). Panel A shows the results for the CAPM. Panel B is a CAPM extension with a single sentiment risk factor based on the excess return of a long-positive/short-negative portfolio (PMNSNT-CAPM). Panel C uses the excess returns of all portfolio sorts based on negative, neutral, and positive sentiment (SNT-CAPM). The sample period is from January 1998 to December 2017. All plots use the same scale to favor direct visual comparisons.



5.2 Benchmarking MarketPsych against other sentiment proxies

In this section, we demonstrate the superiority of the RMI indicators by benchmarking linear regression models against a variety of well-known and more-established sentiment measures. Given that these models tend to emphasize the pricing effects of negative sentiment, we employ a sentiment-augmented CAPM based on the negative RMI scores only (NSNT-CAPM) in Panel A of Table E.1 of the Appendix. Negative sentiment is highly statistically significant with the exception of the Chinese market (CN300) and shows the highest explained variation for all models involved in the comparison (unreported).[23] Fundamental-based sentiment proxies are represented by the orthogonalized BW index

[23]Relative to the full sentiment model SNT-CAPM in Table E.5, this model shows an inferior performance because the explanatory power of positive sentiment is neglected. However, given that the negative sentiment factor is only orthogonalized against the market risk premium, it captures some of the variation of the positive sentiment factor it is correlated to.

proposed by Baker and Wurgler (2006).[24] The BW index is significant for all of the equity markets but the explained variation is marginally higher than the CAPM and worse than the negative sentiment CAPM, supporting our hypothesis in Section 3.1 that BW based on fundamental information may capture a different lifecycle of sentiment. Panel C shows the results for the improved BW index by Huang et al. (2015) applying partial least squares (PLS). We note a further enhancement of the goodness-of-fit measure but the model remains inferior to our sentiment-augmented CAPM with only negative sentiment. The downside risk factor $DR \times MRP$ from Lettau et al. (2014) in Panel D turns out to be insignificant for all markets.

An alternative bottom-up search based sentiment proxy is represented by the FEARS index proposed by Da et al. (2015).[25] The FEARS factor is statistically significant for all markets and achieves a performance between the BW-CAPM and PLS-CAPM. The explained variation is approximately 50% smaller than the RMI model, on average. This may reflect MarketPsych's superior human language processing system that is able to capture a wider range of negative emotions than FEARS as well as the fact that FEARS is similar to BW purely based on U.S. data. The results of the closest benchmark, RavenPack, in Panel E are surprisingly poor even though the variable is significant for most countries except China. On average, the explained variation is 40% lower than for the NSNT-CAPM. It appears that the performance of the RavenPack indicator comes closest to MarketPsych demonstrating the superiority of bottom-up news-based sentiment compared to other proxies.

[24] We note two limitations of the BW index: (i) the BW index is purely based on U.S. data but we see the benchmark as appropriate due to the linkage of significant capital flows between the U.S. and the world's biggest economies as stated by Baker et al. (2012). U.S. sentiment does predict the cross-section of country returns for Canada, France, Germany, Japan, and the U.K.. Baker et al. (2012) also construct local sentiment indices for those markets but the time series data is not publicly available. Re-constructing those indicators is beyond the scope of this paper. (ii) the BW index is only available on a monthly basis. We use a cubic spline to interpolate the time series to weekly.

[25] Because FEARS is publicly available only for a July 2004-December 2011 sample, the estimates in Table E.1 refer to this sub-sample period only.

5.3 Global versus local sentiment

In the spirit of Baker et al. (2012) and benefiting from the broad international coverage of RMI's sentiment indicator, in this section we study whether there is an alternative view to our global sentiment factors based on portfolio sorts. In particular, we (i) examine the performance of local sentiment for CAPM models of international equity indices and (ii) we apply an alternative specification of global sentiment using the first and second principal components of local sentiment measures. Compared to our previous hypothesis that sentiment is a global risk factor that drives aggregate equity returns, we temporarily assume that each market is (also) driven by its individual sentiment. Table E.2 provides in Panel A the results of a CAPM which adds the RMI sentiment change of each country to the corresponding predictive regressions that implement linear factor pricing.[26] The local sentiment factors are significant for all equity markets, with positive estimated predictive coefficients. The Chinese index CN300 shows the highest estimated sensitivity of 0.14. Compared to our CAPM specification with global sentiment the results are similar to the PMNSNT-CAPM with a global long-positive/short-negative sentiment factor in terms of goodness-of-fit but the relative importance of the local sentiment factor is considerably higher than in the case of PMNSNT. However, the results are inferior to our extended specification with separate positive and negative sentiment. We find that local sentiment matters in the absence of global sentiment. Panel B thus adds our two global factors $s(+)$ and $s(-)$ based on portfolio sorts to the local sentiment factors, creating a specification of "total" (i.e., both global and local) sentiment.[27] Excluding Brazil, Canada and the UK – the latter two are also in the sample of countries investigated by Baker et al. (2012) – we find that local sentiment remains significant for most markets while the coefficients of the global factors remain as strong and precisely estimated as in the SNT-CAPM. Local sentiment reaches the highest importance in the case of the Chinese CN300 index while not exceeding an R^2 contribution of 2.5% for any other national stock market. These

[26] In unreported results, we also estimate the regressions including local sentiment variables in levels rather than their change but find no remarkable differences.

[27] Both the global and local sentiment variables are orthogonalized to market excess returns and to each other in order to address concerns of multicollinearity.

findings are in line with Baker et al. (2012) who report that country-level results are mainly driven by global sentiment.

Additionally, we have also used PCAs instead of portfolio sorts to construct the global sentiment factor(s). The factor loadings are reported in Table E.3 in the Appendix and the estimation results are instead in Panels C and D of Table E.2. First, we construct a single global sentiment factor similar to the one in Baker et al. (2012) but based on the local sentiment change of the RMI indicators. The first component is significant for all markets except Brazil and the UK, with positive but miniscule estimated coefficients. Its relative importance exceeds that of the local sentiment factors and commands a total explained variation in line with the PMNSNT-CAPM model. Second, in Panel D, we use the first two principal components which can be interpreted as a benchmark for the SNT-CAPM which includes both positive and negative sentiment factors computed from the portfolio sorts. However, the second component is precisely estimated with small positive coefficients only in the case of Brazil, Germany, and the United States. The goodness-of-fit measures are marginally better than before but inferior to our two sentiment-factor model. As such, the principal components are unable to cover the same notions of sentiment as our sentiment-mimicking portfolio sorts which may be partially due to PC1 and PC2, and yet only explain 42% of the entire variation in the local sentiment indicators.

6 Sentiment risk premia

In order to quantify the additional return demanded by investors for investments in sentiment-responsive assets, in this section we proceed to estimate the price of global sentiment risk. In fact, we compute a separate risk premium for positive and negative changes in sentiment. We follow the two-stage regression methodology in Fama and MacBeth (1973) (henceforth FMB) to compute the sentiment price of risk and systematically compare our approach with the in-sample fit of the CAPM and FF5 using standard error corrections to account for cross-asset correlation and heteroscedasticity. In line with Ang

et al. (2006), we argue that in times of increased uncertainty and fear, investors expect to be compensated for the additional risk from investing in sentiment-sensitive assets. We also conjecture that investors fear over-optimism and market overheating, and hence also demand a positive risk premium in the case of positive sentiment-exposed assets.

The time series regressions of the first stage in the FMB procedure yield the point estimates of the market $\hat{\beta}_i$ and sentiment betas. These are then used as explanatory variables in a second-stage, cross-sectional regression of the average return of the international indices on the first-stage sentiment betas. This two-stage approach of using estimated betas of the first-stage as variables in the second, introduces a generated regressor bias which we correct using the standard error adjustments in Shanken (1992). In the case of our sentiment-augmented model, the second-stage regression is

$$r^e_{i,t} = \hat{\beta}_i \bar{r}^e_m + \hat{\beta}_{s(-),i}\lambda_{s(-)} + \hat{\beta}_{s(+),i}\lambda_{s(+)} + \epsilon_{SNT,i,t}, \forall t \in T, \tag{5}$$

where $r^e_{i,t}$ is the asset excess return, $\hat{\beta}_{s(-),i}$ and $\hat{\beta}_{s(+),i}$ are the betas on the excess return of the positive and negative sentiment mimicking portfolios at time t, respectively. $\lambda_{s(-)}$ and $\lambda_{s(+)}$ are the prices of risk for positive and negative sentiment. Of course, this framework is easy to adapt to the PMNSNT-CAPM, which would yield a unique estimate of the price of sentiment risk, λ_{PMNSNT}. This regression is estimated at each time t. As pointed out by Cochrane (2009), large sampling errors, resulting from cross-sectional correlation of asset returns, are a key obstacle when producing inferences in cross-sectional analyses. Performing the recursive estimation on sub-samples and averaging the statistics accounts for this cross-sectional correlation and reduces the sampling error. The FMB approach takes this idea further and calculates the cross-sectional regression for each period t.

We jointly estimate the first-stage, full-sample betas for the market, $s(-)$ and the $s(+)$ factors. We price the market return and exclude an intercept term such that an asset with zero beta has a zero excess return as assumed in Lettau et al. (2014). As in Fama and French (1992), we regress the indices directly on the factors, not on the long-short mimicking portfolios. This is justified by the fact that the indices represent diversified

portfolios by definition, even though they tend to be inevitably biased towards large caps. This procedure presumably may lead to more noisy estimators with higher standard errors, but it seems more appropriate in the case of our empirical exercise. To reduce the omitted variable bias and address a concern that sentiment may measure omitted statistical properties of excess equity returns, such as idiosyncratic volatility $iv_{i,t}$, skewness $is_{i,t}$, and kurtosis $ik_{i,t}$ (that are left unexplained by the "fundamental" factors in either the CAPM or FF5), we additionally control for such features in our implementation of the FMB methodology.[28] Following Boyer, Mitton, and Vorkink (2010), we define these control variables as:

$$iv_{i,t} = \left(\frac{1}{N(t)} \sum_{d \in S(t)} \epsilon_{i,d}^2 \right)^{\frac{1}{2}}, is_{i,t} = \frac{1}{N(t)} \frac{\sum_{d \in S(t)} \epsilon_{i,d}^3}{iv_{i,t}^3}, \text{ and } ik_{i,t} = \frac{1}{N(t)} \frac{\sum_{d \in S(t)} \epsilon_{i,d}^4}{iv_{i,t}^4}, \quad (6)$$

where $\epsilon_{i,d}$ is the residual of either the CAPM or the FF5 models. Given that our study offers an application to international equity returns, we also control for the effects of major global currencies. Therefore, we use the U.S. Dollar index returns, which represent the log-changes of the USD versus a basket of currencies like EUR, JPY, GBP, CAD, SEK, and CHF. The full specification also controls for momentum.

6.1 Risk premia for global sentiment

We compare eight different models: i) a standard CAPM, used as a naïve benchmark, ii) a standard Fama-French five-factor model FF5, iii) a PMNSNT-CAPM model that augments the CAPM model by adding a long-positive/short-negative sentiment factor, iv) a PMNSNT-FF5 which includes the PMNSNT factor in a Fama-French five-factor specification, v) a PMNSNT-FF5x model that additionally controls for $iv_{i,t}$, $is_{i,t}$, $ik_{i,t}$, and the log returns on the U.S. Dollar index, vi) a SNT-CAPM that prices positive and negative sentiment separately, vii) a SNT-FF5 Fama-French five-factor specification including two sentiment factors, and finally viii) a SNT-FF5x model that also incorporates

[28] Because these properties are idiosyncratic factors that should not be priced in the cross-section, we do not include them in the second-stage, but only in the time-series regression. A similar rationale applies to momentum which is why we refrain from estimating a separate model.

the same control variables listed above.

In Table 4, the R^2 of the CAPM is 36.34%, supported by a price of market risk of 0.22%, i.e. it is priced higher in the cross-section than the average weekly excess percentage return for the MSCI World index of 0.12%. Estimates of the Fama-French model show that the size (SMB) and investment (CMA) factors do not yield a positive risk premium in contrast to the market, value (HML) and profitability (RMW) factors. Global equity indices select the largest companies with the largest investment outlays and because these companies represent the majority share of the indices, the related size and investment factors may not be separately priced. The explained variation is 35.11% and not significantly different to the CAPM-based benchmark. With 42.58%, the PMNSNT-CAPM model exceeds the cross-sectional R^2 delivered by the FF5 model, which remains stable when the five Fama-French factors are added. Market risk is priced slightly higher in the cross-section. We note the positive and significant risk premia of the PMNSNT factor, estimated at 0.09%. If we add the fundamental control variables, the estimated premium and explained variation remain stable while the risk premium is reduced to 0.02% when further control variables are added. Thus, the idiosyncratic control variables affect the pricing of the global sentiment factor. Yet, the Sentiment-CAPM including disaggregated sentiment indicators (SNT-CAPM) performs markedly better than the CAPM, FF5, PMNSNT-CAPM, and PMNSNT-FF5(x) models and implies a significantly negative sentiment risk premium of -0.06% for negative sentiment. This is in line with the results from the portfolio sorts. Negative sentiment does not seem to be a compensated factor. Positive sentiment achieves a weekly average excess return of 0.12% and, as such, compensates investors for bearing such risk. If we control for the typical "fundamental factors" or for factors that reflect the statistical properties of returns, the estimated sentiment premia keep the same sign. Interestingly, with the introduction of the higher moment factors in the first stage, the absolute values of the estimated premia increase for all the sentiment factors to levels similar to the market risk premium. In terms of explained variation, we observe a jump to 48.18% from when the two sentiment factors are added. Additional fundamental variables do not have a significant effect. The

adjusted R^2 slightly decreases from 48.18% for the simple SNT-CAPM model to 47.18% for the SNT-FF5 and increases to 49.11% for the model including the control variables. The negative sentiment premium is reduced in absolute terms to -0.04% while positive sentiment becomes even higher priced at 0.18%. Overall, the results for positive and negative sentiment are in line with our previous findings and support the addition of sentiment factors to asset pricing models.[29] However, whether the sentiment premium persists in the long-term remains open and is explored in the following section.

6.2 Long-term sentiment premia for developed and emerging markets

Tetlock (2007) discusses two key questions. The first concerns the timing of media pessimism associated with low investor sentiment, while the second asks whether high media pessimism forecasts or reflects past investor sentiment. While our study primarily focuses on the hypothesis that sentiment forecasts investor actions, we cannot rule out the reverse effect. In this case, low sentiment would forecast low returns in the short-term but higher returns in the long-term, as visually shown by Tetlock (2007) in Figure 1. This so-called sentiment theory is contrasted to the information theory. The second question asks whether media pessimism as a proxy for negative information about fundamental values of equities. If the sentiment theory holds, low sentiment predicts low returns in the short-term, which we find to be the case for our data, but a reversal of returns at longer horizons. If the information theory prevails instead, the negative (positive) returns will persist indefinitely. We study these hypotheses by splitting our dataset into developed

[29] For robustness, we also estimated the SNT-FF5x model with each sentiment factor separately, i.e., when the second-stage FMB regression only accounts for one single sentiment factor at the time. When we only price positive sentiment, the risk premium is higher at 0.21% with a reduced R^2 of 42.19% which is in the range of the PMNSNT-FF5; Pricing negative sentiment in isolation, leads to a switch in the sign of the premium to positive 0.12% with R^2 of 41.47%. That means that negative sentiment bears some positive risk premia if positive sentiment is neglected. We trace this back to the high variability in the negative sentiment factor mimicking portfolio and to a significant correlation between the time series of negative and positive sentiment betas of 0.81 from the first-stage regression of the SNT-CAPM model.

Table 4: Cross-sectional pricing comparison of the CAPM, FF5, and sentiment-augmented models

This table shows the results of a comparison between eigth different models: i) a standard CAPM, used as a benchmark, ii) a standard Fama-French five-factor model FF5, iii) a PMNSNT-CAPM model that extends the CAPM model to include a long-positive/short-negative sentiment factor, iv) a PMNSNT-FF5 which adds the PMNSNT factor to a Fama-French five-factor specification, v) a PMNSNT-FF5x model that additionally controls for $iv_{i,t}$, $is_{i,t}$, $ik_{i,t}$, and the log returns on the U.S. Dollar index, vi) a SNT-CAPM that prices positive and negative sentiment separately, vii) a SNT-FF5 Fama-French five-factor specification with two sentiment factors, and viii) a SNT-FF5x model that also incorporates the same control variables listed above. We report the market risk premium as λ and its standard error. For the CAPM we assume market risk to be priced correctly and hence equal to the average weekly excess return on the MSCI World portfolio, so that no standard error of the estimate is provided. λ_{PMNSNT} is the price of the single sentiment risk factor PMNSNT. The two estimates $\lambda_{s(+)}$ and $\lambda_{s(-)}$ are the sentiment premia for positive and negative sentiment change, respectively. Additional rows indicate whether we also control jointly for the Fama-French factors for size, value, profitability and investment, the change in idiosyncratic volatility $iv_{i,t}$ (dVol), skewness $is_{i,t}$ (dSkew), kurtosis $ik_{i,t}$ (dKurt), and the log returns on the U.S. Dollar index computed from a basket of international currencies. Standard errors of the estimates are provided in parentheses and incorporate the Shanken (1992) errors-in-variables correction as well as the Newey-West correction for heteroscedasticity. P-values are corrected by the Bonferroni-Holm method to account for multiple testing bias. The last row reports the adjusted R^2 of each model for the cross-section of returns.

	Baseline		Aggregated Sentiment Factor PMNSNT			Disaggregated Sentiment Factors SNT		
	CAPM	FF5	CAPM	FF5	FF5x	CAPM	FF5	FF5x
λ	0.2178***	0.2174***	0.2315***	0.2418***	0.2148***	0.1063***	0.0978***	0.0145***
	(0.0029)	(0.0031)	(0.0033)	(0.0038)	(0.0027)	(0.0006)	(0.0005)	(0.0003)
λ_{PMNSNT}			0.0872***	0.0984***	0.0157***			
			(0.0011)	(0.0013)	(0.0011)			
$\lambda_{s(+)}$						0.1220***	0.1559***	0.1796***
						(0.0006)	(0.0010)	(0.0013)
$\lambda_{s(-)}$						-0.0646***	-0.0967***	-0.0372***
						(0.0008)	(0.0009)	(0.0007)
FF5 Factors				Yes	Yes		Yes	Yes
dVol					Yes			Yes
dSkew					Yes			Yes
dKurt					Yes			Yes
USD Index					Yes			Yes
Adj. R2	35.34	35.11	42.58	42.51	43.53	48.18	47.18	49.11

* p $\leq$ 0.1, ** p $\leq$ 0.05, *** p $\leq$ 0.01

markets[30] versus emerging markets.[31] We limit the analysis to the period 2006-2017 due to limited availability of sentiment data for emerging markets for earlier periods. The construction of the sentiment factor separates accordingly between developed and emerging markets. We recap the information provided in Panels B and C of Table 1 which reports negative returns for negative sentiment for developed markets as shown for the global version of the portfolio sorts. For emerging markets, positive and negative sentiment are both associated with positive returns. The spread still accounts for 0.27%, which is similar to 0.28% for developed markets but is no longer statistically significant. The mean return following negative return is still positive with 0.54% weekly. To differentiate the information and sentiment theory, we compute the sentiment risk premium for the PMNSNT and positive/negative sentiment factor at two, four, and eight weeks ahead. If our earlier theory that sentiment has a higher impact in inefficient markets held, we would expect to obtain evidence favorable to the information theory in the case of emerging markets and evidence supporting the sentiment theory from developed markets' data. Table 5 shows a reduction from a positive PMNSNT premium of 0.0288% for one week to a weekly average of 0.0155%, respectively 0.0020%, for two and four weeks and an ultimate shift to a negative risk premium of -0.0045% for eight weeks. For positive and negative sentiment the same reversal is shown already for the fourth week, when the premium for positive sentiment switches from positive to negative and for negative sentiment vice versa. This is the expected pattern for developed markets if the sentiment theory holds. On the contrary, for emerging markets we observe two things; first, that negative sentiment has a positive risk premium as suggested by the portfolio sorts which is four to five times higher than for positive sentiment. Investors in emerging markets are compensated for negative sentiment risk exposure. Second, that average weekly sentiment premia for PMNSNT, positive and negative sentiment slightly increase the longer the forecast horizon and remain positive. This is consistent with the hypothesis of inefficient, emerging markets and the

[30]The developed markets comprise the country indices of the United States (US500), Australia (AU500), Canada (CA250), Switzerland (CH20), Germany (DE30), Spain (ES35), France (FR40), Great Britain (GB100), Hong Kong (HK50), Japan (JP225), and Singapore (SG30).

[31]The emerging markets comprise the stock indices for the four BRIC countries, i.e., Brazil (BR50), Russia (RU50), India (IN50), and China (CN300).

information theory of Tetlock (2007).

7 Robustness checks

7.1 State-dependent Fama-French five-factor model

In order to improve the fairness of the horse race among models with and without sentiment, we implement a conditional version of the baseline Fama-French five-factor model with state-dependent interaction terms. The idea is that investors may simply price common risk drivers like the five Fama-French factors differently in negative and positive sentiment states. We therefore code a dummy sentiment variable, which takes a value of one if the change in our global RMI sentiment indicator is negative, i.e., when abnormal negative sentiment is recorded, and interact this proxy with each classical factor.[32] The results are recorded in Panel B of Table E.4 and show that the interaction is only significant once for the value factor for UK (GB100). In most of the markets, the coefficients of the fundamental factors remain of the same sign and of similar amplitude compared with earlier results in Section 5. We observe that the effects of a number of factors stop being significant in some markets as soon as the sentiment interaction is included in the model. Compared to the SNT-FF5 model, which considers sentiment directly as a priced factor, the FF5 model with sentiment-driven interaction terms performs considerably worse. In light of these results, we conclude that there is not enough evidence to reject the hypothesis of sentiment being a priced risk factor but rather a state variable that modifies the pricing of other, well-known factors.

[32] In unreported results, we also apply the sentiment level, the CAPM- and FF5 orthogonalized PMNSNT factor as an alternative proxy for the sentiment state and find that our main results hold independently of the specification of the interaction term.

Table 5: Cross-sectional pricing comparison between developed and emerging markets

This table shows the results of a comparison between two different models with different forecasting horizons spanning one, two, four, and eight weeks: i) a PMNSNT-FF5 which adds the PMNSNT factor to a Fama-French five-factor specification, ii) a SNT-FF5 model that uses both positive and negative sentiment separately. We report λ_{PMNSNT}, the price of the single sentiment risk factor PMNSNT over the corresponding forecasting horizon, $\lambda_{s(+)}$ and $\lambda_{s(-)}$ for the postive and negative sentiment premium. For the two- to eight-week forecast horizon, the premia are given as a weekly average for better comparison with the one-week ahead prediction. Standard errors of the estimates are provided in parentheses and incorporate the Shanken (1992) errors-in-variables correction as well as the Newey-West correction for heteroscedasticity. P-values are corrected by the Bonferroni-Holm method for multiple testing bias. The last row reports the adjusted R^2 of each model for the cross-section of returns. The data comprises January 2006 to December 2017 due to the limited availability of sentiment data for emerging markets beforehand. To disentangle emerging and developed markets, we split the dataset and sentiment factor accordingly and use a market-specific (either emerging or developed) factor. Panel A uses the developed market sentiment factors on the related indices. Panel B constructs the sentiment factor only on emerging markets and augments the FF5 models for the same index universe.

	1 Week		2 Weeks		4 Weeks		8 Weeks	
	PMNSNT-FF5	SNT-FF5	PMNSNT-FF5	SNT-FF5	PMNSNT-FF5	SNT-FF5	PMNSNT-FF5	SNT-FF5
Panel A: Developed Markets with Developed Market Sentiment Factor								
λ_{PMNSNT}	0.0288***		0.0155***		0.0020***		-0.0045***	
	(0.0006)		(0.0004)		(0.0003)		(0.0001)	
$\lambda_{s(+)}$		0.0587***		0.0105***		-0.0098***		-0.0051***
		(0.0009)		(0.0005)		(0.0003)		(0.0001)
$\lambda_{s(-)}$		-0.0644***		-0.0173***		0.0069***		0.0067***
		(0.0010)		(0.0006)		(0.0004)		(0.0001)
Adj. R2	49.17	53.41	47.85	51.49	47.87	51.46	50.55	53.52
Panel B: Emerging Markets with Emerging Market Sentiment Factor								
λ_{PMNSNT}	0.2710***		0.2900***		0.3082***		0.4019***	
	(0.0034)		(0.0034)		(0.0039)		(0.0104)	
$\lambda_{s(+)}$		0.3204***		0.4000***		0.4331***		0.4973***
		(0.0041)		(0.0085)		(0.0127)		(0.0335)
$\lambda_{s(-)}$		1.7082***		1.8331***		1.8870***		1.9275
		(0.2213)		(0.4388)		(0.6451)		(1.3643)
Adj. R2	28.11	45.67	30.62	47.53	30.24	47.78	29.12	49.74

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

7.2 Tests on an extended set of global equity indices

We extend our empirical tests by considering additional equity indices of medium- and small-sized companies and aggregate indices for both the Eurozone and emerging markets to show that our results are more generally applicable. The specialty of this analysis is that the local RMI sentiment indicator of those indices have not been considered for the construction of our global sentiment factors. First, we use a range of alternative U.S. and UK equity indices and compare them to our previous regression results for the S&P 500 and FTSE 100. The findings, summarized in Table E.5 of the Appendix, emphasize that all of our key results hold for the Dow Jones Industrial Average (US30), the Russell 2000 (USMID2000) of mid-sized companies, the tech-oriented NASDAQ 100 (USNAS100), and the index ranking the 101st-350th size-ranked, UK-based, London stock exchange (LSE) listed companies ($GBMID250$).[33] We also extend the analysis to aggregate indices like the top 50 emerging markets companies, measured by the MSCI 50 (EM50), and the top 50 pan-European companies, covered by the EURO STOXX 50 (EU50). We note the high exposure of emerging markets companies to positive sentiment, which further supports our hypothesis that sentiment is more applicable to inefficient markets.

7.3 Sub-samples analysis

We also evaluate our models on sub-samples to validate our initial intuition that sentiment is more important during crisis regimes, as classical asset pricing models are well-known to fail in extreme, turbulent market regimes (see, e.g., Hillert, Jacobs, & Müller, 2014; Yu & Yuan, 2011). We split the data into the 1998-2001, 2002-2006, 2007-2011, and 2012-2017 periods.[34] The aggregate regression results in Table E.6 in the Appendix show that adding the two novel sentiment factors to both a CAPM or FF5 specification approximately

[33] In unreported results we also include those factors into the global sentiment construction and arrive at similar findings. The reduction in the reported R^2 can further be explained by the lower media coverage characterizing smaller companies.

[34] The splits were motivated economically to differentiate between alternative bull-bear cycles in international financial markets but also to include a similar number of observations in each phase. As such, we focus on sub-periods that span both bull market regimes like the 2002-2006 and 2012-2017 periods, as well as crisis regimes, like 1998-2001 (the Dot.Com crisis) and, of course, 2007-2011 (the Global Financial crisis, encompassing also the European sovereign debt crisis).

doubles the average R^2 in all market phases. The single sentiment models, either PMNSNT-CAPM or PMNSNT-FF5, only reveal a marginal contribution by sentiment below 10% on average across all sub-periods. The period 2007-2011 is marked by a high average R^2 across all international markets. This confirms our hypothesis that sentiment as a priced risk factor becomes stronger during crises. This result is consistent with earlier evidence shown in García (2013) on the impact of news on stock returns but differs from the conclusions drawn by Antoniou, Doukas, and Subrahmanyam (2016) who find that investor behavior appears more rational during pessimistic periods due to reduced noise trading. We also note a very high contribution of sentiment to the R^2 in the 2012-2017 sub-sample. Yet, given that our sentiment sorts exhibit higher average excess returns for positive sentiment, this is unsurprising. It seems that sentiment has become more relevant in recent years, which we trace back to both the occurrence of improvements in the precision of the human-language processing engine applied by MarketPsych as well as to a better coverage of financial markets in both the news and social media channels. An alternative analysis on sub-samples with similar results but based on the categorization of recessions offered by the National Bureau of Economic Research (NBER) can be found in the Appendix F.

7.4 Separating the effects of news from social media channels

MarketPsych also disentangles the signals from both news and social media channels offering distinct sentiment measures as recently emphasized by Jiao et al. (2020) with reference to volatility and trading volumes for individual U.S. stock returns. We hypothesize that social media may be more relevant to retail investors, whereas news may be preferred by institutional traders to inform their investment decisions because these are generally presumed to be less volatile and more reliable. Given the efficient market hypothesis by which all information is instantaneously reflected in market prices, we expect that social media-based sentiment may either contain more timely information or proxy for the prevalence of private investors who are more exposed to sentiment and exuberance, compared to widely available public news. Therefore, we estimate our linear factor models

separately for the two channels and observe that social media-based models perform slightly better than news-based or aggregated models covering both sources. We evaluate the minimum, maximum, average, and median R^2 and the average, relative importance of sentiment. Sentiment-augmented models are uniformly better for social media-only sentiment for the 2-factor sentiment CAPM, the FF5 specification, and the single factor sentiment-augmented models. Overall results are shown in Table E.7 in the Appendix and appear to be in line with previous findings on similarly classified sentiment data, e.g., by Nooijen and Broda (2016).

7.5 Fama-MacBeth without orthogonalization of the factors

The applied orthogonalization of all factors comes at the cost that the estimated risk premia may no longer be easily interpretable. The results for removing this orthogonalization are summarized in Table E.8 in the Appendix. The empirical results concerning the CAPM and FF5 models remain identical as they were not affected by orthogonalization. In the case of the PMNSNT-based models, the estimated risk premium is slightly lower. For the extended PMNSNT-FF5x that includes control variables the estimated risk premium even declines from 0.02% to -0.01%. The estimates of the models including the two sentiment variables, however, change as expected due to correlations between the returns on pairs of sentiment factor-mimicking portfolios. The risk premia on the two sentiment indicators become positive with the positive sentiment risk premium climbing much higher than the negative sentiment premia. This represents evidence that the orthogonalization of the factors is essential to capture the marginal effects of the sentiment portfolios.

8 Conclusion

Using recently popularized, proprietary sentiment measures based on MarketPsych's human-language processing engine applied to news and social media feeds, this paper proves the existence of a strong empirical relationship between sentiment and the excess returns of a number of international equity indices. Moreover, we uncover strong evidence

that sentiment is a priced risk factor. We sort the test equity indices according to the methodology popularized by Fama and French (1993, 2015, 2017) into quartiles according to the deviation of their sentiment score from its country/index-specific long-term average. We show that global long/short portfolios generated on the basis of these sorts generate a significant risk-adjusted outperformance over market- and factor-based normal returns. The outperformance is positive for positive deviations of sentiment from its long-term average, and negative for negative deviations. This represents a new finding, qualitatively different from the existing evidence in Baker and Wurgler (2006, 2007), by which positive excess returns are accompanied by negative sentiment shocks. While Baker and Wurgler's and other fundamental-based indices capture market outcomes like turnover and stock issuance that follow from investment behaviors, our sentiment anticipates investors' actions and is therefore in line with psychological evidence, by which business activity tends to follow, rather than lead, social moods.

These insights are used to benchmark multiple linear regression models including sentiment as a priced risk factor against the standard CAPM and Fama and French's five-factor model. Our specifications consistently yield a better goodness-of-fit from the sentiment-augmented models versus the benchmarks. We demonstrate the superiority of our sentiment indicators against a variety of alternative proxies formed from fundamental information, market data, internet searches or by competing human-language processing engines. We report that sentiment cannot be fully captured by a single risk factor, because positive and negative deviations in sentiment are differently priced by the cross-section of international equity indices. Our time series regressions emphasize that sentiment is asset-specific and has more explanatory power for stocks in less efficient markets where it also carries lower correlations with traditional factors. While yielding precisely estimated coefficients, sentiment is more relevant as a priced risk factor in the case of less developed countries and emerging markets. This confirms the conjecture that sentiment may play a more prominent role in informationally inefficient markets. We find further empirical support to the sentiment theory by Tetlock (2007) in the case of developed markets, while emerging markets appear to be ruled by the information theory. We show that in

line with Baker et al. (2012), country-level results are mainly driven by global instead of local sentiment and that our methodology of portfolio sorts is superior to global sentiment extracted from principal components. Using the FMB technique applied to sentiment-augmented models, we compute a positive conditional sentiment risk premium for positive changes in sentiment, while the estimated risk premium turns out to be negative and significant for negative sentiment. Compared to the standard CAPM and a Fama-French five-factor specification, our sentiment-augmented model significantly improves the in-sample predictive power even in the presence of additional controls for idiosyncratic volatility, skewness, kurtosis, and the log-returns on a basket of international currencies. In summary, we report empirical evidence supporting our main hypothesis that different tones in the change in sentiment are priced differently with a positive (negative) premium for positive (negative) sentiment.

A rich set of robustness checks confirms our results and provides valuable insights into the mechanics of sentiment formation and its relationships to asset prices. First, we fail to corroborate the findings from Kozak et al. (2018) that sentiment should only be relevant if correlated to fundamental risk factors as it would be otherwise subject to arbitrage. This does not apply if a sentiment risk premium persists on aggregated markets as shown theoretically by Shefrin and Belotti (2008). Our evidence suggests that (orthogonal) sentiment is a separately priced, global risk factor and not a state variable that changes the pricing of fundamental variables. Second, our global sentiment theory is further reinforced as signed measures of change in sentiment deviations from long-run means also improve the performance of linear factor models applied to markets whose individual sentiment is not considered during the factor construction. Third, social media-driven sentiment provides stronger signals than news-only or aggregate indicators that reflect both, indicating a potentially stronger sentiment bias by retail investors. Fourth, sentiment provides additional explanatory power for the cross-section of excess stock returns in all market regimes and is not limited to crisis periods only. Recent technological enhancements and wider news and social media coverage have increased the measurable contribution of sentiment to the explained variation of various asset-pricing

models. Finally, we show that our findings are robust to different specifications without orthogonalization and additional controls for idiosyncratic risks, currency returns, and momentum.

Further work could also explore whether MarketPsych's sentiment indicators may be related to trading volume and stock price volatility (as in Jiao et al., 2020) where we expect that sentiment may carry a significant negative relation to trading volume and volatility across markets, in line with evidence from psychology, where negative sentiment causes investors to trade and speculate more, as they look to overcome their negative sentiment with a positive outcome from an alternative activity. In fact, it would be intriguing to test such associations separately across positive, neutral, and negative sentiment factors. We leave this for additional future, exciting research endeavors.

A Supplementary tables

Table A.1: Global equity indices list

This table contains the full set of international equity indices with asset code and a short description, split into developed (Panel A) and emerging markets (Panel B), and a control set of composite and mid-cap equity indices (Panel C).

Asset Code	Description	Resembling Index
Panel A: Developed equity markets		
AU500	Top 500 Australian companies	ASX All Ordinaries
CA250	Top 250 Canadian companies	S&P/TSX Composite
CH20	Top 20 Swiss companies	Swiss Market
DE30	Top 30 German companies	DAX 30
ES35	Top 35 Spanish companies	IBEX 35
FR40	Top 40 French companies	CAC 40
GB100	Top 100 UK companies	FTSE 100
HK50	Top 50 Hong Kong companies	Hang Seng
JP225	Top 225 Japanese companies	Nikkei 225
SG30	Top 30 Singaporean companies	FTSE Straits Times
US500	Top 500 US companies	S&P 500
Panel B: Emerging equity markets		
BR50	Top 50 Brazilian companies	IBRX 50
CN300	Top 300 Chinese companies	CSI 300
IN50	Top 50 Indian companies	Nifty 50
RU50	Top 50 Russian companies	RTS
Panel C: Control set of international equity markets		
EM50	Top 50 Emerging markets companies	MSCI 50
EU50	Top 50 Pan-European companies	EURO STOXX 50
US30	Top 30 US companies	Dow Jones Industrial Avg
USMID2000	Ranks 2001-3000 of US companies	Russell 2000
USS100	Top 100 Nasdaq companies	Nasdaq 100
GBMID250	Ranks 101-350 of UK companies	FTSE Mid 250

Table A.2: Pearson correlation analysis between predictor Variables

This table shows the Pearson correlation between the two global portfolio sorts based on postive $s(+)$ and negative $s(-)$ sentiment change, the long-positive/short-negative factor (PMNSNT), the market risk premium (MRP), the four developed Fama-French factors (SMB, HML, RMW, CMA), and the Momentum factor (MOM). ***, **, and * indicate significance at the 0.1%, 1%, and 5% level.

	$s(+)$	$s(-)$	PMNSNT	MRP	SMB	HML	RMW	CMA	MOM
$s(+)$	1.00	0.73***	0.25***	0.56***	0.12***	-0.06*	-0.30***	-0.35***	-0.18***
$s(-)$		1.00	-0.48***	0.58***	0.05	-0.07**	-0.29***	-0.34***	-0.19***
PMNSNT			1.00	-0.11***	0.08***	0.02	0.03	0.04	0.04
MRP				1.00	-0.30***	-0.13***	-0.42***	-0.47***	-0.30***
SMB					1.00	0.08**	-0.03	0.07**	0.23***
HML						1.00	0.05	0.65***	-0.16***
RMW							1.00	0.23***	0.27***
CMA								1.00	0.18***
MOM									1.00

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

B Details on Refinitiv-MarketPsych's sentiment construction

MarketPsych's language processing engine goes beyond traditional textual sentiment analysis based on a one-dimensional output of positive or negative sentiment and a notion of neutrality, but exploits a broad range of human emotions. A common classification system of human emotions uses two dimensions known as valence and arousal[35] as psychological research has demonstrated that more than just one dimension has predictable effects on investor behavior (see, e.g., Peterson, 2007; Shu, 2010). Besides positivity or negativity in terms of valence, the level of arousal has been shown to map directly to cognitive performance through an inverse U-curve relationship, the so-called Yerkes-Dodson's Law, capturing both the reduction in complex problem solving skills when stress levels are high and the reduction in attention and reaction times when arousal levels are low (see, e.g., Diamond, Campbell, Park, Halonen, & Zoladz, 2007; Yerkes & Dodson, 1908).

MarketPsych uses this classification system following the affective circumplex model of sentiment by Russell (1980) and constructs RMI indicators spanning the entire multi-dimensional plane of human emotions. Among others, their aggregated sentiment reflects notions of fear, optimism, and joy. As argued by Shen, Najand, Dong, and He (2017), those are the three most commonly documented emotions in the finance literature. Optimism is generally defined as the tendency to overestimate the future payoffs of a financial asset. Such an overconfidence from investors may result in deviations of asset prices from intrinsic values as observed during extreme bullish or overheated markets. Odean (1998) finds that overconfidence leads to the entry in the market by retail investors, driving up liquidity. Ciccone (2003) reports lower returns for firms characterized by optimistic versus those driven by pessimistic expectations. On the contrary, fear leads to demand shocks, driven by investors' emotional stress, increasing market uncertainty and volatility. Da et al. (2015) establish a daily fear index based on the online searches of U.S. households, predicting

[35] Valence refers to the positive or negative affectivity, while arousal measures the level of calmness or excitement of a statement or news item.

return reversals and volatility. Strongly negative emotions like anger, fear, and gloom, all of which are captured by RMI, bias human decision making and lead to a range of behaviors like herding or panic and affect trading activities, for instance by triggering either under- or over-reactions (see, e.g., Daniel, Hirshleifer, & Subrahmanyam, 1998; Lerner & Keltner, 2001; Lerner, Small, & Loewenstein, 2004; Winkielman, Berridge, & Wilbarger, 2005; Winkielman, Knutson, Paulus, & Trujillo, 2007). Wright and Bower (1992) found that pleasant emotions like bliss, joy, and optimism affect the subjective probability assessments of uncertain outcomes and therefore influence investors' decision-making as documented by Dolan (2002). Figure B.1 depicts several among the RMI sentiments that are described in detail in Table B.2 devoted to explaining the affective circumplex. Each dot in the figure corresponds to the emotion's location on the circumplex, whereby RMI indicators are themselves hybrids of multiple emotions according to the original framework. The thin grey line connects the positive and negative poles of matching indicators. The RMI sentiment indicator itself spans the entire multi-dimensional plane of the circumplex as described in detail in Table B.1. The table shows that RMI's construction of the sentiment indicator is tilted towards capturing negative statements because, as confirmed in private exchanges, MarketPsych research on business and financial language has found a prevalence of concepts with negative versus positive valence. As a result, the sentiment indicator is usually negative in net terms.

To provide insights about the mechanism of sentiment indicator construction, MarketPsych provides an example on the complex language processing system that reveals how they address some common pitfalls in news and social media sentiment analysis. Figure B.2 evaluates the opinion of a Goldman Sachs' analyst about his expectations of the following day's quarterly call of Apple Inc. of increasing profit margins. MarketPsych is able to differentiate between forward-looking statements and general chatter by breaking down concepts into forecasts (future tense) versus present or past observations. For instance, the PriceForecast category is a future-tense subset of PriceDirection. "The price of Apple rose last week" is a PriceDirection-only reference while "The price of Apple will rise" would be attributed to both PriceDirection and PriceForecast.

Table B.1: Construction of the Refinitiv-MarketPsych sentiment indicator

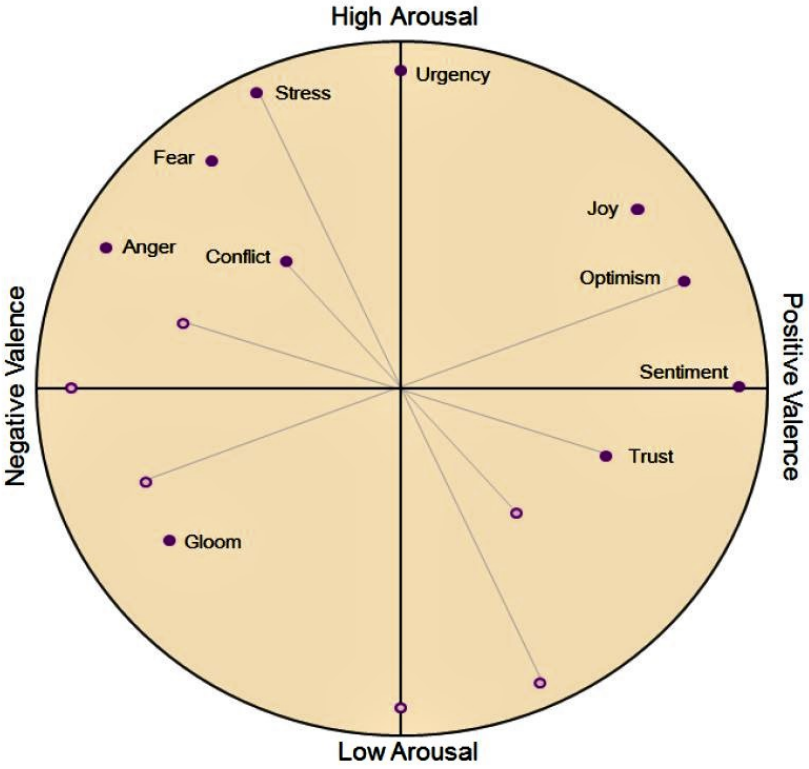
This table provides the construction details for the Refinitiv-MarketPsych (RMI) Sentiment indicator as the net of positive and negative references along the valence-arousal classification system.

Positive References	Negative References
Positive	Negative
AccountingGood	AccountingBad
Upgrade	Downgrade
EconomicPositive	EconomicNegative
EconomistPositive	EconomistNegative
EconomicActorsPositive	EconomicActorsNegative
ManagementGood	ManagementBad
BullVerbs	BearVerbs
ExcitementPos	
FearDown	FearUp
AngerDown	AngerUp
HappyUp	HappyDown
GloomDown	GloomUp
OptimismUp	OptimismDown
PessimismDown	PessimismUp
LoveUp	LoveDown
HateDown	HateUp
InnovativeUp	InnovativeDown
EarningsSurprisePos	EarningsSurpriseNeg
EarningsUp	EarningsDown
EarningsExpectationsUp	EarningsExpectationsDown
EarningsGuidanceUp	EarningsGuidanceDown
GuidanceUp	GuidanceDown
	ProfitWarning
	CatastropheConcept
	DeclareBankruptcy

Source: MarketPsych

Figure B.1: Sentiment classification system: valence and arousal

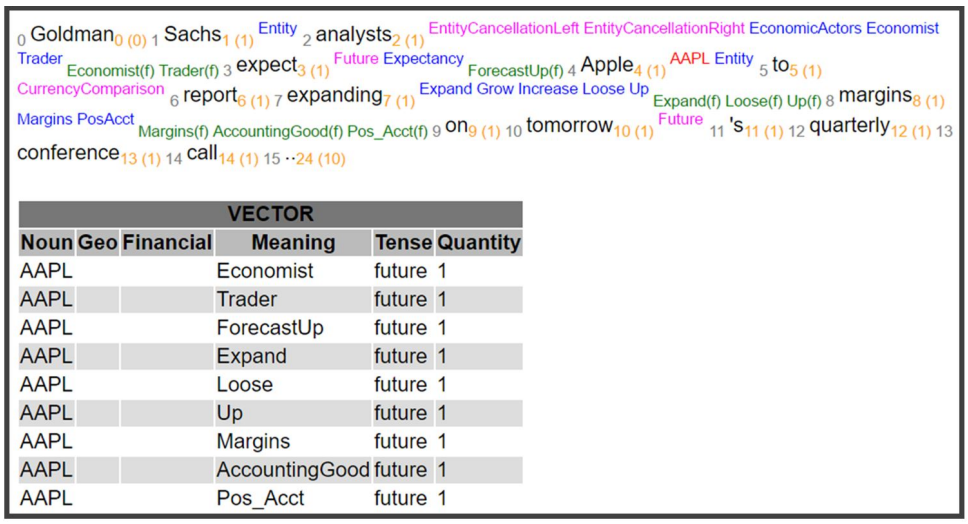
This figure plots a common classification system for human emotions along two dimensions: valence and arousal. MarketPsych uses this classification system following the affective circumplex model of sentiment by Russell (1980) and constructs RMI indicators spanning the plane of human emotions. The figure depicts several of the RMI sentiment indicators on the affective circumplex. Each dot corresponds to the emotion's location on the circumplex, whereby RMI indicators are themselves hybrids of multiple emotions according to the original framework. The grey lines connect the positive and negative poles of matching indicators.



Source: MarketPsych

Figure B.2: Example of MarketPsych's human language processing system

This figure depicts an example of how MarketPsych processes news and evaluates human emotions. Each term is annotated by MarketPsych. Complex meanings such as $AccountingGood_f$ are extracted. This is a forward looking assessment based on the attribute "tomorrow". "Goldman Sachs" is ignored as an irrelevant entity because it relates to the analyst, while "Apple" is correctly recognized as the object of interest. MarketPsych differentiates between value-adding statements as above versus irrelevant terms. Those irrelevant terms are excluded from the score vector and are not used in RMI calculations.



VECTOR					
Noun	Geo	Financial	Meaning	Tense	Quantity
AAPL			Economist	future	1
AAPL			Trader	future	1
AAPL			ForecastUp	future	1
AAPL			Expand	future	1
AAPL			Loose	future	1
AAPL			Up	future	1
AAPL			Margins	future	1
AAPL			AccountingGood	future	1
AAPL			Pos_Acct	future	1

Source: MarketPsych

In order to have a correct attribution of articles to the right time window MarketPsych also limits article consumption to those less than 2,500 words as longer articles usually take longer to write and are unlikely to be timely. In order to avoid the impact of stale news, content that has been published more than 24 hours before a given time t, is excluded. That means all content drops out of the 24 hours averages when it has been more than 24 hours since its publication. Articles that are more than 98% similar to articles recorded in the previous 24 hours are removed from analysis to avoid double-counting.[36]

Various sources, though limited to English, are used to inform the data feed of the language processing system used by MarketPsych. These include news publishers like Refinitiv and Bloomberg, electronic databases like the U.S. Securities and Exchange

[36]In the case of social media for which the concepts of re-tweeting, re-posting, and commenting are defined, MarketPsych employs a rigorous approach to cleanse the data. RMI indicators do not include re-tweets, unless they include additional commentary or remarks about the original tweet. RMI does not include comments with the same title that are repeated multiple times; however, they do include commentary text when it changes from post to post.

Table B.2: Description of Refinitiv-MarketPsych indicators

This table provides a detailed description of Refinitiv-MarketPsych (RMI) indicators to better understand the aggregated sentiment measure.

RMI COMMON NAME	ANTICIPATED MARKET IMPACT
Sentiment	There are several important research findings related to sentiment and price movement. Based on academic research on Thomson Reuters News Analytics sentiment scores, positive and negative sentiment in the news about individual stocks extend price momentum, which is supported by additional evidence that traders collectively under-react to negative sentiment in news reports. Another study finds that market sentiment improves factor weighting in some models. In foreign exchange, news sentiment was found to influence volatility.
Optimism	There is empirical evidence that proxies for optimism correlate with positive price behavior and that bullish comments in financial social media precede higher trading volume. Optimism in earnings press releases was found correlated with future stock price activity.
Fear	Academic researchers who aggregated search terms they deemed reflective of economic fear found short-term mean reversion in prices when fear-related search terms spiked in volume. In experimental markets, fear was found to decrease bid and increase ask prices, leading to less overall trading activity. As a result, we expect wider bid-ask spreads when fear is high.
Joy	Joy is a marker of exuberance. Experimental markets demonstrate higher price peaks and larger collapses during bubble simulations if traders watched a positively exciting movie clip before trading begins.
Trust	Trust was designed specifically for nations and banking and financial groups. Economists have found that national interpersonal Trust levels correlate with future economic growth.
Conflict	The Conflict RMI, which is intended to capture disagreement and dispute, is anticipated to correlate with price volatility. A study of international markets found that global conflicts significantly impact asset prices.
Stress and Urgency	Urgency and Stress are high-arousal indices that vary in valence. Based on evidence that arousal drives cognitive performance in an inverse-U shaped curve, we infer that pricing anomalies are more likely to emerge at low or high arousal values, as seen with both high positive and high negative arousal during research into experimental market bubbles.
Uncertainty	Researchers found that high-uncertainty equities and country indices on average outperform their low-uncertainty peers. 39 In contrast, during speculative bubbles uncertainty amplifies the price momentum of positive sentiment. In emerging fixed income markets, releases of macroeconomic data decrease future volatility.
Gloom	Traders in an experimental market offered lower ask and high bid prices when "sadness" was induced prior to trading, leading to increased transaction volume. If this result transfers into larger market behavior, we expect increased trading volume during periods of high Gloom. Researchers speculate that identified semi-annual variations in country stock index returns - which scale by latitude and reverse from northern to southern hemispheres - may be caused by seasonal changes in affect (the "winter blues") among local traders.
Anger	Traders induced to feel anger in an experimental market decrease both average ask and bid prices. As a result, we speculate that higher RMI Anger readings should lead to increased selling and reduced buying in associated assets, leading to downward pressure on prices during high Anger periods.

Source: MarketPsych

Commission's Edgar repository of company filings, direct press releases by companies, transcripts of conference calls, websites, blogs, and especially posts on social media like Twitter and Yahoo's stock message boards. We use the aggregate measure that reflects activities through all types of channels, news and social media. The indicators are updated at a one-minute frequency and the system works 24/7, continuously scanning all the tracked sources. In order to construct a daily record, 24 hours or 1440 minutes are aggregated into one daily observation. If no records are found for the constituents of a specific equity index, a "N/A" is returned and the observation is not stored. This implies that the retrievable time series of each individual sentiment indicator are not equally spaced over time. In practical terms, if no observation is found, no $Buzz$ is recorded and the time series fails to be updated. Crucially, such a case needs to be differentiated from true "0" values, where positive and negative statements concerning an asset exactly balance each other.[37]

For the purposes of our investigation, we cumulate the RMI index at a lower, weekly frequency thus aggregating the original, higher daily frequency using Equation (7). A weekly frequency appears to strike a reasonable balance between a sufficient granularity of the data and a need to control for the risk of using a noisy estimator of sentiment.[38] There are no missing observations in our weekly sample of international equity indices. Let $Buzz_0$, $Buzz_{-1}$, $Buzz_{-(T-1)}$ and RMI_0, RMI_{-1}, $RMI_{-(T-1)}$ represent the corresponding $Buzz$ RMI data for a given equity market, content source, and timestamp over the past T days. The $Buzz$-weighted average RMI over the trailing T-day window length is then computed as:[39]

$$\frac{(Buzz_0 * RMI_0 + Buzz_{-1} * RMI_{-1} + \ldots + Buzz_{-(T-1)} * RMI_{-(T-1)})}{(Buzz_0 + Buzz_{-1} + \ldots + Buzz_{-(T-1)})}. \tag{7}$$

[37] Positive and negative references that net each other out may still signal increased uncertainty in the market and disagreement between investors and potentially lead to higher trading activity. However, in private exchanges, MarketPsych has confirmed to us that the primary relationship is that sentiment RMI variability rises as the overall $Buzz$ decreases. So $Buzz$ is the primary determinant of sentiment dispersion.

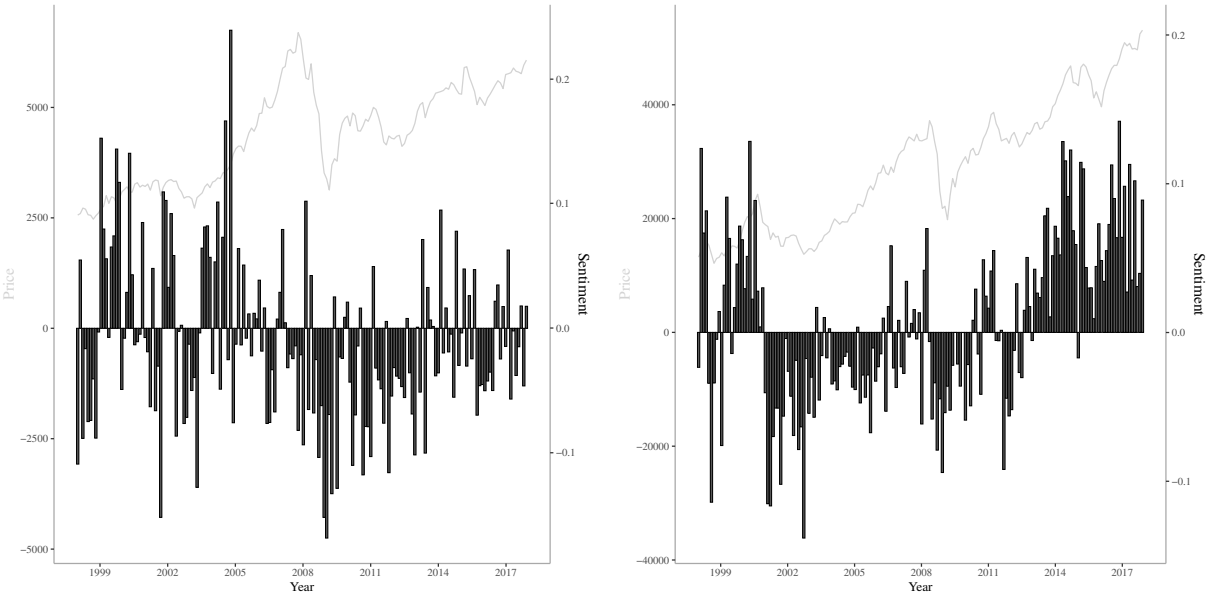
[38] In an unreported exploratory analysis, we checked that sentiment fluctuates massively at daily frequencies, whereas at a monthly frequency it suffers from a loss of valuable information that, however, appears to be manageable. This analysis is available upon request.

[39] This definition ensures comparability of sentiment between different markets as outlined by MarketPsych in their research guidelines, accessible at `https://old.marketpsych.com/guide/`.

The next figures show the sentiment level for each international equity market in context of the development of the country's main equity index.

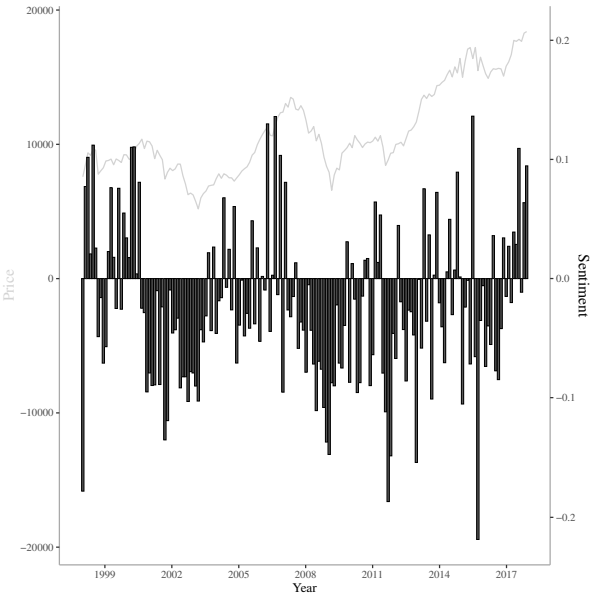
Figure B.3: Equity markets and their local sentiment

This figure plots the development of the international equity markets (light grey) and their local sentiment (black) from January 1998 - December 2017.

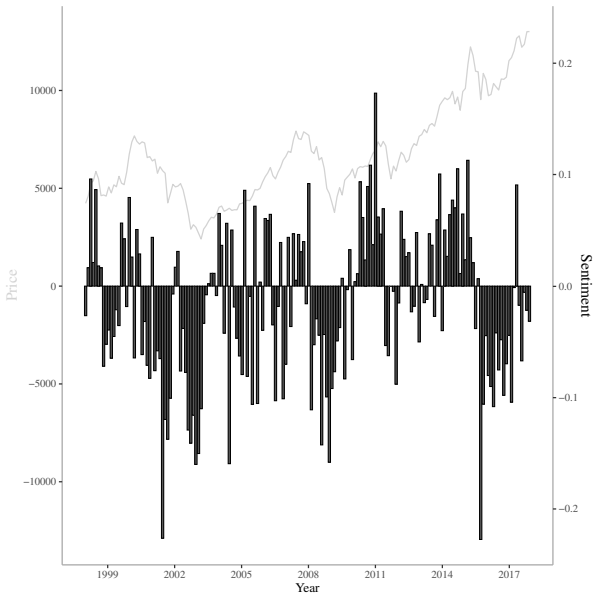


(a) AU50

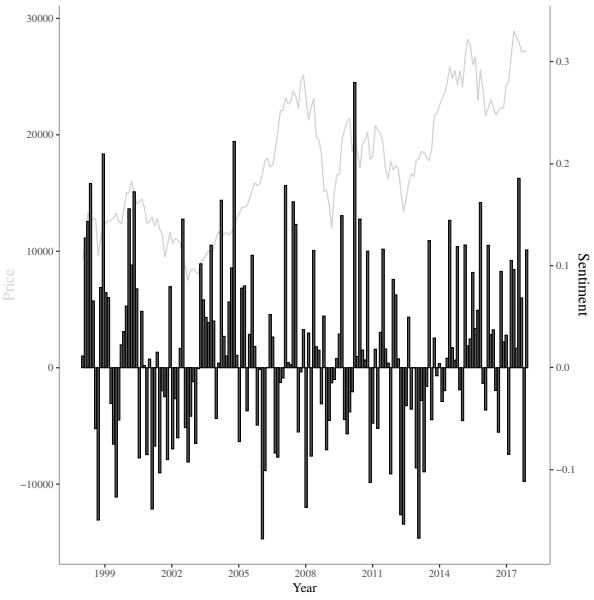
(b) CA250



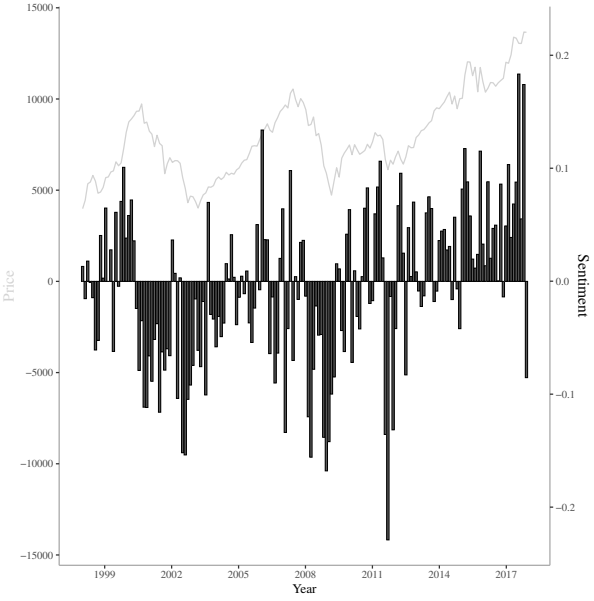
(c) CH20



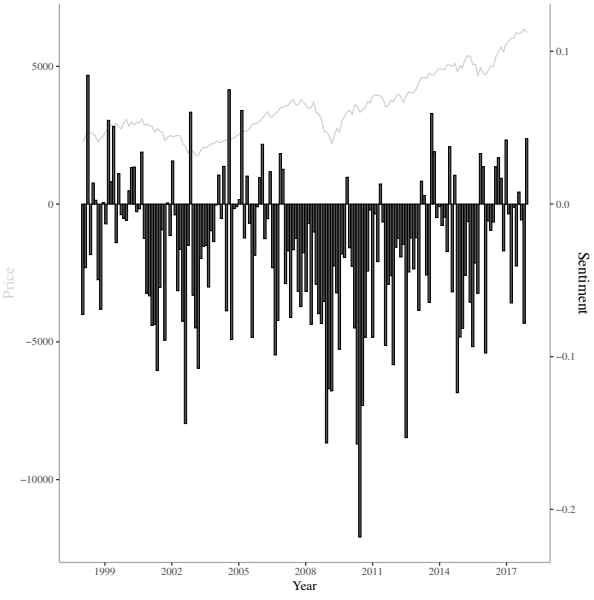
(d) DE30



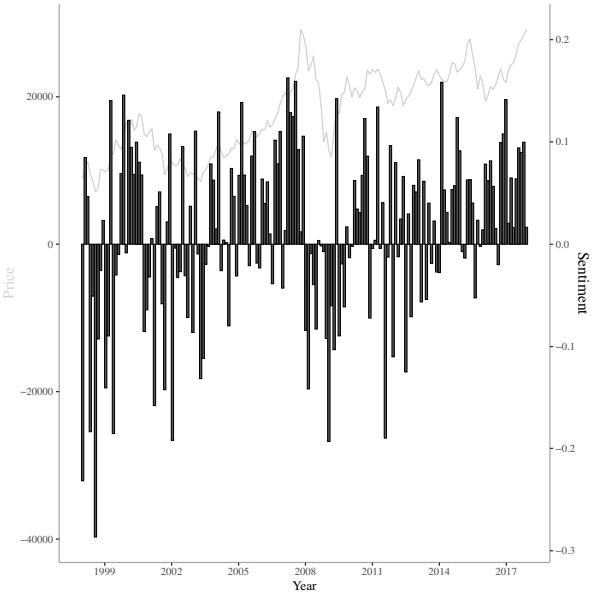
(e) ES35



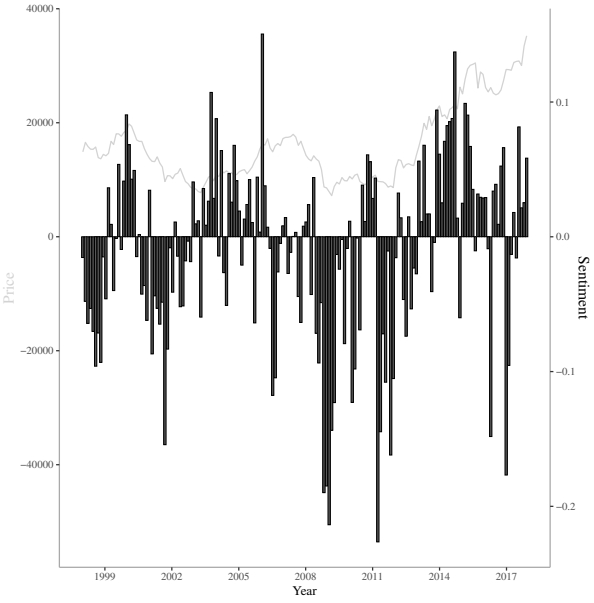
(f) FR40



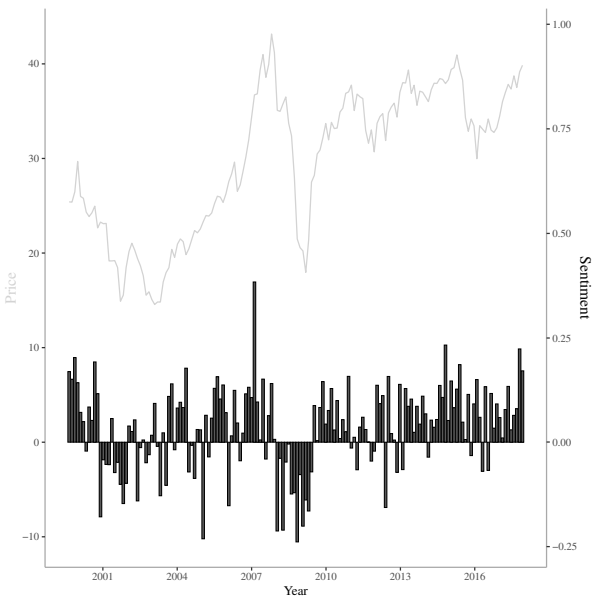
(g) GB100



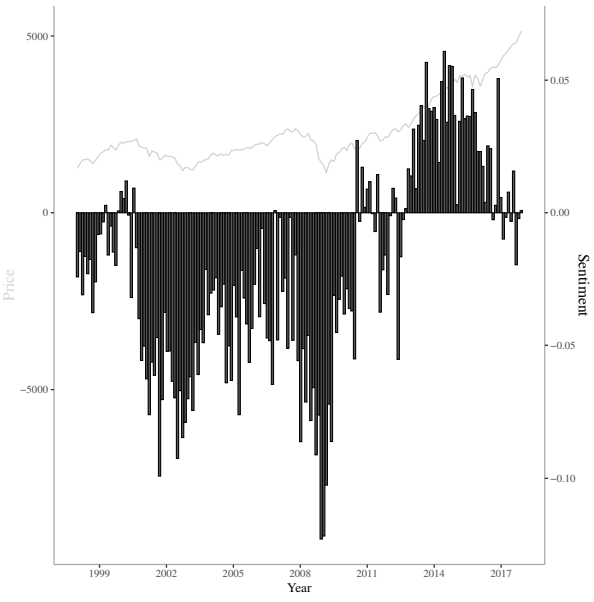
(h) HK50



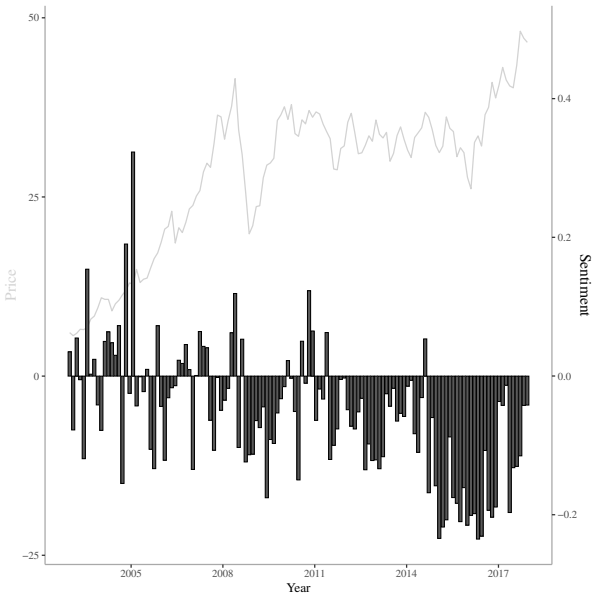
(i) JP225



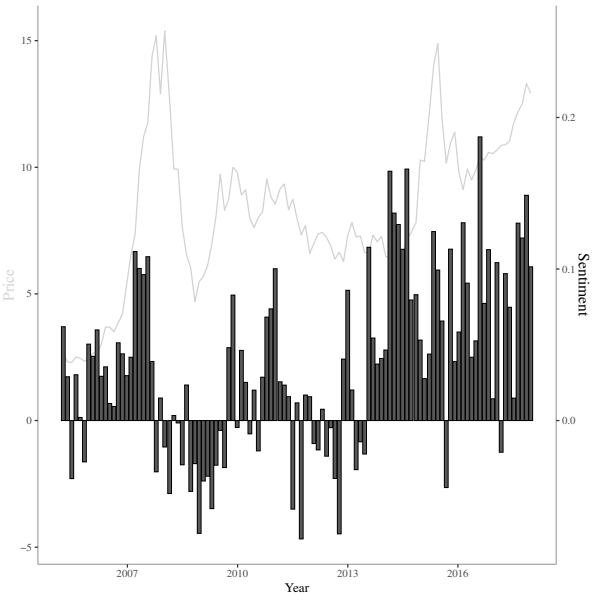
(j) SG30



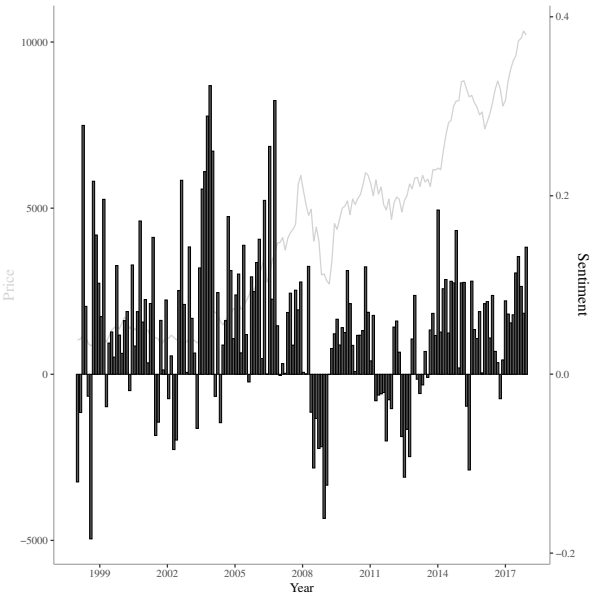
(k) US500



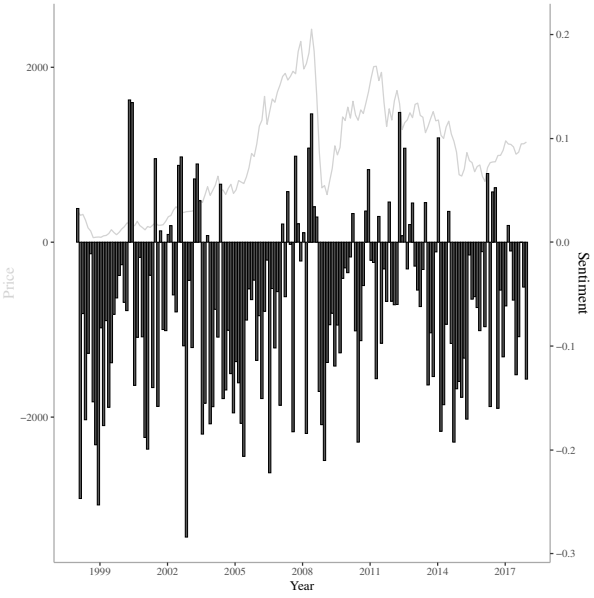
(l) BR50



(m) CN300



(n) IN50



(o) RU50

C Orthogonalization approach

We first orthogonalize positive sentiment by regressing its mimicking excess portfolio returns on market returns:

$$r^e_{s(+),t} = \beta_{CAPM} r^e_{m,t} + \epsilon_{s(+),t}. \tag{8}$$

The residuals represent our positive sentiment indicator when excess market returns are accounted for. Next, we orthogonalize negative sentiment by regressing the mimicking returns on market excess returns and the residuals from Equation (8):

$$r^e_{s(-),t} = \beta_{CAPM} r^e_{m,t} + \beta_{\epsilon_{s(+)}} \hat{\epsilon}_{s(+),t} + \epsilon_{s(-),t}. \tag{9}$$

In case of the FF5 model, the same logic applies. Positive and negative sentiment is then orthogonal to all five fundamental factors: MRP, SMB, HML, RMW, and CMA.

D Market efficiency test

Table D.1: Market efficiency tests

This table shows results for the Spearman rank correlation coefficients for two tests. The first column the aggregated relative importance of all sentiment variables of the SNT-CAPM model with the R^2 of the CAPM market model, known as a naïve market efficiency test. The remaining columns correlate the same aggregated importance of sentiment with the results of the Lo-MacKinlay variance ratio tests for lags $k = 2, 5, 10$. $M1(k)$ refers to the null hypothesis H_0 of a random walk with homoskedastic increments, while $M2(k)$ concerns the heteroskedastic increments. ***, **, and * indicate significance at the 0.1%, 1%, and 5% level.

CAPM	**Lo-MacKinlay variance ratio test**					
	M1(2)	M1(5)	M1(10)	M2(2)	M2(5)	M2(10)
-0.72***	0.81***	0.86***	0.87***	0.58*	0.83***	0.87***

E Additional tables

This sections contains additional tables not shown in the main manuscript.

Table E.1: Comparing the Refinitv-MarketPysch sentiment indicators with alternative sentiment measures

This table compares the sentiment-augmented CAPM based on RMI with benchmark models including a variety of different sentiment measures. Panel A concerns results for the sentiment-augmented CAPM with only negative sentiment (SNT-CAPM). Panel B adds the orthogonalized BW index proposed by Baker and Wurgler (2006) to the market risk premium factor (BW-CAPM). Panel C shows an improvement of the BW index by Huang et al. (2015), separating information from noise using partial least squares (PLS-CAPM). Panel D uses the downside risk model (DR-CAPM) from Lettau et al. (2014). Panel E reports the results for an extended CAPM including the FEARS index (FEARS-CAPM) based on 30 negative terms from Da et al. (2015). In Panel F we use the event sentiment score ESS from RavenPack based on the country under analysis for a RAVENPACK-CAPM. The table reports the coefficient estimates of the sentiment variables, Newey-West corrected standard errors to adjust for autocorrelation and heteroscedasticity in returns in parentheses and the adjusted R^2. The regressions are estimated for each equity index individually on weekly returns from July 2004 to December 2011, restricted by the limited availability of the FEARS index.

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel A: NSNT-CAPM															
Intercept	0.0002	0.0015**	0.0005	0.0001	0.0023*	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017**	0.0002	0.0009	0.0000	0.0004
	(0.0003)	(0.0007)	(0.0004)	(0.0005)	(0.0017)	(0.0005)	(0.0005)	(0.0005)	(0.0003)	(0.0007)	(0.0007)	(0.0007)	(0.0016)	(0.0005)	(0.0003)
$s(-)$	0.4499***	0.9661***	0.4715***	0.5155***	0.4988***	0.7089***	0.7033***	0.6468***	0.5104***	0.8032***	0.6851***	0.6418***	1.4200***	0.7532***	0.4054***
	(0.0658)	(0.0546)	(0.0434)	(0.0463)	(0.0999)	(0.0691)	(0.0609)	(0.0644)	(0.0457)	(0.0840)	(0.0805)	(0.0876)	(0.1240)	(0.0483)	(0.0508)
Adj. R^2	46.19	62.43	57.03	53.68	8.48	62.78	57.24	65.85	65.26	48.11	35.65	37.32	44.13	53.21	69.07
Panel B: BW-CAPM															
Intercept	-0.0001	0.0024	0.0007	0.0000	0.0027	0.0004	0.0004	-0.0004	0.0006	0.0006	0.0025	-0.0004	0.0020	0.0005	0.0000
	(0.0008)	(0.0012)	(0.0007)	(0.0007)	(0.0027)	(0.0008)	(0.0010)	(0.0007)	(0.0004)	(0.0011)	(0.0015)	(0.0014)	(0.0023)	(0.0010)	(0.0005)
BW	0.3908***	0.8094***	0.5371***	0.4378***	0.1927**	0.6232***	0.5719***	0.6058***	0.5357***	0.5558***	0.5289***	0.4883***	0.8583***	0.4653***	0.5999***
	(0.0300)	(0.0599)	(0.0364)	(0.0381)	(0.0828)	(0.0325)	(0.0581)	(0.0335)	(0.0178)	(0.0493)	(0.0578)	(0.0532)	(0.1094)	(0.0508)	(0.0205)
Adj. R^2	25.02	41.60	41.75	32.23	1.73	42.48	33.65	42.94	45.02	25.12	20.88	19.28	24.38	26.59	54.06

* p ≤ 0.1, ** p ≤ 0.05, *** p ≤ 0.01

Table continues on the next page

Table E.1 continues here:

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel C: PLS-CAPM															
Intercept	-0.0014	0.0011	-0.0004	-0.0004	-0.0006	-0.0003	-0.0009	-0.0010	0.0006	-0.0012	0.0007	-0.0016	-0.0008	-0.0014	-0.0004
	(0.0012)	(0.0014)	(0.0010)	(0.0009)	(0.0026)	(0.0010)	(0.0010)	(0.0009)	(0.0007)	(0.0015)	(0.0016)	(0.0016)	(0.0032)	(0.0013)	(0.0008)
PLS	0.3772***	0.7947***	0.5261***	0.4325***	0.1498*	0.6155***	0.5573***	0.5996***	0.5360***	0.5378***	0.5109***	0.4765***	0.8306***	0.4473***	0.5965***
	(0.0344)	(0.0599)	(0.0321)	(0.0418)	(0.0824)	(0.0311)	(0.0483)	(0.0347)	(0.0186)	(0.0552)	(0.0595)	(0.0491)	(0.1006)	(0.0483)	(0.0251)
Adj. R^2	26.19	42.04	42.49	32.40	4.68	42.73	34.50	43.12	44.98	26.25	21.90	19.77	25.47	28.27	54.13
Panel D: DR-CAPM															
Intercept	0.0011	0.0030	0.0013	0.0001	0.0036	0.0016	0.0005	0.0002	0.0008	0.0008	0.0040**	0.0006	0.0032	0.0013	0.0001
	(0.0009)	(0.0015)	(0.0009)	(0.0010)	(0.0032)	(0.0012)	(0.0012)	(0.0011)	(0.0009)	(0.0014)	(0.0017)	(0.0017)	(0.0025)	(0.0011)	(0.0008)
DR.Intercept	-0.0093	-0.0059	0.0006	-0.0002	-0.0003	-0.0028	-0.0121	-0.0096	-0.0057	0.0014	-0.0061	0.0005	-0.0088	-0.0007	0.0014
	(0.0059)	(0.0077)	(0.0083)	(0.0076)	(0.0101)	(0.0070)	(0.0072)	(0.0063)	(0.0062)	(0.0074)	(0.0095)	(0.0082)	(0.0201)	(0.0061)	(0.0033)
$DR \times MRP$	-0.0366	-0.0530	0.0907	0.0058	0.1180	0.1134	-0.2678	-0.1366	-0.1118	0.0681	0.0828	0.1599	-0.0334	0.1054	0.0552
	(0.1310)	(0.2091)	(0.1893)	(0.2458)	(0.2461)	(0.2551)	(0.2190)	(0.2249)	(0.1976)	(0.1684)	(0.2287)	(0.2115)	(0.4985)	(0.1521)	(0.0876)
Adj. R^2	25.88	41.53	41.73	32.06	1.24	42.90	34.12	43.25	45.04	24.89	21.28	19.39	24.26	26.44	53.96
Panel E: FEARS-CAPM															
Intercept	-0.0002	0.0023	0.0006	0.0000	0.0026	0.0004	0.0004	-0.0004	0.0006	0.0005	0.0023	-0.0004	0.0020	0.0004	0.0000
	(0.0008)	(0.0012)	(0.0007)	(0.0007)	(0.0026)	(0.0008)	(0.0009)	(0.0007)	(0.0006)	(0.0012)	(0.0015)	(0.0014)	(0.0024)	(0.0011)	(0.0005)
FEARS	0.3920***	0.8115***	0.5413***	0.4376***	0.1916**	0.6238***	0.5725***	0.6068***	0.5379***	0.5610***	0.5329***	0.4888***	0.8624***	0.4703***	0.6004***
	(0.0302)	(0.0637)	(0.0348)	(0.0397)	(0.0871)	(0.0329)	(0.0448)	(0.0337)	(0.0201)	(0.0513)	(0.0584)	(0.0537)	(0.1154)	(0.0527)	(0.0234)
Adj. R^2	25.20	41.88	42.56	32.23	1.61	42.51	33.68	43.01	45.13	25.67	21.30	19.26	24.34	26.85	54.04
Panel F: RAVENPACK-CAPM															
Intercept	0.0045	0.0084	0.0024	-0.0026	-0.0024	-0.0029	-0.0226*	-0.0028	0.0048	-0.0283	0.0012	0.0179	0.0078	0.0068	0.0123*
	(0.0036)	(0.0085)	(0.0030)	(0.0109)	(0.0085)	(0.0084)	(0.0096)	(0.0064)	(0.0031)	(0.0217)	(0.0103)	(0.0141)	(0.0118)	(0.0086)	(0.0054)
ESS	0.4375***	0.5635***	0.5135***	0.1891*	0.1028	0.6369***	0.6679***	0.7399***	0.5043***	0.6069***	0.5163***	0.4416***	0.8931***	0.3026***	0.5821***
	(0.0761)	(0.0634)	(0.0328)	(0.0737)	(0.0696)	(0.0527)	(0.1493)	(0.0528)	(0.0298)	(0.1561)	(0.0500)	(0.0979)	(0.1713)	(0.0702)	(0.0213)
Adj. R^2	30.45	25.42	39.80	6.12	0.22	47.46	38.77	50.78	48.96	24.17	23.20	10.27	31.64	15.43	53.92

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table E.2: Comparing the CAPM with local and global sentiment

This table compares local and global sentiment factors. Panel A augments the CAPM with the RMI sentiment change to the long-term average since inception for each individual equity market, thus representing a CAPM with local sentiment ($r_i^e = \beta_{CAPM,i} r_m^e + \beta_{SNT,i} \hat{\epsilon}_{SNT,i}$). Panel B adds our two global sentiment factors $s(+)$ and $s(-)$ based on portfolio sorts to the local sentiment factors, thus showing another CAPM specification of total (global and local) sentiment ($r_i^e = \beta_{CAPM,i} r_m^e + \beta_{SNT,i} \hat{\epsilon}_{SNT,i} + \beta_{s(+),i} \hat{\epsilon}_{s(+)} + \beta_{s(-),i} \hat{\epsilon}_{s(-)}$). Panel C and D construct an alternative global sentiment factor by applying a principal component analysis (PCA) on the local sentiment change and either uses only the first (PC1) or the first two components (PC2). The factor loadings are shown in Table E.3. The table reports the coefficient estimates, the Newey-West corrected standard errors to adjust for autocorrelation and heteroscedasticity in returns in parentheses and the adjusted R^2. We also report the relative importance indicator for each coefficient, which estimates its relative contribution to the total explained variation. The regressions are estimated for each equity index individually on weekly returns from 1998 to 2017.

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel A: CAPM with local sentiment change															
Intercept	0.0002	0.0015*	0.0005	0.0001	0.0023*	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017**	0.0002	0.0009	0.0000	0.0004
	(0.0005)	(0.0008)	(0.0005)	(0.0005)	(0.0012)	(0.0006)	(0.0006)	(0.0006)	(0.0005)	(0.0007)	(0.0007)	(0.0007)	(0.0014)	(0.0006)	(0.0004)
MRP	0.3206***	0.7534***	0.5059***	0.5074***	0.2060***	0.6743***	0.6076***	0.6369***	0.5159***	0.5197***	0.4208***	0.4664***	0.8088***	0.4012***	0.5996***
	(0.0194)	(0.0376)	(0.0204)	(0.0227)	(0.0524)	(0.0273)	(0.0278)	(0.0242)	(0.0196)	(0.0317)	(0.0321)	(0.0317)	(0.0600)	(0.0269)	(0.0172)
localSNTchg	0.0438***	0.0418***	0.0444***	0.0339***	0.1410***	0.0409***	0.0474***	0.0429***	0.0267***	0.0958***	0.0807***	0.0676***	0.0765***	0.0331***	0.0507***
	(0.0073)	(0.0096)	(0.0087)	(0.0088)	(0.0220)	(0.0092)	(0.0074)	(0.0083)	(0.0094)	(0.0091)	(0.0082)	(0.0117)	(0.0171)	(0.0065)	(0.0108)
Adj. R^2	22.84	34.91	37.94	33.02	7.67	37.64	33.15	40.88	40.28	26.61	20.37	19.21	16.14	20.50	54.19
RI-MRP	88.34	95.53	95.92	97.11	27.36	96.87	92.07	96.29	98.86	70.61	64.02	86.72	90.08	89.67	98.20
RI-localSNTchg	11.66	4.47	4.08	2.89	72.64	3.13	7.93	3.71	1.14	29.39	35.98	13.28	9.92	10.33	1.80
Panel B: CAPM with both global and local sentiment															
Intercept	0.0002	0.0015**	0.0005	0.0001	0.0023*	0.0002	0.0002	0.0003	0.0002	0.0005	0.0017***	0.0002	0.0010	-0.0001	0.0004
	(0.0003)	(0.0006)	(0.0004)	(0.0004)	(0.0012)	(0.0004)	(0.0005)	(0.0004)	(0.0003)	(0.0005)	(0.0006)	(0.0006)	(0.0011)	(0.0004)	(0.0003)
MRP	0.3197***	0.7531***	0.5054***	0.5074***	0.2055***	0.6743***	0.6055***	0.6360***	0.5150***	0.5171***	0.4221***	0.4610***	0.8106***	0.4009***	0.5990***
	(0.0146)	(0.0276)	(0.0154)	(0.0185)	(0.0505)	(0.0192)	(0.0210)	(0.0168)	(0.0138)	(0.0235)	(0.0274)	(0.0258)	(0.0466)	(0.0188)	(0.0129)
localSNTchg	0.0113*	0.0012	0.0114	0.0176**	0.0663***	0.0286***	0.0150**	0.0199***	0.0100	0.0167**	0.0208***	0.0227**	0.0313**	0.0111**	0.0295***
	(0.0054)	(0.0070)	(0.0065)	(0.0071)	(0.0209)	(0.0064)	(0.0055)	(0.0057)	(0.0065)	(0.0066)	(0.0070)	(0.0094)	(0.0133)	(0.0045)	(0.0080)
s(+)	0.5287***	0.9020***	0.5573***	0.4532***	0.6444***	0.7452***	0.7071***	0.6663***	0.5317***	0.9502***	0.7415***	0.7719***	1.4412***	0.7676***	0.4558***
	(0.0199)	(0.0390)	(0.0208)	(0.0251)	(0.0715)	(0.0259)	(0.0283)	(0.0227)	(0.0186)	(0.0322)	(0.0371)	(0.0349)	(0.0628)	(0.0262)	(0.0175)
s(-)	0.2439***	0.6854***	0.2536***	0.4066***	0.0583	0.4558***	0.4838***	0.4315***	0.3372***	0.4272***	0.4146***	0.3371***	0.9791***	0.4334***	0.2380***
	(0.0226)	(0.0527)	(0.0238)	(0.0286)	(0.1018)	(0.0297)	(0.0324)	(0.0260)	(0.0213)	(0.0362)	(0.0424)	(0.0399)	(0.0720)	(0.0347)	(0.0200)
Adj. R^2	56.19	64.99	64.94	55.52	14.41	69.23	62.04	71.40	70.52	60.07	42.05	46.29	49.50	60.99	74.22
RI-MRP	35.89	51.40	56.03	57.78	14.45	52.78	49.02	55.08	56.38	31.14	31.39	35.57	29.77	30.31	71.58
RI-localSNTchg	1.36	0.09	0.44	1.00	11.43	1.31	0.85	0.81	0.33	2.32	2.19	1.36	0.39	1.16	0.65
RI-s(+)	54.01	36.86	37.60	25.63	73.87	35.76	36.95	33.48	33.16	57.66	53.59	55.06	51.75	58.03	23.03
RI-s(-)	8.74	11.65	5.92	15.60	0.24	10.15	13.18	10.63	10.13	8.88	12.83	8.01	18.09	10.50	4.74

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table continues on the next page

Table E.2 continues here

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel C: CAPM with PCA (PC1) sentiment															
Intercept	0.0001	0.0010	0.0005	0.0006	0.0023*	0.0008	0.0003	0.0003	0.0006	0.0006	0.0019**	0.0012	0.0001	0.0002	0.0008*
	(0.0005)	(0.0009)	(0.0005)	(0.0005)	(0.0018)	(0.0006)	(0.0008)	(0.0005)	(0.0004)	(0.0009)	(0.0010)	(0.0010)	(0.0015)	(0.0007)	(0.0004)
MRP	0.3321***	0.7222***	0.4887***	0.4144***	0.1472**	0.5925***	0.5763***	0.5766***	0.5191***	0.4856***	0.4467***	0.4421***	0.8050***	0.3937***	0.5465***
	(0.0281)	(0.0531)	(0.0339)	(0.0357)	(0.0734)	(0.0334)	(0.0434)	(0.0290)	(0.0208)	(0.0467)	(0.0467)	(0.0466)	(0.0847)	(0.0441)	(0.0243)
PC1	0.0012***	0.0005	0.0008***	0.0010***	0.0022***	0.0011***	0.0009**	0.0012***	0.0005	0.0014***	0.0012**	0.0015***	0.0015*	0.0010***	0.0011***
	(0.0004)	(0.0006)	(0.0003)	(0.0003)	(0.0011)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0005)	(0.0006)	(0.0005)	(0.0010)	(0.0005)	(0.0002)
Adj. R^2	23.63	33.42	40.15	31.66	4.09	40.20	31.58	41.61	43.05	23.85	19.98	18.66	24.04	24.46	54.72
RI-MRP	80.44	93.04	89.35	85.32	37.87	88.15	89.38	87.20	92.32	83.73	84.74	81.42	88.45	85.06	87.65
RI-PC1	19.56	6.96	10.65	14.68	62.13	11.85	10.62	12.80	7.68	16.27	15.26	18.58	11.55	14.94	12.35
Panel D: CAPM with PCA (PC1 & PC2) sentiment															
Intercept	0.0001	0.0010	0.0005	0.0006	0.0023	0.0008	0.0003	0.0003	0.0006	0.0006	0.0020*	0.0012	0.0001	0.0003	0.0008*
	(0.0006)	(0.0010)	(0.0005)	(0.0005)	(0.0019)	(0.0005)	(0.0008)	(0.0005)	(0.0004)	(0.0009)	(0.0010)	(0.0010)	(0.0016)	(0.0007)	(0.0004)
MRP	0.3253***	0.7101***	0.4837***	0.4106***	0.1491**	0.5825***	0.5778***	0.5721***	0.5158***	0.4831***	0.4460***	0.4372***	0.7891***	0.3887***	0.5404***
	(0.0296)	(0.0545)	(0.0346)	(0.0364)	(0.0725)	(0.0336)	(0.0441)	(0.0277)	(0.0213)	(0.0484)	(0.0476)	(0.0471)	(0.0882)	(0.0449)	(0.0258)
PC1	0.0012***	0.0005	0.0008***	0.0011***	0.0022***	0.0012***	0.0009*	0.0012***	0.0005	0.0014***	0.0012**	0.0015***	0.0015*	0.0010***	0.0011***
	(0.0004)	(0.0006)	(0.0004)	(0.0002)	(0.0011)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0006)	(0.0006)	(0.0005)	(0.0011)	(0.0005)	(0.0002)
PC2	0.0009	0.0017*	0.0007	0.0005	-0.0003	0.0014**	-0.0002	0.0006	0.0005	0.0004	0.0001	0.0007	0.0022	0.0007	0.0008*
	(0.0004)	(0.0009)	(0.0004)	(0.0004)	(0.0015)	(0.0004)	(0.0007)	(0.0004)	(0.0003)	(0.0007)	(0.0008)	(0.0007)	(0.0013)	(0.0006)	(0.0002)
Adj. R^2	23.90	33.78	40.25	31.68	3.95	40.64	31.49	41.64	43.05	23.76	19.86	18.63	24.35	24.51	54.95
RI-MRP	77.20	89.63	87.43	83.66	37.94	85.03	88.87	85.69	90.87	82.63	83.96	79.63	84.71	82.90	85.65
RI-PC1	19.33	6.89	10.62	14.66	61.89	11.75	10.61	12.79	7.68	16.27	15.26	18.54	11.40	14.89	12.32
RI-PC2	3.47	3.48	1.95	1.68	0.18	3.22	0.52	1.52	1.45	1.11	0.77	1.83	3.88	2.21	2.03

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table E.3: Factor loadings and statistics for principal component analysis (PCA)

This table shows the factor loadings and summary statistics of the principal component analysis (PCA) on the sentiment change over the long-term mean since inception. We apply the principal component analysis on all available equity indices from our developed and emerging markets universe to construct global sentiment factors. The PCA uses the fully available dataset from 1998 to 2017 and is thus subject to some degree of hindsight bias which we perceive as negligible. The first two principal components (PC1 & PC2) are used as global sentiment factors in the regression models as an alternative specification for our portfolio-based global sentiment factors Panel A shows the factor loadings and Panel B the summary statistics.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
Panel A: Factor Loadings															
AU500	0.29	-0.03	0.07	-0.11	0.23	-0.02	-0.21	0.45	-0.22	0.55	-0.43	-0.15	-0.15	0.10	0.09
BR50	-0.02	-0.68	-0.09	0.18	0.20	0.12	0.30	0.28	-0.21	0.06	0.19	0.40	0.10	-0.12	-0.08
CA250	0.31	0.28	0.23	-0.04	-0.04	0.12	0.10	0.05	-0.20	0.19	0.41	0.08	0.59	0.39	0.04
CH20	0.28	-0.16	-0.14	0.14	0.05	-0.27	0.08	-0.16	0.46	0.44	0.48	-0.19	-0.27	0.05	-0.02
CN300	0.32	0.01	-0.29	-0.36	-0.11	0.24	0.26	0.03	-0.03	-0.14	0.07	-0.07	-0.08	-0.24	0.67
DE30	0.19	-0.35	0.25	0.38	-0.45	-0.04	0.27	-0.05	-0.05	-0.14	-0.28	-0.45	0.08	0.19	0.09
ES35	0.16	0.16	-0.19	0.51	0.29	0.70	-0.05	-0.16	0.14	-0.01	-0.12	-0.01	-0.08	0.09	-0.02
FR40	0.30	0.05	0.25	-0.04	-0.07	-0.13	0.21	-0.34	0.16	0.07	-0.35	0.67	-0.18	0.15	0.07
GB100	0.23	0.07	0.03	0.02	0.66	-0.37	0.26	0.08	0.17	-0.41	-0.14	-0.18	0.22	0.01	0.01
HK50	0.28	-0.12	-0.43	-0.44	-0.16	0.15	0.12	-0.01	0.06	-0.11	-0.15	-0.07	0.03	0.25	-0.60
IN50	0.23	-0.11	-0.34	0.15	0.08	-0.28	-0.31	-0.53	-0.54	0.03	-0.03	-0.01	0.14	-0.11	0.04
JP225	0.31	0.05	0.20	0.06	-0.01	-0.02	-0.20	0.21	-0.30	-0.44	0.33	0.06	-0.57	0.19	-0.09
RU50	-0.04	-0.41	0.47	-0.42	0.30	0.30	-0.20	-0.41	0.04	-0.01	0.06	-0.20	-0.01	0.06	0.04
SG30	0.26	-0.24	-0.04	0.06	-0.16	-0.03	-0.63	0.22	0.43	-0.22	-0.03	0.19	0.32	-0.05	0.13
US500	0.36	0.14	0.31	0.02	-0.10	0.10	0.06	0.00	-0.02	0.07	0.02	-0.06	0.07	-0.76	-0.37
Panel B: Summary Statistics															
Standard deviation	2.1877	1.2076	1.0356	1.0106	0.9585	0.8962	0.8923	0.8552	0.7952	0.7805	0.7502	0.7295	0.6860	0.5906	0.5069
Proportion of Variance	0.3191	0.0972	0.0715	0.0681	0.0612	0.0535	0.0531	0.0488	0.0422	0.0406	0.0375	0.0355	0.0314	0.0232	0.0171
Cumulative Proportion	0.3191	0.4163	0.4878	0.5559	0.6171	0.6706	0.7237	0.7725	0.8146	0.8552	0.8928	0.9283	0.9596	0.9829	1.0000

Table E.4: A Fama-French five-factor model with sentiment interactions

This table compares the Fama-French five-factor model (FF5) with a FF5 model including sentiment interactions. We code a sentiment dummy variable, which takes the value one if sentiment change is negative and interact this proxy with each of the classical five Fama-French factors. Panel A refreshes the analysis for the plain vanilla FF5 model for added comparability. Panel B reports the estimates of the FF5 model including sentiment interactions. The table shows the coefficient estimates, the Newey-West corrected standard errors to adjust for autocorrelation and heteroscedasticity in returns in parentheses, as well as the adjusted R^2. The regressions are estimated for each equity index individually on weekly returns from 1998 to 2017.

	AU500	BR50	CA250	CH20	CN300	DE30	ES35	FR40	GB100	HK50	IN50	JP225	RU50	SG30	US500
Panel A: FF5															
Intercept	-0.0002	0.0012	0.0006	-0.0002	0.0025	0.0004	0.0003	0.0003	-0.0001	0.0006	0.0014	0.0001	0.0006	-0.0003	0.0004
	(0.0004)	(0.0009)	(0.0004)	(0.0006)	(0.0018)	(0.0006)	(0.0006)	(0.0005)	(0.0003)	(0.0008)	(0.0009)	(0.0008)	(0.0017)	(0.0007)	(0.0004)
MRP	0.3937***	0.7326***	0.4978***	0.5483***	0.1640	0.6419***	0.5764***	0.6355***	0.5602***	0.5023***	0.4948***	0.5002***	0.8221***	0.4264***	0.6165***
	(0.0274)	(0.0607)	(0.0352)	(0.0418)	(0.0833)	(0.0443)	(0.0411)	(0.0373)	(0.0261)	(0.0489)	(0.0473)	(0.0438)	(0.0822)	(0.0445)	(0.0327)
SMB	0.5537***	0.6409***	0.5780***	0.1308	0.3399	0.3622***	0.3714***	0.3098***	0.2447***	0.6950***	0.7773***	0.7393***	1.1904***	0.6294***	0.2193***
	(0.0684)	(0.1414)	(0.0626)	(0.0744)	(0.1722)	(0.0870)	(0.0895)	(0.0753)	(0.0646)	(0.1154)	(0.1127)	(0.1299)	(0.2094)	(0.1085)	(0.0708)
HML	0.1141	0.2755	0.0983	0.0812	0.0411	0.1278	0.2830***	0.1559	0.0552	0.2673**	-0.0437	0.1774	0.5389***	0.1979**	-0.0604
	(0.0550)	(0.1524)	(0.0707)	(0.0938)	(0.1996)	(0.1078)	(0.1142)	(0.0980)	(0.0742)	(0.1097)	(0.0939)	(0.1193)	(0.2416)	(0.0933)	(0.0673)
RMW	0.2033**	0.1756	-0.0199	0.0123	-0.2049	-0.2548**	-0.3205***	-0.0729	0.1733*	-0.1535	0.0382	-0.1733	0.0216	0.1080	-0.0412
	(0.0670)	(0.2319)	(0.0909)	(0.0831)	(0.3106)	(0.1061)	(0.1171)	(0.0871)	(0.0656)	(0.1286)	(0.1083)	(0.1135)	(0.2239)	(0.1061)	(0.0676)
CMA	-0.1585	-0.4575**	-0.4611***	0.1004	-0.2613	-0.2914**	-0.2955**	-0.2210*	-0.0640	-0.5549***	-0.1388	-0.2588	-0.9420***	-0.4178***	-0.0084
	(0.0800)	(0.2589)	(0.0965)	(0.1695)	(0.3243)	(0.1629)	(0.1782)	(0.1556)	(0.1129)	(0.1588)	(0.1265)	(0.1528)	(0.3295)	(0.1194)	(0.0998)
Adj. R^2	30.06	36.83	47.13	32.69	2.34	39.34	34.28	41.33	41.27	26.38	20.13	24.44	21.23	27.07	54.37
Panel B: FF5 with Sentiment Interaction															
Intercept	-0.0005	-0.0015	0.0007	-0.0015	0.0038	0.0003	-0.0004	-0.0002	-0.0008	0.0004	0.0000	0.0006	0.0002	-0.0005	0.0004
	(0.0006)	(0.0013)	(0.0006)	(0.0009)	(0.0026)	(0.0010)	(0.0010)	(0.0009)	(0.0007)	(0.0012)	(0.0013)	(0.0011)	(0.0024)	(0.0010)	(0.0006)
MRP	0.4017***	0.7394***	0.5363***	0.5533***	0.0355	0.5977***	0.5683***	0.6304***	0.5468***	0.4566***	0.5192***	0.4821***	0.7932***	0.3767***	0.5909***
	(0.0426)	(0.0894)	(0.0402)	(0.0526)	(0.1501)	(0.0635)	(0.0569)	(0.0570)	(0.0413)	(0.0805)	(0.0748)	(0.0717)	(0.1210)	(0.0629)	(0.0375)
snt.dummy	0.0007	0.0052**	-0.0001	0.0027	-0.0026	0.0004	0.0016	0.0009	0.0014	0.0004	0.0029	-0.0008	0.0012	0.0005	0.0001
	(0.0010)	(0.0017)	(0.0010)	(0.0013)	(0.0028)	(0.0013)	(0.0013)	(0.0012)	(0.0010)	(0.0015)	(0.0016)	(0.0016)	(0.0029)	(0.0012)	(0.0009)
SMB	0.5465***	0.7170***	0.5385***	0.1256	0.1195	0.2726*	0.4715***	0.2895**	0.0950	0.6604***	0.7666***	0.6296***	0.9972***	0.7168***	0.1321
	(0.0912)	(0.2163)	(0.0924)	(0.1157)	(0.2449)	(0.1355)	(0.1285)	(0.1161)	(0.0906)	(0.1890)	(0.1789)	(0.1822)	(0.3011)	(0.1657)	(0.0988)
HML	0.0159	0.2075	0.0505	0.0367	0.1961	0.0633	0.1075	0.0121	0.0688	0.1977	-0.0803	0.1685	0.4636	0.1165	-0.0479
	(0.0900)	(0.2102)	(0.1018)	(0.1286)	(0.2710)	(0.1218)	(0.1409)	(0.1221)	(0.0983)	(0.1617)	(0.1402)	(0.1365)	(0.3898)	(0.1439)	(0.0698)
RMW	0.2614*	-0.0030	0.0732	0.0947	-0.6154	-0.2985	-0.1396	0.0047	0.1900	-0.2896	0.0050	-0.1945	0.0853	0.0359	-0.0459
	(0.0918)	(0.2959)	(0.1012)	(0.1422)	(0.4078)	(0.1599)	(0.1393)	(0.1382)	(0.1062)	(0.2092)	(0.1741)	(0.1792)	(0.3484)	(0.1666)	(0.0977)
CMA	-0.0977	-0.4014	-0.4912***	-0.0472	-0.4673	-0.4510**	-0.4394**	-0.3223	-0.1377	-0.5627**	-0.1729	-0.2674	-0.8179	-0.5286***	-0.0631
	(0.1014)	(0.3575)	(0.1287)	(0.2260)	(0.4226)	(0.1688)	(0.1764)	(0.1562)	(0.1221)	(0.2001)	(0.1789)	(0.1864)	(0.4914)	(0.1702)	(0.0877)
$MRP \times SNT$	-0.0166	0.0024	-0.0726	0.0012	0.2417	0.0869	0.0285	0.0151	0.0232	0.0863	-0.0387	0.0278	0.0408	0.1040	0.0461
	(0.0585)	(0.1153)	(0.0695)	(0.0642)	(0.1868)	(0.0827)	(0.0727)	(0.0750)	(0.0631)	(0.0914)	(0.0881)	(0.1005)	(0.1694)	(0.0696)	(0.0556)
$SMB \times SNT$	0.0185	-0.1533	0.0982	0.0463	0.3566	0.1815	-0.1458	0.0664	0.2958**	0.0487	0.0366	0.2022	0.3549	-0.1673	0.1664
	(0.1167)	(0.2615)	(0.1255)	(0.1395)	(0.3625)	(0.1769)	(0.1538)	(0.1552)	(0.1206)	(0.2056)	(0.2036)	(0.2226)	(0.3658)	(0.1759)	(0.1117)
$HML \times SNT$	0.1944	0.1653	0.0764	0.0607	-0.2775	0.0887	0.3181	0.2502	-0.0424	0.1236	0.0529	0.0095	0.1631	0.1268	-0.0325
	(0.1488)	(0.2874)	(0.1394)	(0.1599)	(0.3626)	(0.1908)	(0.1943)	(0.1749)	(0.1534)	(0.1951)	(0.2025)	(0.2127)	(0.4760)	(0.1878)	(0.1156)
$RMW \times SNT$	-0.1306	0.4241	-0.1874	-0.1467	0.8202	0.0814	-0.3237	-0.1498	-0.0690	0.2612	0.0770	0.0101	-0.2124	0.1825	-0.0127
	(0.1347)	(0.3663)	(0.1666)	(0.1821)	(0.5851)	(0.2074)	(0.1879)	(0.1904)	(0.1450)	(0.2422)	(0.2536)	(0.2432)	(0.4605)	(0.2072)	(0.1220)
$CMA \times SNT$	-0.1370	-0.0263	0.0748	0.3142	0.4341	0.3538	0.2690	0.2129	0.1765	0.0267	0.0943	0.0319	-0.2651	0.2307	0.1230
	(0.1991)	(0.4496)	(0.1824)	(0.2495)	(0.5797)	(0.2628)	(0.2598)	(0.2410)	(0.2149)	(0.2610)	(0.2835)	(0.2708)	(0.6094)	(0.2595)	(0.1686)
Adj. R^2	29.99	37.43	47.32	33.38	2.31	39.55	35.15	41.82	41.74	26.15	20.14	24.14	20.99	27.45	54.30

* p $\leq$ 0.1, ** p $\leq$ 0.05, *** p $\leq$ 0.01

Table E.5: Comparing the CAPM to different sentiment-augmented CAPM for additional indices

This table compares the standard CAPM to a number of sentiment-augmented linear factor models on international indices that are not included in the construction of the sentiment factors, i.e. the six indices from our control group are excluded during the portfolio sorts forming either the PMNSNT factor or the separate (positive) negative sentiment factor $(s(+))$ $s(-)$. Panel A concerns results for the standard CAPM based on the market risk premium (MRP) as its single priced factor. Panel B is the extension of the CAPM to include the return of a long-positive/short-negative sentiment portfolio as a single sentiment risk factor (PMNSNT-CAPM). Panel C uses two separate sentiment risk factors based on the returns of portfolios mimicking negative and positive sentiment (SNT-CAPM). The table reports the coefficient estimates,the Newey-West corrected standard errors to adjust for autocorrelation and heteroscedasticity in returns in parentheses and the adjusted R^2. In the case of models including sentiment, we also report the relative importance indicator for each coefficient, which estimates its relative contribution to the total explained variation. The regressions are estimated for each equity index individually on weekly returns from 1998 to 2017.

	EM50	EU50	GBMID250	US30	USMID2000	USNAS100
Panel A: CAPM						
Intercept	0.0012*	-0.0004	0.0012**	0.0007*	0.0006	0.0005
	(0.0008)	(0.0005)	(0.0005)	(0.0004)	(0.0006)	(0.0007)
MRP	0.6029***	0.6353***	0.4685***	0.5385***	0.6920***	0.7639***
	(0.0585)	(0.0337)	(0.0303)	(0.0203)	(0.0288)	(0.0375)
Adj. R^2	28.55	38.89	30.06	46.06	41.97	37.80
Panel B: PMNSNT-CAPM						
Intercept	0.0012	-0.0004	0.0012**	0.0007	0.0006	0.0005
	(0.0008)	(0.0005)	(0.0005)	(0.0004)	(0.0005)	(0.0008)
MRP	0.6029***	0.6353***	0.4685***	0.5385***	0.6920***	0.7639***
	(0.0590)	(0.0319)	(0.0281)	(0.0197)	(0.0288)	(0.0437)
PMNSNT	0.0024	-0.1213**	-0.0176	-0.0358	-0.0005	0.0274
	(0.0571)	(0.0705)	(0.0306)	(0.0329)	(0.0384)	(0.0638)
Adj. R^2	28.47	39.29	30.02	46.11	41.92	37.76
RI-MRP	100.00	98.80	99.93	99.78	100.00	99.95
RI-PMNSNT	0.00	1.20	0.07	0.22	0.00	0.05
Panel C: SNT-CAPM						
Intercept	0.0012***	-0.0004	0.0012***	0.0007**	0.0006	0.0005
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0005)	(0.0007)
MRP	0.6029***	0.6353***	0.4685***	0.5385***	0.6920***	0.7639***
	(0.0335)	(0.0378)	(0.0213)	(0.0198)	(0.0228)	(0.0415)
$s(+)$	1.0356***	0.7656***	0.6081***	0.4324***	0.6062***	0.6732***
	(0.0305)	(0.0334)	(0.0365)	(0.0324)	(0.0458)	(0.0572)
$s(-)$	0.5018***	0.5281***	0.3032***	0.2427***	0.2824***	0.3195***
	(0.0510)	(0.0588)	(0.0418)	(0.0346)	(0.0464)	(0.0587)
Adj. R^2	80.12	74.81	63.31	66.33	62.64	55.42
RI-MRP	35.69	52.01	47.52	69.42	66.98	68.15
RI-$s(+)$	55.63	37.55	44.16	24.68	28.35	27.86
RI-$s(-)$	8.68	10.44	8.33	5.90	4.67	3.98

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table E.6: Sub-samples comparisons

This table reports the results for a number of asset pricing models on different subsets of the data for various financial cycles. The columns show a CAPM, a long-positive/short-negative sentiment CAPM (PMNSNT-CAPM), a 2-factor sentiment CAPM (SNT-CAPM), a 5-factor Fama-French model (FF5), a long-positive/short-negative sentiment FF5 model (PMNSNT-FF5), and a 2-factor sentiment FF5 model (SNT-FF5). Panel A shows results concerning the full sample from 1998-2017 as comparison, Panel B the dot-com crisis 1998-2001, Panel C the pre-crisis period 2002-2006, Panel D the financial crisis period 2007-2011, Panel E the post-crisis period 2012-2017. We provide the average, minimum, and maximum absolute alpha and adjusted R^2 as well as the relative importance of sentiment as the sum of of positive and negative sentiment.

	Abs. Alpha			Adj. R^2			RI-SNT		
	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min
Panel A: Full Sample 1998-2017									
CAPM	0.23	0.06	0.01	53.25	27.03	2.03			
PMNSNT-CAPM	0.23	0.06	0.01	53.25	27.39	2.99	33.70	3.51	0.00
SNT-CAPM	0.23	0.06	0.01	73.90	57.14	13.26	84.06	55.68	27.94
FF5	0.25	0.06	0.01	54.37	31.93	2.34			
PMNSNT-FF5	0.23	0.06	0.01	54.39	32.37	3.23	25.02	3.02	0.02
SNT-FF5	0.23	0.06	0.01	74.64	57.92	12.86	77.60	47.06	27.03
Panel B: Sub-sample 1998-2001									
CAPM	0.24	0.08	0.02	50.35	25.96	5.99			
PMNSNT-CAPM	0.24	0.09	0.01	50.12	26.62	5.56	27.07	5.40	0.16
SNT-CAPM	0.23	0.09	0.00	65.38	50.17	26.46	80.04	52.46	24.06
FF5	0.31	0.09	0.01	52.49	35.45	14.67			
PMNSNT-FF5	0.22	0.10	0.02	52.30	35.84	15.04	17.87	3.44	0.11
SNT-FF5	0.22	0.10	0.02	66.23	51.61	29.05	72.78	39.50	22.73
Panel C: Sub-sample 2002-2006									
CAPM	0.70	0.20	0.01	55.73	24.26	0.71			
PMNSNT-CAPM	0.70	0.20	0.01	56.68	25.44	−0.34	22.36	7.81	0.53
SNT-CAPM	0.55	0.16	0.00	73.47	54.08	3.99	76.41	54.23	20.98
FF5	0.60	0.13	0.00	56.40	29.64	1.14			
PMNSNT-FF5	0.68	0.15	0.01	57.74	31.31	0.19	16.52	6.76	0.17
SNT-FF5	0.59	0.14	0.00	75.72	56.00	3.07	59.16	44.07	25.70
Panel D: Sub-sample 2007-2011									
CAPM	0.14	0.07	0.00	54.35	32.26	1.51			
PMNSNT-CAPM	0.13	0.07	0.00	54.17	32.65	2.95	52.23	6.71	0.46
SNT-CAPM	0.21	0.08	0.02	82.49	69.08	14.42	88.37	56.33	34.67
FF5	0.17	0.08	0.01	56.14	38.42	0.82			
PMNSNT-FF5	0.16	0.06	0.02	55.97	38.81	2.11	40.19	4.68	0.34
SNT-FF5	0.19	0.07	0.02	84.47	70.57	14.02	85.29	46.16	28.92
Panel E: Sub-sample 2012-2017									
CAPM	0.26	0.11	0.00	52.04	23.31	2.56			
PMNSNT-CAPM	0.27	0.11	0.00	51.92	25.05	3.10	44.72	9.93	0.03
SNT-CAPM	0.29	0.12	0.02	73.95	54.39	11.22	74.30	59.01	29.40
FF5	0.25	0.11	0.01	52.51	25.88	4.67			
PMNSNT-FF5	0.30	0.11	0.03	52.41	27.77	5.07	36.77	8.32	0.01
SNT-FF5	0.28	0.11	0.02	74.84	55.20	11.54	68.47	50.40	26.70

Table E.7: Comparison between news- and social media-based sentiment indicators

This table depicts the results for various asset pricing models based on different sentiment measures and in particular when we differentiate between news-only and social media-only driven sentiment compared to a combined (news and social media) measure of sentiment. The columns show the CAPM, a long-positive/short-negative sentiment CAPM (PMNSNT-CAPM), a 3-factor sentiment CAPM (SNT-CAPM), a 5-factor Fama-French model (FF5), a long-positive/short-negative sentiment FF5 model (PMNSNT-FF5), and a 3-factor sentiment FF5 model (SNT-FF5). We report the average, minimum, maximum, and median absolute alpha and adjusted R^2 as well as the relative importance of sentiment as the sum of of negative, neutral, and positive sentiment. Panel A shows the combined news and social media models, Panel B the news only, and Panel C the social media only. The sample period is January 1998 - December 2017.

	Abs. Alpha			Adj. R^2			RI-SNT		
	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min
Panel A: Combined News and Social Media									
CAPM	0.23	0.06	0.01	53.25	27.03	2.03			
PMNSNT-CAPM	0.23	0.06	0.01	53.25	27.39	2.99	33.70	3.51	0.00
SNT-CAPM	0.23	0.06	0.01	73.90	57.14	13.26	84.06	55.68	27.94
FF5	0.25	0.06	0.01	54.37	31.93	2.34			
PMNSNT-FF5	0.23	0.06	0.01	54.39	32.37	3.23	25.02	3.02	0.02
SNT-FF5	0.23	0.06	0.01	74.64	57.92	12.86	77.60	47.06	27.03
Panel B: News only									
CAPM	0.23	0.06	0.01	53.25	27.03	2.03			
PMNSNT-CAPM	0.23	0.06	0.01	53.21	27.51	2.65	26.21	3.71	0.00
SNT-CAPM	0.23	0.06	0.01	74.49	57.59	15.11	85.95	56.03	28.52
FF5	0.25	0.06	0.01	54.37	31.93	2.34			
PMNSNT-FF5	0.23	0.06	0.01	54.34	32.53	2.89	18.30	3.21	0.00
SNT-FF5	0.23	0.06	0.01	75.31	58.39	14.78	80.33	47.49	27.67
Panel C: Social Media only									
CAPM	0.23	0.06	0.01	53.25	27.03	2.03			
PMNSNT-CAPM	0.23	0.06	0.01	53.43	27.91	2.73	28.00	5.65	0.13
SNT-CAPM	0.23	0.06	0.01	74.44	57.32	9.71	78.50	55.60	28.47
FF5	0.25	0.06	0.01	54.37	31.93	2.34			
PMNSNT-FF5	0.23	0.06	0.01	54.57	32.77	3.00	20.70	4.10	0.06
SNT-FF5	0.23	0.06	0.01	75.11	58.27	9.32	69.99	46.94	27.49

Table E.8: Cross-sectional pricing comparisons of the CAPM and the SNT-CAPM without orthogonalization

This table shows the results of a comparison between eight different models where the variables are not orthogonalized: i) a standard CAPM, used as a naïve benchmark, ii) a standard Fama-French five-factor model FF5, iii) a PMNSNT-CAPM model that extends the CAPM model to include a long-positive/short-negative sentiment factor, iv) a PMNSNT-FF5 which adds the PMNSNT factor to a Fama-French five-factor specification, v) a PMNSNT-FF5x model that additionally controls for $iv_{i,t}$, $is_{i,t}$, $ik_{i,t}$, and the log returns on the U.S. Dollar index, vi) a SNT-CAPM that prices positive and negative sentiment separately, vii) a SNT-FF5 Fama-French five-factor specification with two sentiment factors, and vii) a SNT-FF5x model that also incorporates the same control variables listed above. We report the market risk premium as λ and its standard error. For the CAPM we assume market risk to be priced correctly and hence equal to the average weekly excess return on the MSCI World portfolio, so that no standard error of the estimate is provided. λ_{PMNSNT} is the price of the single sentiment risk factor PMNSNT. The three estimates $\lambda_{s(+)}$ and $\lambda_{s(-)}$ are the sentiment premia for positive and negative sentiment change, respectively. Additional rows indicate whether we also control jointly for the Fama-French factors for size, value, profitability and investment, the change in idiosyncratic volatility $iv_{i,t}$ (dVol), skewness $is_{i,t}$ (dSkew), kurtosis $ik_{i,t}$ (dKurt), and the log returns on the U.S. Dollar index computed from a basket of international currencies. Standard errors of the estimates are provided in parentheses and incorporate the Shanken (1992) errors-in-variables correction as well as the Newey-West correction for heteroscedasticity. P-values are corrected by the Bonferroni-Holm method to account for multiple testing bias. The last row reports the adjusted R^2 of each model for the cross-section of returns.

	Baseline		Aggregated Sentiment Factor PMNSNT			Disaggregated Sentiment Factors SNT		
	CAPM	FF5	CAPM	FF5	FF5x	CAPM	FF5	FF5x
λ	0.2178***	0.2174***	0.2324***	0.2199***	0.1976***	0.0991***	0.1076***	0.3410***
	(0.0029)	(0.0031)	(0.0034)	(0.0029)	(0.0023)	(0.0005)	(0.0006)	(0.0048)
λ_{PMNSNT}			0.0769***	0.0979***	-0.0120***			
			(0.0011)	(0.0013)	(0.0001)			
$\lambda_{s(+)}$						0.1708***	0.1947***	0.2429***
						(0.0012)	(0.0016)	(0.0033)
$\lambda_{s(-)}$						0.0817***	0.0572***	0.2047***
						(0.0004)	(0.0004)	(0.0022)
FF5 Factors				Yes	Yes		Yes	Yes
dVol					Yes			Yes
dSkew					Yes			Yes
dKurt					Yes			Yes
USD Index					Yes			Yes
Adj. R2	35.34	35.11	42.55	42.81	46.15	48.14	48.07	48.82

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$

F Sub-sample analysis using NBER classification

In addition to our naïve classification of bull and bear markets and in consideration of the reference point role of the U.S. financial markets, we also apply the categorization of recessions offered by the National Bureau of Economic Research (NBER) and show the results in Table F.1 of the Internet Appendix. NBER defines a recession as a significant decline in economic activity spread across the economy, lasting for at least two consecutive quarters, normally observable in real GDP, real income, employment, industrial production, and wholesale-retail sales. According to this methodology, our dataset comprises three expansionary cycles, shown in Panels A (Jan 1998 - Feb 2001), C (2001 Dec - Nov 2007), and E (Jul 2009 - Dec 2017) as well as two recession phases, for which results appear in Panels B (Mar 2001 - Nov 2001) and D (Dec 2007 - Jun 2009). First, improvements in the model performance measures are achieved by splitting sentiment into its positive and negative tone, independent of market cycle classification. This means disaggregated sentiment improves standard asset pricing models in both bull and bear markets. Second, if we compare the two recessions, we observe that while the overall goodness-of-fit measures are similar, the relative importance of sentiment considerably increases during the financial crisis of 2007-2009. This is consistent with the conjecture of improved sentiment measures caused by the introduction of social media in 2006 that have increased news coverage and have led to an increased effectiveness by the language processing engine applied by MarketPsych. If we apply the same comparison to the three expansionary periods, the relative importance of sentiment diminishes in the recent bull cycle. These findings suggest that the pricing effects of sentiment may change over time. Third, if we compare the more recent recessionary phases paired with earlier ones, we observe that the effect of sentiment has changed. These remarks beg for further research on the time-varying pricing effects of sentiment. In unreported results (available upon request), we observe that the pricing of positive and negative sentiment is distorted during recessions, indicating that the pricing mechanism of sentiment shifts in such a way that neither positive nor negative sentiment are compensated while there remains a reward to investing in a long-

positive/short-negative sentiment portfolio. Moreover, in bull regimes, the findings from the full-sample analysis are even further pronounced with negative sentiment being even more negatively, and positive sentiment even more positively priced.

Table F.1: Sub-samples Comparisons - NBER Recession Dates

This table reports the results for a number of asset pricing models on different subsets of the data following the classification of recessions by the National Bureau of Economic Research (NBER). The columns show a CAPM, a long-positive/short-negative sentiment CAPM (PMNSNT-CAPM), a 3-factor sentiment CAPM (SNT-CAPM), a 5-factor Fama-French model (FF5), a long-positive/short-negative sentiment FF5 model (PMNSNT-FF5), and 3-factor sentiment FF5 model (SNT-FF5). The dataset comprises three expansionary markets in Panel A (Jan 1998 - Feb 2001), Panel C (2001 Dec - Nov 2007) and Panel E (Jul 2009 - Dec 2017) as well as two recessions in Panel B (Mar 2001 - Nov 2001) and Panel D (Dec 2007 - Jun 2009). We provide the average, minimum, and maximum absolute alpha and adjusted R^2 as well as the relative importance of sentiment as the sum of negative, neutral, and positive sentiment.

	Abs. Alpha			Adj. R^2			RI-SNT		
	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min
Panel A: Jan 1998 - Feb 2001 Bull Market									
CAPM	0.35	0.12	0.00	55.34	27.51	2.92			
PMNSNT-CAPM	0.28	0.11	0.01	55.23	28.37	2.60	31.73	7.11	0.12
SNT-CAPM	0.30	0.11	0.02	64.46	48.22	23.57	89.22	51.40	18.87
FF5	0.39	0.14	0.02	56.10	35.48	9.77			
PMNSNT-FF5	0.29	0.12	0.01	56.11	36.10	9.43	24.22	4.41	0.19
SNT-FF5	0.30	0.12	0.01	66.58	50.27	24.89	72.98	37.42	18.12
Panel B: Mar 2001 - Nov 2001 Bear Market									
CAPM	0.89	0.30	0.04	36.07	21.00	1.15			
PMNSNT-CAPM	1.12	0.41	0.02	45.34	29.20	11.86	78.59	35.67	12.04
SNT-CAPM	1.38	0.27	0.03	81.95	64.84	27.73	84.17	59.23	24.19
FF5	1.17	0.31	0.02	60.77	43.31	21.50			
PMNSNT-FF5	1.40	0.31	0.03	65.74	48.14	28.10	36.19	17.81	3.02
SNT-FF5	1.43	0.28	0.01	87.23	67.71	26.48	62.22	45.80	14.83
Panel C: Dec 2001 - Nov 2007 Bull Market									
CAPM	1.09	0.23	0.00	55.84	25.02	0.22			
PMNSNT-CAPM	1.05	0.22	0.00	56.38	26.16	5.37	81.52	10.61	0.42
SNT-CAPM	0.84	0.18	0.01	74.66	56.54	14.40	94.69	57.27	23.37
FF5	1.13	0.20	0.00	56.53	30.78	−1.33			
PMNSNT-FF5	1.12	0.21	0.00	57.35	32.27	4.42	68.74	9.08	0.15
SNT-FF5	0.88	0.17	0.00	75.96	58.35	14.01	81.15	48.80	27.74
Panel D: Dec 2007 - Jun 2009 Bear Market									
CAPM	0.45	0.23	0.00	53.15	33.76	−0.88			
PMNSNT-CAPM	0.45	0.23	0.04	52.61	33.96	−0.88	81.80	11.35	1.17
SNT-CAPM	0.49	0.14	0.02	82.92	71.77	8.83	97.69	57.65	35.55
FF5	0.45	0.21	0.00	59.06	44.15	−3.74			
PMNSNT-FF5	0.41	0.14	0.01	61.11	45.30	−4.03	27.52	5.67	0.93
SNT-FF5	0.42	0.12	0.00	83.57	73.22	7.38	83.06	42.46	25.66
Panel E: Jul 2009 - Dec 2017 Bull Market									
CAPM	0.23	0.08	0.02	52.54	25.96	4.33			
PMNSNT-CAPM	0.22	0.08	0.01	52.49	26.53	4.50	19.01	3.48	0.06
SNT-CAPM	0.25	0.09	0.01	80.34	59.49	14.25	73.62	56.19	32.62
FF5	0.22	0.08	0.01	52.85	27.74	5.66			
PMNSNT-FF5	0.26	0.08	0.02	52.84	28.46	5.72	16.32	3.29	0.04
SNT-FF5	0.26	0.08	0.02	81.13	60.07	14.39	68.38	49.78	30.45

Essay II: Does social media sentiment matter in the pricing of U.S. stocks?

Does social media sentiment matter in the pricing of U.S. stocks?*

Christian Koeppel†

Abstract

This paper applies a recently developed social media-based sentiment proxy for the construction of a new risk factor for sentiment-augmented asset pricing models on U.S. equities. Accounting for endogeneity, autocorrelation and heteroskedasticity in a GMM framework, we find that the inclusion of sentiment significantly improves the performance of the five-factor model from Fama and French (2015, 2017) for different industry and style portfolios such as size, value, profitability, investment. The sentiment risk premium provides the missing component in the behavioral asset pricing theory of Shefrin and Belotti (2008) and (partially) resolves the pricing puzzles of small extreme growth, small extreme investment stocks and small stocks that invest heavily despite low profitability.

JEL Classification Codes: $C32$, $C53$, $G12$, $G41$

Keywords: *Asset pricing*; *behavioral finance*; *financial markets*; *investor sentiment*; *sentiment risk premium*; *gmm*.

*We are grateful to Richard Peterson from MarketPsych and Elijah DePalma from Refinitiv for granting access to the MarketPsych sentiment indicators. Wharton Research Data Services (WRDS) was used in preparing this paper. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

†University St. Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland; Phone: +41 (0)76 38 30 757, Email: christian.koeppel@live.com

1 Introduction

Shefrin and Belotti (2008) stipulate that the price of any security is determined by the sum of a fundamental and sentiment premium. In their latest seminal work, Fama and French (2015) add profitability (RMW) and investment (CMA) to the market, size (SMB) and value (HML) factors[1] of the famous Fama-French three-factor asset pricing model (henceforth FF3) to describe average stock returns. The formation of this new five-factor model (henceforth FF5) is a response to the findings of Titman, Wei, and Xie (2004), Aharoni, Grundy, and Zeng (2013), and Novy-Marx (2013) and justified by the dividend discount model. Using the Modigliani and Miller (1958) theorem, they formulate the relationship of a share's expected return to its size, price-to-book ratio and expectations of future profitability and investment.

These possible explanations were constrained by the paradigms of neoclassical finance and traditional asset pricing models, which neglect the contribution of the sentiment premium as postulated by Shefrin and Belotti (2008). Those classical models assume that their predictions are the same, whether the price behavior is rational or irrational, and constant over time. Thus, the results show that even the FF5 model only partially describes an asset's expected return and, in particular, fails for presumably irrational situations. Asset pricing puzzles of hard-to-value stocks remain unsolved like explaining average returns of small stocks or companies which invest heavily despite low profitability. When supply and demand determine market prices, those firms' evaluations seem to deviate from their fundamentally explainable values and could be driven by irrational investors. Such behavior cannot be captured by any fundamental pricing model that assumes rational agents in perfect markets. In fact, investors do not necessarily act completely reasonably nor do they only pursue the maximization of their personal utility function. Likewise, they are not unexceptionally risk-averse. They do not have perfect data nor the unlimited cognitive capabilities to gather, absorb and interpret all information

[1]SMB: Mimicking portfolio of small-minus-big market capitalization ($size$) firms; HML: Mimicking portfolio of high-minus-low book-to-market ratio (B/M) firms; RMW: Mimicking portfolio of robust-minus-week profitability (OP) firms; CMA: Mimicking portfolio of conservative-minus-aggressive investment (Inv) firms

immediately and correctly to derive fully rational and optimal decisions. Neither are the markets strongly efficient nor free of arbitrage as widely assumed by traditional finance. In contrast, behavioral theories consider human biases and market imperfections in the analysis of risk and return, and the decision-making process for investments. In addition to the risk-free rate and a fundamental risk premium, the intrinsic value of an asset is further described by sentiment, forming a joint stochastic discount factor for future cash flows. Due to the lack of other sources, the sentiment risk premium has been often estimated as the dispersion of analysts' forecasts.

In this paper, we add the sentiment premium to asset pricing, in particular to address anomalies of hard-to-value portfolios of extreme growth, extreme investment, and small stocks that invest heavily despite weak profitability. We test Shefrin and Belotti's (2008) hypothesis by constructing a novel sentiment risk factor based on a set of recently developed, direct search-based investor sentiment indicators provided by MarketPsych. This new type of measure is derived from a proprietary human language processing that analyzes asset-specific information as it circulates through social media channels (see, e.g., Chen et al., 2014). This study shows that this indicator captures and quantifies investors' sentiment, making it the ideal factor to model the sentiment risk premium in the cross-section of U.S. equity markets.

Our key results can be summarized as follows: By using search-based and bottom-up sentiment indicators to construct a new risk factor, we reveal patterns in average returns related to investors' mood. The sentiment score augments the existing fundamental asset pricing model by Fama and French (2015) for the U.S. equity market and adds the sentiment premium to the pricing paradigm from Shefrin and Belotti (2008). We trace back the significant performance improvements for different factor-mimicking and industry portfolios to investors' activities in social media. The same effect is not found exploiting sentiment extracted from public news channels. Different notions of human mood are expressed in social media and can be used to form a monthly-rebalanced sentiment risk factor which is orthogonal to underlying macroeconomic or business cycle-related developments. It is also unrelated to existing fundamental pricing factors and independent

of momentum.

In two sets of portfolio sorts with either 25 or 36 portfolios using 5x5 double-sorting or 2x2x3x3 quadruple-sorting, we show a high average excess return spread between stocks with positive, and securities with negative sentiment. This spread is significantly higher than for other style tilts like size, value, profitability, or investment. We use those factor-mimicking portfolios in a sentiment-augmented asset pricing model to show that the inclusion of sentiment reduces the intercepts and improves the explained variation. Our model accounts for endogeneity, autocorrelation of returns and heteroskedasticity by mapping it to the GMM framework and generating robust standard errors. We show that similarly to the five-factor model from Fama and French (2015), the GRS test rejects our sentiment-augmented model to fully explain the cross-section of expected returns for different style or industry portfolios. However, it improves the five-factor model in the cross-section along various dimensions and provides an additional piece to solve the pricing puzzle for portfolios of hard-to-value securities like small extreme growth and extreme investment stocks. Sentiment reduces the (absolute) intercepts by 13-17%.

We also directly address the research challenge put forward by Fama and French (2015) regarding behavioral stories for stocks that invest heavily despite low profitability. Sentiment has an indirect contribution by modifying the slopes of other factors. We also report very poor performance of the five-factor model for sentiment-tilted portfolios. Unsurprisingly, the inclusion of the sentiment risk factor here is particularly beneficial as it is predestined to explain the excess returns of those style portfolios. Judged on regression slopes, both largecaps and smallcaps with negative sentiment seem to behave like stocks with weak profitability.

The paper is organized as follows. Section 2 introduces the stock and sentiment dataset with detailed descriptions and summary statistics. We begin our analysis in Section 3 with the formation of factor-mimicking portfolios in order to extract return patterns of the known existing fundamental stock properties and our novel sentiment risk factor. In Section 4, we construct both the fundamental and sentiment factors which are used in Section 5 for our sentiment-augmented GMM pricing model and the five-factor

benchmark model from Fama and French (2015). In Section 6, we explore the relationship between sentiment and fundamental factors before going into the detailed regression analysis on intercepts and slopes for well-known asset pricing puzzles of hard-to-value stocks. Section 8 juxtaposes our findings with the same results for sentiment extracted from news instead of social media. Section 9 concludes.

2 U.S. equity stocks and sentiment indicators

Following Fama and French (2015), our dataset covers all New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automated Quotations (NASDAQ) stocks on both CRSP and Compustat with share codes 10 or 11 for the period from January 1998 - December 2017.[2] The data traditionally excludes Financial Services, i.e. banks, insurance companies, broker/dealers, real estate and other financial services. We use monthly data as in the original paper of Fama and French (2015).

Our search-based sentiment index is the Refinitiv-MarketPsych social media-based sentiment indicator (RMI) covering more than 12.000 U.S. stocks. The automatic language processing system from MarketPsych uses a human-developed lexicon, which associates words and word groups to different kinds of indicators related to the performance of financial assets. Words and word groups in a message are annotated with so-called "Psych Words" (e.g., volatility, conflict, safety, etc.), defining a novel conceptual space. To define groups of words and create relationships, the lexical distance is assessed by applying weights on a scale from 0.0 to 1.0 to account for proximity in the text, while punctuation and additional structures are also taken into account. This process results in tuples, which are then recorded and aggregated as sentiment indicators. The scores are again divided by the total of the scores for all psych categories, called the $Buzz$, i.e., the weight of all messages and phrases of interest over a certain period. This ratio gives an indication of how important (or commonly discussed) a subject was over a given time interval in social

[2]The time period is very limited for an asset pricing study which is due to the unavailability of sentiment data before 1998.

media channels. This normalization allows equally weighted comparisons among numerous topics and nouns. Compared to other sentiment providers like RavenPack (see, e.g., Audrino et al., 2020; Y. Shi et al., 2016), MarketPsych indicators are not pre-calibrated to fit financial market prices and events using a training sample. Hence, we can use back-fitted time-series without any concern for the existence of hindsight biases. Because of this construction method, MarketPsych's approach goes far beyond the often used bag-of-words or similar techniques applied in previous studies (see, e.g., Jiang et al., 2019; Tetlock, 2007). The interested reader is referred to the Appendix A for further details about the construction of the RMI sentiment indicators.

In our empirical analysis, we use the aggregated sentiment indicator Snt from MarketPsych which captures the net positive versus negative references in the social media concerning a given stock. It can be interpreted as an overall sentiment proxy, void of any insight into the fundamental reasons for why references to a security may be positive or negative.[3] It comprises among others, the three most commonly documented emotions in the existing finance literature according to Shen et al. (2017): optimism, joy, and fear. Optimism is defined as the overestimation or overconfidence of investors on the future payoff of a financial asset, and may result in deviations of asset prices from intrinsic values as observed during extreme bullish or overheated markets. Overconfidence leads to the entry of retail investors, driving up liquidity (Odean, 1998). Joy and other pleasant emotions like bliss affect the subjective probability assessments of uncertain outcomes and thus, influence investors' decision-making as documented by Dolan (2002); Wright and Bower (1992).[4] Ciccone (2003) reports lower returns for firms with optimistic compared to those with pessimistic expectations. Fear, on the other hand, leads to demand shocks driven by investors' emotional stress, increasing market uncertainty and volatility. Da et al. (2015) establish a daily fear index based on the online searches of U.S. households, predicting return reversals and volatility.

Weighted by $Buzz$, we refer to abnormally high-positive or high-negative sentiment.

[3]The interested reader is referred to Appendix A for further details about the construction of the RMI sentiment indicators.

[4]In unreported results we also used the individual indicators for optimism, joy, and fear but the results did not justify the higher model complexity.

Since the level of sentiment shows trend behavior and is non-stationary, we compute the change to the long-term mean since inception ("sentiment changes"). Such an average is estimated and updated on a recursive basis to avoid any hindsight biases, i.e., the mean is computed on an expanding window since inception of the sentiment index until t, not the entire period T. As a result, the change in sentiment is comparable across stocks as we account for differences of sentiment perception.[5] This approach is advantageous over a monthly change since it allows us to consider time-varying perception of sentiment and thus better reflects the time-oscillation of sentiment as postulated by Shefrin and Belotti (2008). We treat the sentiment change for each individual company as a property of that particular stock and use it in Section 4 for the construction of Fama-French-like factor-mimicking portfolios and a sentiment risk factor PMN. PMN is the difference between the returns on diversified portfolios comprised of stocks with positive and negative deviation of previous month's sentiment from the expanding window mean from inception up to time t.

In order to resolve concerns that sentiment from social media may still only proxy underlying rational economic and business activities, we regress our constructed risk factor PMN against various economic indicators for the U.S. market. This analysis addresses the hypothesis whether the RMI sentiment indicator only captures fundamental, rational references to the business and economic cycles instead of true sentiment. Table 2 provides an overview of the seven most relevant economic and business factors which go beyond those used in Baker and Wurgler (2006) to orthogonalize their sentiment proxy. The list looks at labor statistics, production and consumption indices, price and inflation indicators, as well as a proxy for the developments in the U.S. housing market. The choice of indicators is motivated by those which most likely affect the mood and sentiment of retail investors, i.e. the same group who is most active in social media channels. The results in Table 3 show that the economic indicators only explain a negligible portion in

[5]This approach is backed by the literature and relates changes in sentiment to demand shocks. DeVault et al. (2019) identify whether trades explained by sentiment metrics are, in the aggregate, initiated by individual or institutional investors. Such a classification exploits the fact that changes in sentiment will be positively related to changes in sentiment traders' demand (i.e., demand shocks) in the case of speculative stocks and inversely related to demand shocks in the case of "safe stocks".

Table 2: List of economic indicators

This table shows the list of economic indicators. We provide the name and source as well as a short description of the data. This is typically taken directly from the source. For each indicator we compute the month-on-month percentage change from January 1998 - December 2017.

Name and Source	Description
Unemployment rate from the U.S. Bureau of Labor Statistics	Seasonally adjusted monthly change in the U.S. national unemployment rate. The unemployment rate represents the number unemployed as a percent of the labor force. Persons are classified as unemployed if they do not have a job, have actively looked for work in the prior 4 weeks, and are currently available for work.
Industrial Production from the Federal Reserve Bank of St. Louis	Seasonally adjusted monthly change in the U.S. industrial production index. The Industrial Production Index (INDPRO) is an economic indicator that measures real output for all facilities located in the United States manufacturing, mining, and electric, and gas utilities (excluding those in U.S. territories).
Consumption from the BEA - Bureau of Economic Analysis, U.S. Department of Commerce	Seasonally adjusted monthly change in the U.S. household consumption. Consumer spending, or personal consumption expenditures (PCE), is the value of the goods and services purchased by, or on the behalf of, U.S. residents.
Consumer price index from the U.S. Bureau of Labor Statistics	Monthly change in the U.S. consumer price index for all urban consumers. The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for a market basket of consumer goods and services.
Durable goods from the U.S. Census Bureau	Seasonally adjusted monthly change in the U.S. durable goods orders. The Manufacturers' Shipments, Inventories, and Orders (M3) survey provides broad-based, monthly statistical data on economic conditions in the domestic manufacturing sector. The survey measures current industrial activity and provides an indication of future business trends.
House price index from the Office of Federal Housing Enterprise Oversight	Seasonally adjusted monthly change in the U.S. FHFA house price index. The house price indexes measure changes in single-family home values based on data from all 50 states and over 400 American cities. The methodology based upon a weighted, repeat-sales statistical technique to analyze house price transaction data.
Non/farm payrolls from the U.S. Bureau of Labor Statistics	Seasonally adjusted monthly change in the U.S. non-farm payroll. Non-farm payrolls is the measure of the number of workers in the U.S. excluding farm workers surveyed from private and government entities throughout the U.S.

Table 3: Regressions of social media-based sentiment against economic indicators

This table shows the output of a regression of the social media-based sentiment factor PMN against well-known economic indicators. We show the coefficient estimates, standard errors, and t-statistics for each variable as well as the overall goodness of fit in terms of adjusted R^2. The details of the economic indicators can be found in Table 2.

	estimate	std.error	t-statistic	p-value
Unemployment_Rate	0.05	0.02	2.23	0.03
Industrial_Production	0.00	0.00	1.08	0.28
Personal_Consumption	-1.68	0.92	-1.83	0.07
CPI_MM	0.79	0.51	1.56	0.12
Durable_Goods	0.44	0.21	2.13	0.03
Home_Prices	0.04	0.25	0.16	0.87
Non.farm_Payrolls	-0.73	0.99	-0.74	0.46
Adj. R^2	0.03			

terms of adjusted R^2 of 0.03 of the PMN variance. The unemployment rate from the U.S. Bureau of Labor Statistics is significant at the 5% confidence level with a small coefficient of 0.05. The same applies for the order of durable goods from the U.S. Census Bureau with a higher estimate of 0.44. Personal consumption from the Bureau of Economic Analysis (BEA) is the last relevant factor at the 10% confidence level and a high negative estimation of -1.68. The remaining four economic indicators are not significant at a meaningful level. We conclude that the RMI factor truly captures sentiment and cannot be explained by economic or business activities, making any orthogonalization unnecessary.

Table 4 shows detailed summary statistics of the sentiment level across sectors, liquidity, size, and stock exchanges. For each category we display the mean, median, and standard deviation (SD). Panel A differentiates sentiment across various industries using the NAICS (North American Industry Classification) from the Compustat and CRSP databases. We see a heterogeneous picture for sentiment with values ranging from 3.09 for Arts, Entertainment, and Recreation to 8.50 for Professional, Scientific, and Technical Services. We also observe that the mean is skewed by outliers as the median tends to lie much lower with very high standard deviation. The median ranges from -1.46 for Other Services (except Public Administration) to 5.71 for Wholesale Trade. Panel B shows the sentiment across five liquidity quintiles. We use the ratio of trading volume over shares outstanding as a naïve proxy for liquidity. The breakpoints only use NYSE

securities[6] in order to prevent the bias of data with micro stocks. Sentiment is lowest for very liquid, and highest for illiquid stocks. Low-liquidity assets are also notably diverse with a standard deviation twice as large as for highly liquid securities. It seems that less liquid stocks are more sensitive to sentiment. Panel C differentiates firms by size using market equity from the CRSP database. Small firms tend to have lower sentiment than large companies but with a higher variance. Panel D differentiates sentiment across the three exchanges NYSE, AMEX, and NASDAQ. While NYSE and AMEX seem to behave very similarly, the mean, median, and standard deviation of sentiment is higher for the NASDAQ.

3 Factor-mimicking portfolios

Our empirical tests apply the sentiment-augmented asset pricing models to portfolios designed to produce large excess return spreads in size, value, profitability, investment, and sentiment. The sorting mechanism for the non-sentiment related portfolios closely follows the original methodology of Fama and French (2015).[7] For sentiment portfolios we apply the sorting more frequently to exploit the higher frequency. In contrast to fundamental data, our sentiment sorts are rebalanced on a monthly basis. Table 5 shows average monthly excess returns over the one-month U.S. treasury bill rate for 25 value-weighted (henceforth VW) portfolios from independent sorts of stocks into five size groups and an additional variable of interest. In Panels A-C, we first examine whether the well-known Fama-French patterns are also shown in our dataset. In Panel D, we apply the sorting with our social media-based sentiment variable which measures the deviation of sentiment in the previous month from the long-term mean since inception.

Panel A of Table 5 sorts the stocks into five $Size$ and five B/M groups, called 5x5

[6] As mentioned by Fama and French (1992) the use of NYSE stocks only prevents that breakpoints are biased by the large variability in characteristics of microcaps which comprise more than 60% of the investable universe but only 3% of total market value.

[7] The interested reader is referred to the third chapter in Fama and French (2015) for a detailed description of the applied methods.

Table 4: Descriptive statistics of social media-based sentiment by groups

This table shows simple summary statistics of social media-based sentiment for different categories. Panel A differentiates sentiment across various industries using the NAICS (North American Industry Classification) used in the Compustat and CRSP databases. Panel B shows the sentiment across five liquidity quantiles. We use the ratio of trading volume over shares outstanding as a naïve proxy for liquidity. The breakpoints only use NYSE in order not to bias the data with small stocks that only comprise a small portion of the traded volume. Panel C clusters sentiment by size using market capitalization with NYSE breakpoints from the CRSP database. Panel D differentiates sentiment across the three exchanges NYSE, AMEX, and NASDAQ. For each panel, we display the mean, median, and standard deviation of sentiment for stocks forming this group. The data covers the full time horizon January 1998 - December 2017 with rolling classifications and mappings to accomodate for changes over time. Values are scaled by a factor 100 for better visualization, now forming a range between -100 and 100.

Name	Mean	Median	SD
Panel A: Sentiment by Industries			
Agriculture, Forestry, Fishing and Hunting	4.05	0.52	26.96
Mining, Quarrying, and Oil and Gas Extraction	5.64	2.94	22.79
Utilities	4.84	2.20	29.49
Construction	7.59	3.53	28.12
Manufacturing	6.65	3.66	26.83
Wholesale Trade	8.00	5.71	32.50
Retail Trade	5.17	2.43	24.75
Transportation & Warehousing	4.70	1.21	28.59
Information	7.02	3.39	27.16
Finance and Insurance	6.66	3.33	34.39
Real Estate and Rental and Leasing	6.61	2.76	29.84
Professional, Scientific, and Technical Services	8.50	5.26	29.43
Admin. & Support & Waster Mgmt & Remediation Svcs	5.23	2.33	30.12
Educational Services	4.20	0.00	31.82
Health Care and Social Assistance	5.77	2.96	30.19
Arts, Entertainment, and Recreation	3.09	1.14	25.34
Accommodation and Food Services	6.55	4.35	23.81
Other Services (except Public Administration)	3.12	-1.46	33.45
Panel B: Sentiment by Liquidity			
Liquid	3.77	1.62	20.02
Q2	8.60	5.56	34.56
Q3	7.47	4.28	29.98
Q4	6.53	3.87	26.46
Illiquid	8.66	6.52	40.99
Panel C: Sentiment by Size			
Big	5.01	2.22	22.07
Q2	7.21	4.17	34.13
Q3	8.01	4.93	32.87
Q4	7.46	4.60	30.14
Small	3.46	1.12	30.38
Panel D: Sentiment by Exchange			
NYSE	5.65	2.42	26.94
AMEX	5.92	3.42	28.20
NASDAQ	7.35	4.40	29.88

$Size - B/M$ portfolios.[8] The data exhibits the size effect, for which the average return for each column of B/M falls from small to big stocks. As in the original paper of Fama and French (1993) the first column of low B/M stocks, also known as extreme growth stocks, is an exception and does not show a clear relation between $Size$ and B/M. Extreme growth stocks traditionally pose an asset pricing challenge which we address in detail later. The average monthly return of microcaps is smaller than for large stocks. In addition, the portfolios also show the value effect where average returns tend to increase from low to high B/M portfolios per each $Size$ row. However, this relation is not as clear for the two largest size quintiles.[9] Panel B shows the results for 25 VW portfolios based on stocks sorted into five $Size$ and five $Profitability$ quantiles.[10] Without any exception, we observe the $Size$ effect with decreasing returns. We also find the effect of Novy-Marx (2013) and Fama and French (2015) with increasing average returns from low to high profitability OP for each $Size$ row. The effect is primarily driven by the small (large) returns for low (high) profitability firms. In Panel C, we sort the stocks into five $Size$ and five $Investment$ portfolios.[11] The size effect is consistently confirmed. The investment factor itself shows decreasing returns if $Size$ is kept stable and investment activities are reduced from aggressively to conservatively investing firms. This suggests that average excess returns are lower for firms that invest more heavily. High Inv portfolios usually display significantly lower average returns. Firms with extremely high investment activities have comparatively small returns as investors might relate these to weak cash flows or

[8]The $Size$ and B/M quintile breakpoints use only stocks listed on the NYSE. However, the stock sample is all NYSE, AMEX, and NASDAQ stocks on both CRSP and Compustat with share codes 10 or 11 and data for $Size$ and B/M. As mentioned by Fama and French (1992) the use of NYSE stocks only prevents that breakpoints are biased by the large variability in characteristics of microcaps which comprise more than 60% of the investable universe but only 3% of total market value. We use the same methodology of NYSE breakpoints for sentiment as the approach ensures consistency and social media coverage also tends to be better for larger companies.

[9]Similar disturbances in the value effect are shown in the original monthly 5x5 portfolios from Kenneth's homepage for the reduced observation period January 1998 - December 2017.

[10]The operating profitability measure, OP, is computed every June of year t based on accounting data for the fiscal year ending in $t-1$ as annual revenues minus cost of goods sold, interest expense, and selling, general, and administrative expenses, divided by book equity at the end of fiscal year $t-1$. Breakpoints use only NYSE firms.

[11]Investment Inv is computed in June of year t as the growth of total assets for the fiscal year ending in $t-1$ divided by total assets at the end of $t-2$. We acknowledge that the Miller-Modigliani theorem refers to growth of equity instead of growth of assets, but we follow the rationale of Fama and French (2015) that the lagged growth of assets might be a better proxy for the infinite growth in book equity. It may also produce slightly larger spreads in average returns.

dividend outlooks. Similar to extreme growth stocks, extreme investment firms are one of the most difficult conundrums for asset pricing models. We will explore this phenomenon in detail later.

With Panel A-C, we are able to reproduce and confirm most of the results of Fama and French (2015) concerning size, value, profitability, and investment. In Panel D we show the real contribution with portfolio sorts on size and sentiment change in the previous month calculated as the deviation of sentiment from the long-term mean since inception. The 25 VW portfolios still confirm the size effect with decreasing returns when sentiment is kept roughly constant. The sentiment effect shows increasing returns from negative to positive sentiment in each $Size$ row. The positive relationship means that stocks for which the investors have a positive (negative) view, have a higher (lower) return. The spreads are much larger than in any of the three other panels. In concrete figures, small firms with positive sentiment outperform their peers with negative by 58.6bps per month (177.22 - 118.61). The effect becomes smaller with bigger firms and diminishes for large caps. Here, low and high sentiment firms have similar small excess returns. We argue that this is due to strong fundamentals while smaller firms seem to be more sensitive to sentiment change.

Given that this double-sorting does not jointly control for multiple fundamental variables, we are unable to disentangle the impact of value, profitability, or investment on each other and jointly on sentiment. Previous studies show that the Fama-French variables are indeed correlated (see, e.g., Fama & French, 1995, 2015). It may also be argued that our social media-based sentiment indicator purely reflects the information in prices and publicly available news. In order to isolate the various effects and to particularly resolve concerns that our sentiment indicator only measures fundamental information that is contained in the other variables, we choose another sorting approach. We compromise on the FF3 factors, $Size$ and B/M, which we split into two groups, and use each additional FF5 factor, profitability OP and investment Inv, to disentangle the sentiment effect. In order to manage the complexity, we use low, medium, and high quantiles for each variable. While we apply the same logic to sentiment (Negative $s(-)$, Neutral $s(0)$, and Positive

Table 5: 5x5 portfolios using social media-based sentiment

This table shows average monthly excess returns in basis points (bps) for portfolios formed on $Size$ and book-to-market ratio B/M, operating profitability OP, investment Inv, and sentiment Snt for January 1998 - December 2017. The sorting follows Fama and French (2015) using only NYSE breakpoints: at the end of each June, stocks are allocated to five Size groups (Small to Big) or five B/M groups (Low to High). The intersections produce 25 value-weight $Size - B/M$ portfolios. The $Size - OP$ and $Size - Inv$ portfolios are formed analogously with the second sort variable being either operating profitability or investment. For the sentiment-mimicking sorts $Size - Snt$, we sort the stocks more frequently based on the previous month's deviation of social media-based sentiment from the long-term mean since inception until time t.

Low / Negative	2	3	4	High / Positive
Panel A: Size-B/M Portfolios				
Small	41.30	84.62	86.76	99.36	97.74
2	68.13	85.71	87.60	77.67	81.32
3	68.69	78.87	77.99	72.23	100.23
4	82.18	75.32	85.20	74.84	76.81
Big	55.49	62.21	64.02	43.65	47.28
Panel B: Size-OP Portfolios				
Small	72.47	100.82	93.69	85.76	89.46
2	58.91	74.33	92.54	98.08	102.98
3	58.42	74.51	76.16	76.70	102.52
4	51.56	80.21	83.33	80.88	95.99
Big	21.58	44.21	49.36	56.75	58.80
Panel C: Size-Inv Portfolios				
Small	108.60	100.96	108.55	80.54	47.44
2	76.58	85.91	95.16	92.71	65.09
3	80.28	84.28	90.45	86.79	62.29
4	74.51	84.89	78.75	94.17	69.18
Big	63.35	52.43	59.90	59.74	49.92
Panel D: Size-SNT Portfolios				
Small	118.61	169.04	192.82	251.76	177.22
2	84.30	123.55	112.67	144.17	127.37
3	80.71	91.67	120.53	94.43	109.19
4	76.33	71.49	93.62	118.87	78.19
Big	66.78	32.77	48.14	64.62	61.66

$s(+)$), we form 2x2x3x3 = 36 portfolios.[12] Each sort is independent of the other for a cleaner distribution of stocks across the portfolios. However, this does not imply zero correlation.

Panel A of Table 6 shows the so called $Size - B/M - OP - Snt$ portfolios, sorting for size and value using the NYSE median and three clusters of each profitability and sentiment. For Panel B, we replace the third term, profitability OP, with investment Inv to form 36 $Size - B/M - Inv - Snt$ portfolios. First, size (small versus big) and value portfolios (low versus high B/M) monthly average excess returns increase from negative to positive sentiment. The spread is higher for smaller and growth companies in both panels. Small growth stocks which tend to be a challenge to price with traditional models show clear sentiment patterns. Second, there is a material difference in returns between small and large stocks with the latter having significantly smaller returns for the size effect. This confirms the previous findings that small stocks are more susceptible to sentiment. On the contrary, the value effect is again disturbed such that firms with high B/M ratio do not necessarily outperform their peers with low B/M. The profitability effect is thus mostly confirmed. Robust firms with high profitability outperform weak firms with low OP even if controlled for the other factors including sentiment. Exceptions occur for the group of small growth stocks where weakly profitable firms outperform their peers with robust OP, another asset pricing puzzle we explore later. In Panel B, we evaluate the relationship between investment and sentiment. The investment effect states that conservative firms have higher returns than aggressively investing companies. In terms of sentiment, the results confirm a clear pattern of primarily increasing returns with positive sentiment. Even if controlled for fundamental factors, portfolios formed on stocks with positive sentiment relate to higher returns than peers with negative sentiment.

[12] Due to our focus on sentiment, we only report sorts containing sentiment. Additional results for the other factors are available upon request.

Table 6: 2x2x3x3 portfolios using social media-based sentiment

This table shows average monthly excess returns in basis points (bps) for portfolios formed on $Size$ and book-to-market ratio B/M, operating profitability OP, investment Inv, and sentiment Snt for January 1998 - December 2017. The sorting follows Fama and French (2015) using only NYSE breakpoints: at the end of each June, stocks are allocated to two Size groups (Small and Big), two B/M groups (Low and High), three groups of either profitability or investment (Low, Medium, and High) and three groups of Snt (negative s(-), neutral s(0), and positive s(+)). For sentiment, we sort the stocks based on the previous month's deviation of social media-based sentiment from the long-term mean since inception until time t. The intersections produce 36 portfolios.

	Small						Large					
	Low B/M			High B/M			Low B/M			High B/M		
	$s(-)$	$s(0)$	$s(+)$	$s(-)$	$s(0)$	$s(+)$	$s(-)$	$s(0)$	$s(+)$	$s(-)$	$s(0)$	$s(+)$
Panel A: Size-B/M-OP-SNT Portfolios												
Low	128.12	110.97	140.50	79.19	143.59	131.55	42.32	85.18	16.04	22.28	49.45	79.17
Medium	100.45	104.48	121.21	78.13	143.32	99.98	66.23	46.36	86.54	81.40	66.33	77.34
High	104.55	128.69	156.78	90.68	114.08	146.29	55.78	54.70	69.29	69.16	97.89	104.84
Panel B: Size-B/M-Inv-SNT Portfolios												
Low	132.23	126.10	199.97	72.19	179.22	114.13	68.36	64.54	56.10	33.27	70.73	74.50
Medium	86.01	95.22	120.62	78.43	117.90	146.15	69.63	49.02	74.82	67.92	64.48	92.73
High	114.07	118.54	120.75	85.96	123.80	118.79	45.49	58.41	66.08	25.88	31.77	92.36

4 Risk factor construction and summary statistics

Fama and French (2015) examine the sensitivity of risk factors to the specifics of factor construction. The authors state that the original 2x3 sorts for the size and value factors have been an arbitrary choice. This sorting mechanism weights small and big stocks equally, i.e. they are roughly neutral with respect to size. However, if the model is extended by the additional variables profitability and investment, the constructed factors HML, RMW, and CMA do not control for the other aspects. For example, HML is neutral in terms of size but does not control for operating profitability nor investment. That means that the average HML return combines the premia of value, profitability and investment. Analogous remarks apply to the construction of RMW and CMA. In order to overcome this challenge, we jointly control for all factors and isolate the premia in average returns in five sorts related to size, value, profitability, investment, and sentiment. Even a minimalistic sorting of each variable into low and high using the median of all observations, results in a high number of portfolios (2x2x2x2x2 = 32). To be more precise, we form two groups of $Size$ for small and big stocks, two groups on low and high book-to-market ratio B/M, two groups on robust and week profitability OP, two groups on conservative and aggressive investments Inv, and finally two groups on positive $s(+)$ and negative $s(-)$ sentiment Snt. Overall, the intersections of these groups form 32 VW portfolios and the returns of the relevant factors are the difference of the averages of the two groups the factors account for. The average return of the SMB factor, for example, is the average of the 16 small portfolios minus the average of the 16 big portfolios. In terms of sentiment, PMN is the average of 16 positive sentiment portfolio minus 16 negative.[13] Further details can be found in Table 7. As the individual portfolios tend to be poorly diversified, we also choose the 2x3 sorts to achieve better diversification and allow for benchmarking with Fama and French (2015) at the expense that the average returns of factors may not entirely control for other variables.

[13]Sentiment here refers to the positive and negative deviation of previous month's sentiment from the long-term mean since inception until time t.

Table 7: Factor construction

This table shows the construction of size, value, profitability, investment, and sentiment factors. We use independent sorts to assign stocks to two size, value (B/M), operating profitability (OP), investment (Inv), and sentiment (SNT) groups. The value-weighted portfolios defined by the intersections of the groups are the building blocks for the factors, comprising in total 32 value-weighted portfolios. We label these portfolios with five letters. The first always describes the size, small (S) or big (B), the second the value, high (H) or low (L), the third the profitability, robust (R) or weak (W), the fourth the investment, conservative (C) or aggressive (A), and the fifth the sentiment group, positive (P) or negative (N). The building blocks then form the factors for size (SMB), value (HML), profitability (RMW), investment (CMA), and sentiment (PMN). Those are constructed as the difference between the 16 building blocks of one property minus the other 16. Exemplary for SMB these are all 16 portfolios with S for small stocks minus 16 portfolios with B for big stocks. The sorts with two groups use the NYSE median, the sorts with three groups the 30th and 70th NYSE percentiles as common in the literature citep[see, e.g.,]Fama1993, Fama2015, Fama2017.

Breakpoints		Construction
Panel A: 2x3 sorts on Size and B/M, OP, Inv, or PMN		
Size:	$SMB_{B/M}$ =	(SH + SN + SL)/3 - (BH + BN + BL)/3)
	SMB_{OP} =	(SR + SN + SW)/3 - (BR + BN + BW)/3)
	SMB_{Inv} =	(SC + SN + SA)/3 - (BC + BN + BA)/3)
	SMB =	$(SMB_{B/M} + SMB_{OP} + SMB_{Inv})/3)$
Value:	HML =	(SH + BH)/2 - (SL + BL)/2
Profitability:	RMW =	(SR + BR)/2 - (SW + BW)/2
Investment:	CMA =	(SC + BC)/2 - (SA + BA)/2
Sentiment:	PMN =	(SP + BP)/2 - (SN + BN)/2
Panel A: 2x2x2x2x2 sorts on Size, B/M, OP, Inv, and PMN		
Size:	SMB =	(SHRCP + SHRCN + SHRAP + SHRAN + SHWCP + SHWCN + SHWAP + SHWAN + SLRCP + SLRCN + SLRAP + SLRAN + SLWCP + SLWCN + SLWAP + SLWAN) / 16 -(BHRCP + BHRCN + BHRAP + BHRAN + BHWCP + BHWCN + BHWAP + BHWAN + BLRCP + BLRCN + BLRAP + BLRAN + BLWCP + BLWCN + BLWAP + BLWAN) / 16
Value:	HML =	(SHRCP + SHRCN + SHRAP + SHRAN + SHWCP + SHWCN + SHWAP + SHWAN + BHRCP + BHRCN + BHRAP + BHRAN + BHWCP + BHWCN + BHWAP + BHWAN) / 16 -(SLRCP + SLRCN + SLRAP + SLRAN + SLWCP + SLWCN + SLWAP + SLWAN + BLRCP + BLRCN + BLRAP + BLRAN + BLWCP + BLWCN + BLWAP + BLWAN) / 16
Profitability:	RMW =	(SHRCP + SHRCN + SHRAP + SHRAN + BHRCP + BHRCN + BHRAP + BHRAN + SLRCP + SLRCN + SLRAP + SLRAN + BLRCP + BLRCN + BLRAP + BLRAN) / 16 -(SHWCP + SHWCN + SHWAP + SHWAN + BHWCP + BHWCN + BHWAP + BHWAN + SLWCP + SLWCN + SLWAP + SLWAN + BLWCP + BLWCN + BLWAP + BLWAN) / 16
Investment:	CMA =	(SHRCN + BHRCN + SLWCN + BLWCN + SHRCN + BHRCN + SLWCN + BLWCN + SHRCP + BHRCP + SLWCP + BLWCP + SHRCP + BHRCP + SLWCP + BLWCP) / 16 -SHRAN + BHRAN + SLWAN + BLWAN + SHRAN + BHRAN + SLWAN + BLWAN + SHRAP + BHRAP + SLWAP + BLWAP + SHRAP + BHRAP + SLWAP + BLWAP) / 16
Sentiment:	PMN =	(SHRCP + SHRAP + BLRCP + BLRAP + SHWCP + SHWAP+ BLWCP + BLWAP + SHRCP + SHRAP+ BLRCP + BLRAP + SHWCP + SHWAP+ BLWCP + BLWAP) / 16 -(SHRCN + SHRAN + BLRCN + BLRAN + SHWCN + SHWAN + BLWCN + BLWAN + SHRCN + SHRAN + BLRCN + BLRAN + SHWCN + SHWAN + BLWCN + BLWAN) / 16

Table 8 shows summary statistics for factor returns. The average market return above the risk-free rate $Mkt - RF$ is 55.16bps per year with a high dispersion of 446.16bps. These values match the statistics given by Fama and French (2015). The average return of SMB is 20.27bps for the 2x3 sorts, 49.71bps respectively for the more granular approach. With more than 2.4 standard errors, it is only in the latter case significantly different from zero. The dispersion is higher than in the original paper which is due to the reduced dataset between 1998 and 2017. This also applies to the other three Fama/French factors which are, at most, less than 1.8 standard errors away from zero. The small average HML returns of 14.39bps, -3.07bps respectively, with high standard deviations of 344.94bps and 262.95bps pose challenge as in Fama and French (2015). This also applies for average RMW and CMA returns with 4.08 and 6.82bps for the 2x2x2x2x2 sorts. The dispersion of the factors is overall smaller than in the 2x3 sorts due to a better diversification effect. The 2x3 sorts neglect the stocks in the middle 40%. This applies to the fundamental factors only. In fact, the standard deviation of the PMN factor is comparable with 172.61 (2x2x2x2x2) to 166.49 (2x3). Overall, the dispersion is also the lowest of all factors which indicates that stocks sorted along the change of sentiment perform similarly. Average PMN returns are 22.28bps and 30.43bps for the two sorting approaches and more than 2.0, respectively 2.7, standard errors away from zero.

The correlations between the 2x3 Fama-French factors are mostly in line with the original version in terms of value and sign. Exception to this is the high correlation between the market risk premium and RMW of -0.54 compared to -0.21 in Fama and French (2015). Our sentiment factor, PMN, is significantly negatively correlated with -0.32 to the market and -0.16 to HML. The relation to RMW is significantly positive at 0.24. The alternative 2x2x2x2x2 sorting approach effectively neutralizes the factors and also reduces the correlation despite not being perfectly orthogonal. While the 2x3 sorts only control for size and one other factor, the 2x2x2x2x2 jointly neutralize for size, value, profitability, and investment with the largest impact on the PMN factor. The correlation of this factor between the 2x3 and the 2x2x2x2x2 approach is with 0.71 the lowest, followed by CMA with 0.77 and RMW with 0.80. The correlation for different

Table 8: Summary statistics for factor returns using social media-based sentiment

This table shows the summary statistics for monthly factor returns in bps using the 2x3 and 2x2x2x2x2 sorting approach for January 1998 - December 2017. $MktRF$ is the value-weighted return on the market portfolio of all sample stocks minus the one-month Treasury bill rate. For the five Fama French factors SMB, HML, RMW, and CMA we follow their exact methodology. At the end of each June, stocks are assigned to two $Size$ groups using the NYSE median market capitalization as the breakpoint. Stocks are also assigned independently to two book-to-market equity (B/M), operating profitability (OP), and investment (Inv) groups, using NYSE medians of B/M, OP, and Inv. Sentiment is constructed as the social media sentiment change compared to the rolling mean. As such, this sort is more frequently constructed than those on fundamental factors, given the higher frequency of changes in sentiment. Panel A of the table shows average monthly returns (Mean), the standard deviations of monthly returns (Std dev.) and the t-statistics for the average returns. Panel B shows the correlations between factors and Panel C the corresponding p-values. Panel D demonstrates the correlation between the different versions of the same factor from the 2x3 and 2x2x2x2x2x2 sorts.

	2x3						2x2x2x2x2					
	$Mkt-Rf$	SMB	HML	RMW	CMA	PMN	$Mkt-Rf$	SMB	HML	RMW	CMA	PMN
Panel A: Average, standard deviation and one-sample t-statistics for monthly returns												
Mean	55.16	20.27	14.39	33.27	19.38	22.28	55.16	49.71	-3.07	4.08	6.82	30.43
Std. dev	446.16	344.94	338.16	298.61	185.66	166.49	446.16	312.03	262.95	205.87	142.74	172.61
t-Statistic	1.92	0.91	0.66	1.73	1.62	2.07	1.92	2.47	-0.18	0.31	0.74	2.73
Panel B: Correlation between different factors												
$Mkt-Rf$	1.00	0.29	-0.23	-0.47	-0.28	-0.29	1.00	0.24	-0.23	-0.36	-0.19	-0.13
SMB		1.00	-0.34	-0.62	-0.03	0.11		1.00	-0.14	-0.34	0.13	0.15
HML			1.00	0.58	0.56	-0.19			1.00	0.73	0.14	-0.35
RMW				1.00	0.15	0.02				1.00	-0.03	-0.23
CMA					1.00	-0.03					1.00	-0.09
PMN						1.00						1.00
Panel C: P-value of correlations between different factors in Panel B												
$Mkt-Rf$		0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.04
SMB			0.00	0.00	0.67	0.08			0.04	0.00	0.04	0.02
HML				0.00	0.00	0.00				0.00	0.03	0.00
RMW					0.02	0.71					0.68	0.00
CMA						0.68						0.17
PMN												
Panel D: Correlation between different versions of the same factor												
2x3							1.00	0.92	0.91	0.80	0.77	0.71

versions of SMB and HML are above 0.90. Within the 2x2x2x2x2 sorts, the correlation between PMN and RMW is no longer significant and reduced to a negative relation of -0.02. The remaining correlations of sentiment to the FF5 factors remain similar in value and sign as for the 2x3 sorts, all below 0.25. The correlations between the fundamental factors are similar to the most granular sorting of 2x2x2x2 in the original paper by Fama and French (2015). SMB is positively related to the market, confirming that small stocks tend to have higher market betas than big. The correlation of HML to RMW is positive and significant confirming the original authors who explicitly outline the high correlation between RMW and HML as a negative feature of factors constructed by joint controls. This is due to the extraordinarily high return of small stocks with low B/M, weak OP, and low or high Inv which are held short in the HML and RMW factor. Overall, the results of the correlation analysis show beneficial properties for linear regressions as sentiment seems to capture different information than the fundamental factors.

In Figure 1 we plot the cumulative returns of the risk factors based on the 2x2x2x2x2 sorts from January 1998 - December 2017. During this period investors with exposure to the size factor had gained the highest yield, followed by the market and sentiment factor. For sentiment we observe spikes shortly before the Dotcom crisis in the late 90s as well as during and after the great financial crisis in 2008. Outside these windows, the sentiment premium was relatively stable compared to the other factors and most importantly always positive. The value factor has a negative cumulative return, while investment and profitability exposure gained small positive returns. This provides further guidance for incorporating sentiment into the asset pricing models in the following section.

5 Sentiment-augmented asset pricing models

In this section we turn our focus to pricing models that include our novel sentiment risk factor. The baseline is the FF5 model which is designed to model the fundamental and rational relations between average stock excess returns and market, size SMB (market capitalization in terms of price time shares outstanding), value HML (book-to-market

Figure 1: Cumulative returns of the Fama-French and sentiment risk factors

This figure plots the cumulative returns of the Fama-French and sentiment risk factors constructed by the 2x2x2x2x2 sorts from January 1998 - December 2017. SMB (small minus big market capitalization) is the size factor; HML (high minus low book-to-market ratio B/M) is the value factor; RMW (robust minus weak operating profitability OP) is the profitability factor; CMA (conservative minus aggressive investment Inv) is the investment factor. PMN (positive minus negative sentiment Snt) is the social media-based sentiment factor.



ratio), profitability RMW (operating profit), and investment CMA (growth in book equity) factors. We add the sentiment factor based on a positive-minus-negative sentiment sorts of stocks to Equation (10) in order to cover the sentiment premium for excess returns of stock portfolios:

$$\begin{aligned} R_{i,t} - R_{f,t} =& a_i + b_i(R_{m,t} - R_{f,t}) + s_i SMB_t + h_i HML_t \\ &+ r_i RMW_t + c_i CMA_t + p_i PMN_t + e_{i,t}, \end{aligned} \tag{10}$$

where $R_{i,t}$ is the return on security or portfolio i for period t, $R_{f,t}$ is the risk-free return, $R_{m,t}$ is the excess return on the value-weighted (VW) market portfolio, SMB_t is the return on a diversified portfolio of small stocks minus the return of a diversified portfolio of big stocks, HML_t is the difference between the returns of diversified portfolios of high and low B/M stocks, RMW_t is the spread between the returns of diversified portfolios with robust and weak operating profitability, CMA_t is the return difference between firms with conservative and aggressive investment activities, and $e_{i,t}$ is a zero-mean residual. The additional sentiment factor PMN_t is the difference between the returns on diversified portfolios comprised of stocks with positive and negative deviation of previous month's sentiment from the long-term mean since inception until time t.

We map the above OLS regression to the GMM framework to correct the standard errors for autocorrelation and conditional heteroskedasticity and account for potential endogeneity. The objective of GMM matches our goal to minimize the sum of squared pricing errors without the usual distributional assumptions of i.i.d., normal distribution, or homoskedasticity of the OLS framework. It also accounts for risks of endogeneity between sentiment and returns. Both the portfolio and factor sorts are based on the change in sentiment of individual securities and are rebalanced on a monthly basis. The factor-mimicking portfolio and sentiment factor PMN are thus constructed at a higher frequency than the fundamental FF5 factors. We cannot completely rule out that a sentiment shock may have an impact on the existing risk factors and change their pricing which leads back to the perpetual debate on whether sentiment can be truly seen as an independent risk factor or rather as an instrumental or state variable that affects the

fundamental pricing factors. As previously outlined, we follow the theoretical framework of Shefrin and Belotti (2008) who argue in favor of a sentiment-based risk premium. Both this theoretical evidence as well as our empirical specification elevate the GMM framework with an individual sentiment risk factor. In line with Cochrane (2009), our factor-mimicking portfolios are predestined for the GMM estimator as those portfolios are more stationary than individual stocks. Statistical characteristics of portfolio returns are more constant than the properties of individual securities which are selected and dropped from the portfolios at each rebalancing step. With sentiment-mimicking portfolios we achieve a smoothing effect which is particularly important for our specification with sentiment of individual stocks that changes very frequently, may rise steeply and drop immediately for the next rebalancing window.

We use the new sentiment risk factor to explain average excess returns of factor-mimicking and industry portfolios and benchmark our model against the FF5 model. We consider the following pricing models: i) the standard FF5 model with $R_m - R_f$, SMB, HML, RMW, CMA, and ii) a modified FF5 (FF5-SNT) including aggregate sentiment PMN. The two models are applied to the 25 VW portfolios of Table 5 and 36 VW portfolios of Table 6 as well as to the ten and 30 industry portfolios from Kenneth's homepage as an additional robustness check.

In accordance with Fama and French (2015), we apply the following test statistics:[14] i) the average absolute intercept $A|\alpha_i|$ to have a relative measure for the dispersion of intercepts. We are interested in relative performance improvements if sentiment is added to the FF5 model and test whether sentiment adds explanatory power to traditional asset pricing models. ii) The GRS statistic of Gibbons et al. (1989) to test for the hypothesis of a zero intercept.[15] If an asset pricing model is able to completely capture the variation in

[14] Given that R^2 is not applicable in the GMM framework, none is provided. Also no J-test is provided as our model specification has an exact identification with as many instruments as endogenous variables, i.e. the J-test is zero by definition.

[15] We are aware that the GRS test assumes normally distributed error terms, that are homoskedastic and serially uncorrelated but it is provided for comparative reasons and relative assessment of the two models. We do not show the p-value of getting a GRS statistic larger than the one observed if the true intercepts are all zero to avoid false inferences due to aforementioned assumptions. We are less interested whether the GRS test assesses that our models are complete descriptions of expected returns than their relative performance to the FF5 model.

expected excess returns, the intercept should be close to zero. Thus, a lower GRS score of a model points towards a comparative advantage of that particular specification over the benchmark. iii) We compute a ratio $\frac{A|\alpha_i|}{A|\bar{r}_i|}$ to estimate the unexplained proportion in the cross-section of returns. The numerator is the mean absolute intercept. The denominator measures the dispersion of dependent expected returns. We compute the difference of the time-series mean excess return on portfolio i and the cross-sectional mean of excess returns. A value greater than one suggests that intercepts are more dispersed than average returns. iv) We use the squared versions of the previous metric $\frac{M|\alpha_i^2|}{A|\bar{r}_i^2|}$, following Fama and French (2015), in order to account for biases due to measurement errors which inflate both the median absolute intercept $M|\alpha_i|$ and the average absolute deviation $A|\bar{r}_i|$. As stated by Fama and French (2015) this metric is similar to an adjusted version of $1 - R^2$ and as such a goodness-of-fit measure.[16] We use the median over the mean for the absolute intercepts considering the remarks by Fama and French (2015) that the results are distorted by extreme alphas from hard-to-value portfolios (for more details see Section 7).

Table 9 shows the results of this analysis for different factor-mimicking and industry portfolios on the left-hand side (LHS). The sentiment-augmented model generates lower absolute intercepts than the FF5 model for all LHS portfolios except the $Size - B/M$ portfolios in Panel A. The intercepts are unsurprisingly lowest for the FF5 factor-mimicking portfolios for which the right-hand side (RHS) factors are targeted. They are sensitive to the factor construction and 30-50% higher for the 2x2x2x2x2 sorts which contradicts the pattern shown by Fama and French (2015). In their study, the more granular factors tend to outperform the 2x3 sorts. We thus lead this back to the inclusion of sentiment in the extended 2x2x2x2x2 sorting. The improvement of the sentiment-augmented model is also larger for the more granular approach compared to 2x3. The highest intercepts are shown for the sentiment-mimicking portfolio which suggests that the FF5 performs particularly poorly for portfolios with a sentiment tilt. These portfolios are also the only ones for which the neutralization of the fundamental factors of the 2x2x2x2x2 sort is beneficial for the results. The reduction in average absolute intercepts amounts to 25-30%. The 36

[16] The interested reader is referred to Section 6 in Fama and French (2015) for an in-depth discussion of the last two metrics, their construction and motivation.

Table 9: Summary statistics for regressions using social media-based sentiment

This table shows the summary statistics for tests of the sentiment-augmented model benchmarked against the FF5 for January 1998 - December 2017. We test the ability of the models to explain monthly excess returns on 25 $Size - B/M$ portfolios (Panel A), 25 $Size - OP$ portfolios (Panel B), 25 $Size - Inv$ portfolios (Panel C), 25 $Size - Snt$ portfolios (Panel D), 36 $Size - B/M - OP - SNT$ Portfolios (Panel E), 36 $Size - B/M - Inv - SNT$ Portfolios (Panel F), 10 Fama-French Industry Portfolios (Panel G), and 30 Fama-French Industry Portfolios (Panel H). We show the average absolute value of the intercepts, the GRS statistic testing whether the expected values of all intercept estimates are zero, $A|\alpha_i|$, the median absolute value of the intercept over the mean absolute return on portfolio i minus the mean of the cross-sectional portfolio returns $\frac{M|\alpha_i|}{A|\bar{r}_i|}$, and the squared version of the previous ratio $\frac{M|\alpha_i|^2}{A|\bar{r}_i^2|}$ to account for biases due to measurement errors.

	2x3				2x2x2x2x2			
	$A\|\alpha_i\|$	GRS	$\frac{M\|\alpha_i\|}{A\|\bar{r}_i\|}$	$\frac{M\|\alpha_i\|^2}{\bar{r}_i^2}$	$A\|\alpha_i\|$	GRS	$\frac{M\|\alpha_i\|}{\bar{r}_i}$	$\frac{M\|\alpha_i^2\|}{A\|\bar{r}_i^2\|}$
Panel A: Size-B/M portfolios								
FF5	10.43	1.96	0.65	0.26	15.33	2.34	1.05	0.66
FF5-SNT	10.65	1.83	0.79	0.38	15.67	2.11	1.02	0.63
Panel B: Size-OP portfolios								
FF5	7.20	1.10	0.39	0.10	15.26	1.68	0.82	0.44
FF5-SNT	6.74	1.09	0.26	0.04	13.95	1.53	0.69	0.31
Panel C: Size-Inv portfolios								
FF5	10.91	1.69	0.54	0.20	15.07	2.26	0.80	0.45
FF5-SNT	9.18	1.47	0.46	0.15	13.79	1.89	0.68	0.33
Panel D: Size-SNT portfolios								
FF5	35.70	2.72	0.52	0.16	26.74	2.86	0.49	0.14
FF5-SNT	30.88	2.45	0.71	0.30	22.08	2.45	0.37	0.08
Panel E: Size-B/M-OP-SNT Portfolios								
FF5	25.63	1.81	0.59	0.24	24.50	1.71	0.85	0.50
FF5-SNT	23.53	1.65	0.60	0.24	23.17	1.47	0.66	0.30
Panel F: Size-B/M-Inv-SNT Portfolios								
FF5	24.92	1.47	0.64	0.27	23.55	1.35	0.64	0.27
FF5-SNT	20.94	1.30	0.44	0.13	21.14	1.10	0.60	0.24
Panel G: 10 Industry Portfolios								
FF5	24.25	4.58	2.27	3.79	32.39	6.28	2.77	4.98
FF5-SNT	22.86	4.91	1.91	2.60	30.79	6.67	2.78	5.02
Panel H: 30 Industry Portfolios								
FF5	23.20	1.66	0.95	0.54	33.20	2.14	1.50	1.35
FF5-SNT	22.02	1.89	1.03	0.64	33.63	2.49	1.60	1.54

factor-mimicking portfolios of the quadruple 2x2x3x3 sorts show intercepts with similar values to the industry portfolios. Here, sentiment achieves improvements between 5-20%. The GRS score is better for the sentiment-augmented models for all factor-mimicking portfolios but not for the two industry portfolios. Table 9 also shows that intercepts are more dispersed than average returns for the two industry portfolios. When we exclude those for the different 5x5 factor-mimicking portfolios, the FF5 factors based on 2x3 sorts leave between 39-65% of the average returns unexplained. This ratio is higher for the 2x2x2x2x2 sort with 49-105%.

The addition of sentiment is able to achieve remarkable improvements. The minimum of the range reduces to 26% for the 2x3, respectively 37% for the 2x2x2x2x2. Interesting is the higher value of the FF5-SNT model for the $Size - SNT$ portfolios of 71% compared to 52%. This is, however, only the case if the factors are not neutralized to each other. In fact, if we correct for estimation errors we observe the lowest ratio of only 8% for the $Size - SNT$ portfolios using the FF5-SNT model. Also the two sets of 36 factor-mimicking portfolios including sentiment show very low values of 24-30%. In general, the correction for estimation errors tend to significantly improve the evaluation and is in favor of the sentiment-augmented model. That means that the added value of the FF5-SNT model is even higher if the metric accounts for estimation errors.

6 Relation between fundamental factors and sentiment

We previously elaborated on the idea that the addition of sentiment benefits the FF5 asset pricing model. In order to better understand the relationship between fundamental factors and sentiment, we now run linear regressions for each factor against the five others. This should, in particular, answer the question of whether sentiment (partially) captures existing information in fundamental variables (and vice versa), or indeed contains novel information. In case of the excess market return, $Mkt - Rf$, the explanatory factors are SMB, HML, RMW, CMA, and PMN. Further regressions swap the dependent with one of the explanatory variables. In Panel A of Table 10, we observe that for the 2x3

sorts, PMN is, with a coefficient of -0.68, significantly negatively related to the market. PMN has a positive relation to the size factor with a coefficient of 0.32. The relation to HML is negative with -0.31. In line with the results from Fama and French (2015), HML seems to be a redundant factor as the other factors can explain around 60% of its variation, driven by the positive slopes of RMW and CMA.[17] Interestingly, the high R^2 of 63% for RMW is clearly above the 21% given by Fama and French (2015). This is not driven by the inclusion of PMN with an estimate of 0.14 and a t-statistic of 1.83 but rather by the chosen time period of January 1998 - December 2017 and the high slope of the HML factor. If we exclude PMN from the regression, the R^2 does not change to the third decimal. For our variable of interest, PMN, the results show low coefficients smaller than 0.20. The FF5 factors only explain 17% of the PMN variation, the lowest value across all regressions. This suggests that stocks with positive sentiment behave fundamentally different which cannot be explained by traditional security properties like size, value, profitability, or investment. In Panel B with the neutralized risk factors of the 2x2x2x2x2 sort, the coefficients, t-statistics, and goodness-of-fit measures tend to be even smaller. The neutralization of PMN reduces the correlation to RMW at the expense of increasing the one to CMA. The overall R^2 is with 19% on a similar level. The more granular sorting approach has a higher effect on SMB and CMA for which the R^2 levels are now even below the ones outlined by Fama and French (2015) in their 2x2x2x2 sort.

These results raise the question if there are other relevant factors that may impact the contribution of sentiment. Fama and French (2015) discuss that neither the addition of the momentum factor from Carhart (1997) nor the liquidity factor from Pástor and Stambaugh (2003) lead to significant performance improvements of the FF5 model. However, due to concerns that sentiment, in particular positive mood, may only proxy stock momentum, we extend the factor analysis accordingly in Table C.1 in the Appendix. The results show that the average return of the momentum factor UMD[18] does not significantly differ from

[17] The interested reader is referred to Section 7 from Fama and French (2015) for an in-depth discussion of the HML factor.

[18] Up-minus-down (UMD) captures the return spread between firms with high prior returns over the last twelve months (up) and low prior returns (down). The monthly prior (2-12) return breakpoints are the 30th and 70th NYSE percentiles as described on Kenneth French's homepage. The sorts comprise all NYSE, AMEX, and NASDAQ stocks with clean price history for the end of month $t-13$ to $t-2$.

Table 10: Regressions on factors using social media-based sentiment

This table shows the coefficients and t-statistics for using five factors in regressions to explain average monthly returns on the sixth for January 1998 - December 2017. $Mkt - RF$ is the value-weighted return on the market portfolio of all sample stocks minus the one-month Treasury bill rate; SMB (small minus big) is the size factor; HML (high minus low book-to-market ratio B/M) is the value factor; RMW (robust minus weak operating profitability OP) is the profitability factor; CMA (conservative minus aggressive investment Inv) is the investment factor. PMN (positive minus negative sentiment Snt) is the social media-based sentiment factor. Panel A shows the 2x3 factors from two Size groups and 3 groups of the other variable of interest. Panel B shows the 2x2x2x2x2 factors which are constructed using separate sorts of stocks into two Size groups, two B/M groups (HML), two OP groups (RMW), two Inv groups (CMA), and two Snt groups (PMN).

	α	$Mkt - Rf$	SMB	HML	RMW	CMA	PMN	R^2
Panel A: 2x3 sorts								
MktRf								
Coef	0.01		0.10	0.28	-0.73	-0.78	-0.68	0.35
t-statistic	4.35		1.20	2.57	-6.11	-4.94	-4.62	
SMB								
Coef	0.00	0.06		0.01	-0.70	0.16	0.32	0.40
t-statistic	1.63	1.20		0.08	-8.24	1.26	2.81	
HML								
Coef	-0.00	0.10	0.00		0.64	0.93	-0.31	0.60
t-statistic	-1.60	2.57	0.08		10.02	11.81	-3.46	
RMW								
Coef	0.00	-0.19	-0.32	0.47		-0.37	0.14	0.63
t-statistic	3.90	-6.11	-8.24	10.02		-4.63	1.83	
CMA								
Coef	0.00	-0.12	0.04	0.40	-0.22		0.03	0.42
t-statistic	2.72	-4.94	1.26	11.81	-4.63		0.52	
PMN								
Coef	0.00	-0.12	0.10	-0.16	0.10	0.04		0.17
t-statistic	2.43	-4.62	2.81	-3.46	1.83	0.52		
Panel B: 2x2x2x2x2 sorts								
MktRf								
Coef	0.01		0.26	0.02	-0.80	-0.79	-0.68	0.24
t-statistic	2.76		2.98	0.11	-4.12	-4.31	-4.32	
SMB								
Coef	0.00	0.14		0.29	-0.61	0.31	0.32	0.18
t-statistic	1.76	2.98		2.69	-4.31	2.26	2.76	
HML								
Coef	-0.00	0.00	0.10		0.94	0.24	-0.29	0.60
t-statistic	-0.44	0.11	2.69		15.45	2.97	-4.27	
RMW								
Coef	0.00	-0.08	-0.12	0.54		-0.19	0.00	0.63
t-statistic	2.08	-4.12	-4.31	15.45		-3.22	0.05	
CMA								
Coef	0.00	-0.09	0.07	0.15	-0.22		-0.10	0.12
t-statistic	1.44	-4.31	2.26	2.97	-3.22		-1.80	
PMN								
Coef	0.00	-0.11	0.10	-0.25	0.00	-0.14		0.19
t-statistic	3.08	-4.32	2.76	-4.27	0.05	-1.80		

zero. The t-statistic reports 1.12 standard deviation for the 2x3 sort, respectively 0.53 for the 2x2x2x2x2 approach, with high dispersion of returns of 5.27% or 3.26%. The correlation to sentiment is moderately significant with 0.56 and 0.48 if the factors are neutralized to each other. If we include sentiment in the regression of PMN, the coefficient is with 0.14 similarly low as for other factors.

We further investigate the relation between sentiment and liquidity given that overconfidence leads to the entry of retail investors, driving up liquidity, as shown by Odean (1998). In this case, our indicator may only be a proxy for the change in liquidity. As outlined above we construct the liquidity factor (LMI) as the difference between two groups of liquid and illiquid stocks, measured as the change in monthly trade volume in relation to shares outstanding. Table C.2 in the Appendix shows that for both factor constructions the average return of the LMI factor is not significantly different from zero (2x3: 31.05, t=0.97, 2x2x2x2x2: 22.07, t=1.64). The correlation to sentiment is close to zero. Along Fama and French (2015), these results do not suggest the addition of momentum or liquidity to the pricing models.

7 Regression details of hard-to-value stock portfolios

We now turn to an elaborate discussion of the regression details from Section 5. In particular, we examine the intercepts and coefficient estimations of the sentiment-augmented pricing model. To keep the analysis clear, we hereby focus on the well-known pricing puzzles of hard-to-value stocks. These comprise extreme growth and extreme investment portfolios as outlined in Fama and French (1993, 2015), small stocks that invest heavily despite low profitability, and sentiment-mimicking portfolios. Our research hypothesis states that sentiment can help to (partially) solve these puzzles that all seem to be related to irrational investor behavior.

7.1 Extreme growth stocks

Fama and French (2015) report that the FF5 model helps explain the returns of extreme growth stocks, but cannot provide all the missing pieces to resolve the puzzle. Both in the FF3 model as well as in the FF5 model the intercepts for small extreme growth stocks are significantly negative while being positive for large extreme growth stocks. In Panel A of Table 11, we clearly depict the significantly negative intercepts for small extreme growth stocks and the positive alphas of large extreme growth firms of the 25 $Size - B/M$ portfolios for the FF5 model. The large negative intercept of the growth smallcap portfolio is sufficient to reject the FF5 model to describe the expected returns of the 25 $Size - B/M$ portfolios overall. The sentiment-augmented model reduces the problem by increasing the intercept by 10bps from -0.79 to -0.69. It also increases the intercepts for the next two size quintiles of extreme growth portfolios but cannot completely explain the negative intercept for growth smallcaps and positive for growth largecaps. Panel B shows the further coefficients of the FF5-SNT model. The market slope is always close to one. As in Fama and French (2015) the SMB slope is positive for small (1.21, t=11.74) and negative for big stocks (-0.19, t=-8.98). The HML (-0.53, t=-2.13) and CMA (-0.38, t=-2.58) slopes are comparable across all growth portfolios but the RMW is steepest for growth smallcaps (-0.49, t=-2.32) and positive for growth largecaps (0.20, t=4.12). In terms of sentiment, the small growth portfolio has the most negative slope (-0.29, t=-2.53). This suggests that the returns of this portfolio are driven by smallcaps which behave like unprofitable, aggressively investing firms for which investors have a negative sentiment outlook. The sentiment coefficient for growth largecaps is zero and insignificant to further prove our point.

7.2 Extreme investment stocks

Fama and French (2015) describe the FF5 asset pricing problem of negative intercepts for high investment small stock and positive intercepts for high investment big stock portfolios. Table 9 indicates that sentiment generally improves the model performance for

Table 11: Regressions on Size-B/M portfolios using social media-based sentiment

This table shows the coefficients and t-statistics of the GMM estimator using the independent 2x2x2x2x2 factors on 25 $Size - B/M$ portfolios constructed by the 5x5 sorts. The left-hand side (LHS) variables are the monthly excess returns on the 25 $Size - B/M$ portfolios which sorts the stocks at the end of June each year into five size groups (Small to Big) and five B/M groups (Low B/M to High B/M). The right-hand side (RHS) variables are the excess market return, $Mkt - Rf$, the size factor, SMB, the value factor, HML, the profitability factor, RMW, the investment factor, CMA, and the sentiment factor, PMN, constructed using independent 2x2x2x2x2 sorts on $Size$ and each of book-to-market ratio B/M, operating profitability OP, investment Inv, and sentiment Snt. Table A shows the intercepts from the five-factor model, Panel B the intercepts and coefficients of the sentiment-augmented FF5-SNT model of the form: $R(t) - R_f(t) = a + b(Mkt - Rf) + sSMB + hHML + rRMW + cCMA + pPMN + \epsilon(t)$. The market beta is not shown to save space.

	Low	P2	P3	P4	High	Low	P2	P3	P4	High
Panel A: FF5 intercepts										
			a					**t(a)**		
Small	-0.79	-0.28	-0.09	0.03	0.01	-3.46	-1.55	-0.66	0.22	0.04
S2	-0.40	-0.13	-0.02	-0.16	-0.19	-2.87	-1.22	-0.15	-1.79	-1.58
S3	-0.27	-0.07	0.01	-0.05	0.18	-2.14	-0.58	0.12	-0.41	1.27
S4	0.00	0.05	0.20	0.13	0.11	0.03	0.40	1.44	0.96	0.73
Big	0.10	0.19	0.24	0.00	0.14	1.63	2.16	2.12	0.01	0.79
Panel B: FF5-SNT coefficients										
			a					**t(a)**		
Small	-0.69	-0.25	-0.04	-0.01	0.06	-4.96	-2.68	-0.49	-0.06	0.43
S2	-0.33	-0.12	0.01	-0.13	-0.15	-2.07	-1.31	0.15	-1.55	-1.75
S3	-0.24	-0.05	0.06	-0.03	0.24	-1.70	-0.61	0.60	-0.32	1.56
S4	-0.02	0.09	0.21	0.12	0.19	-0.22	0.99	1.67	0.93	1.31
Big	0.10	0.21	0.27	0.05	0.24	1.82	3.21	1.76	0.34	1.83
			s					**t(s)**		
Small	1.21	1.25	0.97	0.99	0.90	11.74	14.07	20.70	25.87	27.95
S2	1.01	0.96	0.76	0.85	0.89	17.70	20.12	18.73	28.31	20.38
S3	0.74	0.59	0.44	0.46	0.49	12.79	20.31	9.42	9.69	9.29
S4	0.48	0.24	0.16	0.20	0.19	9.66	5.64	3.59	3.92	7.08
Big	-0.19	-0.21	-0.22	-0.11	-0.25	-8.98	-7.52	-8.69	-4.01	-3.74
			h					**t(h)**		
Small	-0.53	-0.22	0.20	0.34	0.60	-2.13	-1.53	2.56	3.56	3.79
S2	-0.40	0.03	0.42	0.60	0.81	-6.65	0.51	6.32	13.91	8.45
S3	-0.65	0.10	0.48	0.80	0.77	-8.71	0.88	3.55	6.77	9.74
S4	-0.52	0.23	0.57	0.76	0.98	-10.67	2.07	4.74	4.87	8.98
Big	-0.40	0.14	0.62	1.01	1.20	-8.74	1.70	4.36	6.23	8.99
			r					**t(r)**		
Small	-0.49	-0.43	-0.24	-0.13	-0.19	-2.32	-4.73	-2.20	-1.62	-1.28
S2	-0.40	-0.03	0.00	0.00	-0.13	-5.60	-0.23	-0.02	0.00	-1.02
S3	-0.06	0.08	0.07	-0.13	0.04	-0.96	0.93	1.09	-1.38	0.28
S4	-0.07	0.11	0.01	-0.28	-0.21	-1.14	0.90	0.07	-3.05	-1.86
Big	0.20	0.08	-0.25	-0.23	-0.93	4.12	1.56	-2.07	-2.37	-6.68
			c					**t(c)**		
Small	-0.38	-0.22	-0.03	0.13	0.18	-2.58	-1.79	-0.39	1.72	2.02
S2	-0.38	-0.23	-0.05	0.13	0.12	-2.59	-1.86	-0.65	2.10	1.02
S3	-0.39	-0.09	-0.03	0.27	0.25	-2.52	-1.32	-0.43	3.58	2.08
S4	-0.34	0.05	0.07	0.14	0.24	-2.11	0.71	0.87	1.50	2.04
Big	-0.09	0.21	0.07	0.01	0.30	-2.68	1.97	0.82	0.12	2.06
			p					**t(p)**		
Small	-0.29	-0.10	-0.14	0.11	-0.16	-2.53	-0.87	-3.18	1.46	-2.24
S2	-0.20	-0.04	-0.09	-0.09	-0.14	-1.96	-0.76	-1.17	-1.64	-3.47
S3	-0.08	-0.05	-0.15	-0.05	-0.17	-1.13	-0.77	-2.01	-0.76	-1.52
S4	0.07	-0.12	-0.05	0.01	-0.25	1.07	-1.83	-0.60	0.12	-2.98
Big	0.00	-0.05	-0.09	-0.15	-0.31	0.03	-0.82	-0.92	-1.14	-2.46

the 25 $Size - Inv$ portfolios. Panel A of Table 12 shows the problem of highly significant negative intercepts for the three small stock portfolios in the highest investment quintile (Small: -0.59, t=-4.49; S2: -0.40, t=-3.22; S3: -0.29, t=-2.09). The FF5-SNT model increases the intercepts by 10bps, 7bps, and 4bps (Small: -0.49, t=-4.81; S2: -0.33, t=-2.81; S3: -0.25, t=-1.67). This represents a reduction of 13-18% and can be traced back to the negative slopes for PMN, in particular for smallcaps (-0.33, t=-3.45). This estimate is the lowest for all high investment portfolios. Similar to extreme growth stocks the consideration of sentiment in pricing models cannot fully explain the mechanisms behind these extreme investments but at least further our knowledge to better understand them.

7.3 Small stocks that invest heavily despite weak profitability

Fama and French (2015) describe the FF5 asset pricing problem of negative intercepts for portfolios of stocks that invest heavily despite low profitability for the 36 $Size - B/M - OP - Inv$ portfolios. Panel A of Table 13 shows the problem of highly significant negative intercepts for small growth (-0.82, t=-5.61) and small value stocks (-0.29, t=-3.09). Both portfolios contain firms that invest heavily (aggressive) despite low (weak) profitability. The coefficients are most negative if size and value are kept stable. The FF5-SNT model increases the intercepts by 10bps for small growth stocks (-0.72, t=-4.17), respectively 7bps for small value stocks (-0.22, t=-1.88). This represents an absolute reduction of 12-23%. However, the slopes for PMN are not significant. This is an interesting result as for both portfolios the slopes of the other factors change drastically. It appears that in this asset pricing puzzle, the inclusion of a sentiment factor is indirectly beneficial by changing the pricing of the existing factors. Similarly to other hard-to-value stocks the addition of sentiment to pricing models cannot resolve but significantly reduce the problem of negative intercepts of firms that invest heavily despite low profitability.

Table 12: Regressions on Size-Inv portfolios using social media-based sentiment

This table shows the coefficients and t-statistics of the regressions using the independent 2x2x2x2x2 factors on 25 $Size - Inv$ portfolios constructed by the 5x5 sorts. The left-hand side (LHS) variables are the monthly excess returns on the 25 $Size - Inv$ portfolios which sorts the stocks at the end of June each year into five size groups (Small to Big) and five investment groups (Low Inv to high Inv). The right-hand side (RHS) variables are the excess market return, $Mkt - Rf$, the Size factor, SMB, the value factor, HML, the profitability factor, RMW, the investment factor, CMA, and the social media-based sentiment factor, PMN, constructed using independent 2x2x2x2x2 sorts on $Size$ and each of book-to-market ratio B/M, operating profitability OP, investment Inv, and sentiment Snt. Table A shows the intercepts from the five-factor model, Panel B the intercepts and coefficients of the sentiment-augmented FF5-SNT model of the form: $R(t) - R_f(t) = a + b(Mkt - Rf) + sSMB + hHML + rRMW + cCMA + pPMN + \epsilon(t)$. The market beta is not shown to save space.

	Low	P2	P3	P4	High	Low	P2	P3	P4	High
Panel A: FF5 intercepts										
			a					**t(a)**		
Small	-0.15	0.04	0.16	-0.11	-0.59	-0.95	0.46	1.52	-1.07	-4.49
S2	-0.35	-0.05	0.07	-0.03	-0.40	-3.29	-0.72	0.97	-0.30	-3.22
S3	-0.09	0.03	0.10	0.03	-0.29	-0.70	0.37	0.90	0.29	-2.09
S4	0.02	0.20	0.15	0.26	-0.10	0.21	1.41	1.50	1.64	-0.70
Big	0.19	0.08	0.15	0.15	0.01	1.98	1.03	1.50	0.99	0.15
Panel B: FF5-SNT coefficients										
			a					**t(a)**		
Small	-0.15	0.04	0.16	-0.10	-0.49	-0.61	0.38	1.23	-0.88	-4.81
S2	-0.31	-0.01	0.08	-0.03	-0.33	-3.66	-0.09	1.29	-0.28	-2.81
S3	-0.01	0.03	0.10	-0.01	-0.25	-0.05	0.36	0.79	-0.14	-1.67
S4	0.10	0.27	0.16	0.19	-0.09	0.87	1.61	1.39	1.23	-0.61
Big	0.21	0.08	0.12	0.10	0.03	2.18	0.87	1.48	0.72	0.25
			s					**t(s)**		
Small	1.29	0.97	0.93	0.92	1.08	10.69	17.37	22.64	26.05	13.52
S2	0.98	0.73	0.83	0.89	1.01	18.44	12.88	32.96	38.36	18.09
S3	0.55	0.60	0.50	0.59	0.71	9.73	12.82	13.41	12.87	12.23
S4	0.19	0.12	0.22	0.34	0.42	3.48	2.00	7.87	5.62	8.44
Big	-0.22	-0.13	-0.16	-0.16	-0.20	-4.60	-3.79	-5.55	-2.77	-6.27
			h					**t(h)**		
Small	0.00	0.39	0.26	0.28	-0.08	0.01	3.62	1.94	1.86	-0.45
S2	0.38	0.34	0.37	0.39	-0.09	6.66	6.05	9.60	6.64	-1.44
S3	0.29	0.43	0.22	0.26	-0.08	3.44	4.58	3.11	2.39	-1.05
S4	0.40	0.37	0.37	0.27	-0.04	6.76	2.81	5.12	2.23	-0.77
Big	0.39	0.12	0.29	0.10	-0.42	3.55	2.07	3.23	1.34	-3.86
			r					**t(r)**		
Small	-0.58	-0.24	-0.03	-0.18	-0.11	-4.26	-2.81	-0.20	-1.09	-0.55
S2	-0.41	0.11	-0.05	0.05	-0.14	-5.35	0.85	-0.52	0.50	-1.13
S3	-0.27	-0.12	0.23	0.03	-0.07	-3.37	-1.66	2.42	0.41	-0.65
S4	-0.11	0.09	0.05	-0.09	-0.24	-0.87	0.64	0.46	-0.87	-3.52
Big	-0.30	0.05	0.02	-0.02	0.19	-2.33	1.04	0.53	-0.27	2.07
			c					**t(c)**		
Small	0.19	0.30	0.18	-0.02	-0.41	1.21	3.74	1.47	-0.20	-5.02
S2	0.34	0.34	0.11	-0.18	-0.48	3.77	3.53	1.74	-1.61	-4.24
S3	0.42	0.35	0.27	-0.12	-0.41	5.80	4.22	3.30	-1.59	-2.63
S4	0.40	0.32	0.17	-0.03	-0.59	5.23	2.76	2.58	-0.22	-5.08
Big	0.66	0.31	0.34	0.04	-0.61	3.87	4.03	3.12	0.35	-5.11
			p					**t(p)**		
Small	0.01	-0.01	-0.01	-0.06	-0.33	0.05	-0.13	-0.10	-0.67	-3.45
S2	-0.12	-0.13	-0.06	0.01	-0.20	-1.49	-1.57	-1.66	0.14	-1.33
S3	-0.26	0.01	0.00	0.13	-0.11	-3.50	0.15	0.01	2.48	-0.98
S4	-0.24	-0.22	-0.04	0.20	-0.03	-3.07	-2.12	-0.60	1.74	-0.34
Big	-0.06	0.00	0.08	0.14	-0.06	-0.49	-0.07	0.82	1.32	-0.60

Table 13: Regressions on Size-B/M-OP-Inv portfolios using social media-based sentiment

This table shows the coefficients and t-statistics (in parentheses to save space) of the regressions using the independent 2x2x2x2x2 factors on 36 $Size-B/M-OP-Inv$ portfolios constructed by the 2x2x3x3 sorts. The left-hand side (LHS) variables are the monthly excess returns on the 25 $Size - B/M - OP - Inv$ portfolios which sorts the stocks at the end of June each year into two size groups (Small to Big), two value groups (Low to High), three profitability groups (Weak to Robust), and three investment groups (Conservative to Aggressive). The right-hand side (RHS) variables are the excess market return, $Mkt - Rf$, the size factor, SMB, the value factor, HML, the profitability factor, RMW, the investment factor, CMA, and the social media-based sentiment factor, PMN, constructed using independent 2x2x2x2x2 sorts on $Size$ and each of book-to-market ratio B/M, operating profitability OP, investment Inv, and sentiment Snt. Table A shows the intercepts from the five-factor model, Panel B the intercepts and coefficients for PMN of the sentiment-augmented FF5-SNT model of the form: $R(t) - R_f(t) = a + b(Mkt - Rf) + sSMB + hHML + rRMW + cCMA + pPMN + \epsilon(t)$. The market beta is not shown to save space.

	Small						**Big**					
	Low			High			Low			High		
	Weak	P2	Robust	Weak	P2	Robust	Weak	P2	Robust	Weak	P2	Robust
Panel A: FF5 intercepts												
						a						
Conservative	-0.39	-0.12	0.09	-0.14	-0.08	-0.24	-0.08	0.03	0.20	0.04	0.40	-0.04
	(-2.06)	*(-0.81)*	*(0.82)*	*(-1.46)*	*(-0.8)*	*(-1.21)*	*(-0.23)*	*(0.16)*	*(1.43)*	*(0.24)*	*(3.2)*	*(-0.16)*
S2	-0.06	0.06	0.11	0.09	0.08	-0.03	-0.06	0.18	0.14	0.27	0.29	0.26
	(-0.29)	*(0.89)*	*(0.96)*	*(0.8)*	*(1.04)*	*(-0.15)*	*(-0.22)*	*(1.64)*	*(1.84)*	*(2.3)*	*(2.44)*	*(1.16)*
Aggressive	-0.82	-0.14	-0.20	-0.29	-0.07	0.04	-0.14	-0.08	0.14	-0.05	0.02	0.02
	(-5.61)	*(-0.87)*	*(-3.59)*	*(-3.09)*	*(-0.65)*	*(0.19)*	*(-0.99)*	*(-0.99)*	*(1.54)*	*(-0.39)*	*(0.12)*	*(0.07)*
Panel B: FF5-SNT coefficients												
						a						
Conservative	-0.36	-0.10	0.18	-0.16	-0.01	-0.11	0.03	-0.01	0.27	0.11	0.41	0.03
	(-1.91)	*(-0.64)*	*(1.35)*	*(-1.39)*	*(-0.11)*	*(-0.53)*	*(0.14)*	*(-0.08)*	*(2.43)*	*(0.75)*	*(2.63)*	*(0.11)*
S2	-0.07	0.09	0.16	0.07	0.09	0.01	-0.05	0.17	0.10	0.23	0.28	0.21
	(-0.36)	*(0.9)*	*(1.3)*	*(0.7)*	*(0.96)*	*(0.08)*	*(-0.17)*	*(1.43)*	*(1.18)*	*(1.72)*	*(1.99)*	*(0.86)*
Aggressive	-0.72	-0.13	-0.17	-0.22	-0.05	0.07	-0.01	-0.09	0.12	0.05	0.00	0.07
	(-4.17)	*(-1.1)*	*(-1.49)*	*(-1.88)*	*(-0.44)*	*(0.32)*	*(-0.03)*	*(-0.8)*	*(1.21)*	*(0.32)*	*(0.01)*	*(0.29)*
						p						
Conservative	-0.76	-0.27	-0.32	-0.25	-0.10	-0.61	-1.21	-0.16	-0.41	-0.01	-0.68	-0.86
	(-4.4)	*(-3.45)*	*(-3.54)*	*(-2.14)*	*(-0.96)*	*(-4.13)*	*(-5.73)*	*(-2.00)*	*(-3.44)*	*(-0.09)*	*(-3.62)*	*(-5.07)*
S2	-0.11	-0.04	-0.28	0.06	-0.23	-0.41	-0.36	0.13	-0.21	-0.24	-0.06	-0.22
	(-0.59)	*(-0.36)*	*(-2.33)*	*(0.59)*	*(-2.82)*	*(-2.9)*	*(-1.49)*	*(0.85)*	*(-2.39)*	*(-2.32)*	*(-0.55)*	*(-1.01)*
Aggressive	0.03	-0.10	-0.16	0.05	-0.04	-0.14	-0.05	0.01	0.13	0.12	0.05	0.15
	(0.23)	*(-1.32)*	*(-1.99)*	*(0.71)*	*(-0.51)*	*(-0.87)*	*(-0.22)*	*(0.1)*	*(2.06)*	*(0.99)*	*(0.44)*	*(1.1)*

7.4 Sentiment-mimicking portfolios

We now evaluate the model performance on sentiment-mimicking portfolios. Table 9 shows that the 25 $Size - Snt$ portfolios are much more difficult to price than the other factor-mimicking portfolios and that the inclusion of sentiment is greatly beneficial. This is not surprising since the sentiment factor PMN is predestined to explain the returns of such portfolios. In Panel A of Table 14, the intercepts of the FF5 model show that the fundamental factors do a surprisingly good job for portfolios with a negative sentiment tilt. The portion of expected returns left unexplained is close to zero and ranges from -0.08 (t=0.69) to 0.23 (t=1.81). Only the intercept for largecaps with negative sentiment is significant. Portfolios with a positive sentiment tilt are, however, more difficult to value and the intercepts range from 0.07 (t=0.59) to 0.74 (t=4.24). Smaller stocks with positive sentiment outlook are harder to price than larger stocks. The inclusion of sentiment improves the results with less significant intercepts across all 25 $Size - Snt$ portfolios. A portfolio of small stocks with positive sentiment only has a significant exposure to the size (1.08, t=13.52), sentiment (-0.33, t=-0.45) and investment factor (-0.41, t=-5.02). Large stocks with a positive sentiment tilt are negatively exposed to the size factor (-0.20, t=-6.27), value (-0.42, t=-3.86), profitability (0.19, t=2.07), and investment (-0.61, t=-5.11). This suggests that this portfolio is fundamental- and not sentiment-driven which is similar for big stocks with a negative sentiment tilt (SMB: -0.22, t=-4.60; HML: 0.39, t=3.55; CMA: 0.66, t=3.87). Different is the slope of RMW (-0.30, t=-2.33), i.e., that this portfolio of big stocks with negative sentiment behaves like big stocks with weak profitability. The coefficient to sentiment is insignificant (-0.06, t=-0.49). Small stocks with negative sentiment are surprisingly not susceptible to PMN either (0.01, t=0.05), but have a significant positive slope for SMB (1.29, t=10.69) and negative for RMW (-0.58, t=-4.26). Similar to big stocks with negative sentiment being like big stocks with weak profitability, smallcaps with low sentiment behave like small stocks with weak profitability. Overall, portfolios with sentiment tilt are challenging to price, even when a factor that particularly targets this portfolio style is included.

Table 14: Regressions on Size-Snt portfolios using social media-based sentiment

This table shows the coefficients and t-statistics of the regressions using the independent 2x2x2x2x2 factors on 25 $Size-Snt$ portfolios constructed by the 5x5 sorts. The left-hand side (LHS) variables are the monthly excess returns on the 25 $Size-Snt$ portfolios which sorts the stocks at the end of June each year into five size groups (Small to Big) and on a monthly basis into five sentiment groups (Positive to Negative Snt). The right-hand side (RHS) variables are the excess market return, $Mkt-Rf$, the size factor, SMB, the value factor, HML, the profitability factor, RMW, the investment factor, CMA, and the social media-based sentiment factor, PMN, constructed using independent 2x2x2x2x2 sorts on $Size$ and each of book-to-market ratio B/M, operating profitability OP, investment Inv, and sentiment Snt. Table A shows the intercepts from the five-factor model, Panel B the intercepts and coefficients of the sentiment-augmented FF5-SNT model of the form: $R(t)-R_f(t) = a+b(Mkt-Rf)+sSMB+hHML+rRMW+cCMA+pPMN+\epsilon(t)$. The market beta is not shown to save space.

	Negative	P2	P3	P4	Positive	Negative	P2	P3	P4	Positive
Panel A: FF5 intercepts										
			a					**t(a)**		
Small	0.09	0.43	0.81	1.16	0.74	0.69	2.03	2.53	4.25	4.24
S2	-0.08	0.15	0.21	0.32	0.33	-0.61	0.82	1.35	2.32	3.50
S3	0.03	0.00	0.21	0.07	0.33	0.24	-0.02	1.16	0.62	2.10
S4	0.09	-0.06	0.18	0.45	0.07	0.78	-0.51	1.35	4.14	0.59
Big	0.23	-0.16	0.05	0.24	0.18	1.81	-1.50	0.73	3.43	2.10
Panel B: FF5-SNT coefficients										
			a					**t(a)**		
Small	-0.15	0.04	0.16	-0.10	-0.49	-0.61	0.38	1.23	-0.88	-4.81
S2	-0.31	-0.01	0.08	-0.03	-0.33	-3.66	-0.09	1.29	-0.28	-2.81
S3	-0.01	0.03	0.10	-0.01	-0.25	-0.05	0.36	0.79	-0.14	-1.67
S4	0.10	0.27	0.16	0.19	-0.09	0.87	1.61	1.39	1.23	-0.61
Big	0.21	0.08	0.12	0.10	0.03	2.18	0.87	1.48	0.72	0.25
			s					**t(s)**		
Small	1.29	0.97	0.93	0.92	1.08	10.69	17.37	22.64	26.05	13.52
S2	0.98	0.73	0.83	0.89	1.01	18.44	12.88	32.96	38.36	18.09
S3	0.55	0.60	0.50	0.59	0.71	9.73	12.82	13.41	12.87	12.23
S4	0.19	0.12	0.22	0.34	0.42	3.48	2.00	7.87	5.62	8.44
Big	-0.22	-0.13	-0.16	-0.16	-0.20	-4.60	-3.79	-5.55	-2.77	-6.27
			h					**t(h)**		
Small	0.00	0.39	0.26	0.28	-0.08	0.01	3.62	1.94	1.86	-0.45
S2	0.38	0.34	0.37	0.39	-0.09	6.66	6.05	9.60	6.64	-1.44
S3	0.29	0.43	0.22	0.26	-0.08	3.44	4.58	3.11	2.39	-1.05
S4	0.40	0.37	0.37	0.27	-0.04	6.76	2.81	5.12	2.23	-0.77
Big	0.39	0.12	0.29	0.10	-0.42	3.55	2.07	3.23	1.34	-3.86
			r					**t(r)**		
Small	-0.58	-0.24	-0.03	-0.18	-0.11	-4.26	-2.81	-0.20	-1.09	-0.55
S2	-0.41	0.11	-0.05	0.05	-0.14	-5.35	0.85	-0.52	0.50	-1.13
S3	-0.27	-0.12	0.23	0.03	-0.07	-3.37	-1.66	2.42	0.41	-0.65
S4	-0.11	0.09	0.05	-0.09	-0.24	-0.87	0.64	0.46	-0.87	-3.52
Big	-0.30	0.05	0.02	-0.02	0.19	-2.33	1.04	0.53	-0.27	2.07
			c					**t(c)**		
Small	0.19	0.30	0.18	-0.02	-0.41	1.21	3.74	1.47	-0.20	-5.02
S2	0.34	0.34	0.11	-0.18	-0.48	3.77	3.53	1.74	-1.61	-4.24
S3	0.42	0.35	0.27	-0.12	-0.41	5.80	4.22	3.30	-1.59	-2.63
S4	0.40	0.32	0.17	-0.03	-0.59	5.23	2.76	2.58	-0.22	-5.08
Big	0.66	0.31	0.34	0.04	-0.61	3.87	4.03	3.12	0.35	-5.11
			p					**t(p)**		
Small	0.01	-0.01	-0.01	-0.06	-0.33	0.05	-0.13	-0.10	-0.67	-3.45
S2	-0.12	-0.13	-0.06	0.01	-0.20	-1.49	-1.57	-1.66	0.14	-1.33
S3	-0.26	0.01	0.00	0.13	-0.11	-3.50	0.15	0.01	2.48	-0.98
S4	-0.24	-0.22	-0.04	0.20	-0.03	-3.07	-2.12	-0.60	1.74	-0.34
Big	-0.06	0.00	0.08	0.14	-0.06	-0.49	-0.07	0.82	1.32	-0.60

8 News-based sentiment

In our last section, we investigate whether sentiment from social media offers unique and more suitable results than other sentiment factors. During the introduction of the RMI indicators, we differentiated between social media- and news-based sentiment. In this section, we re-run most of the analysis, but this time with sentiment extracted from public news. We hypothesize that this source of sentiment captures more factual information than investors' mood.

The 5x5 sentiment-mimicking portfolios in Table B.1 of the Internet Appendix do not show the same return spreads between positive and negative sentiment-tilted stocks. For smallcaps, the portfolio with positive sentiment returns on average 24.87bps per month less than the portfolio with negative stocks. For largecaps the difference still amounts to 5.34bps. The more granular 2x2x3x3 portfolios in Table B.2 of the Internet Appendix are even more negatively-tilted and sometimes outperform portfolios with positive sentiment (and vice versa). If sentiment is neutralized for other fundamental factors, no clear pattern for a sentiment effect can be distinguished.

The factor analysis in Table B.3 reveals that the average return of PMN is not significantly different from zero for neither the 2x3 (3.80bps, t=0.27) nor the 2x2x2x2x2 approach (6.69bps, t=0.69). 2x3 PMN is negatively related to the market (-0.25) and HML (-0.11) at the 10% confidence level, while the 2x2x2x2x2 PMN is negatively related to all factors except SMB. The correlation between different versions of PMN is only 0.66 and as such more dependent on the sorting approach.

Table B.4 of the Internet Appendix summarizes the statistics of regressions with the same LHS portfolios from Table 9, though sentiment-mimicking portfolios are based on news-based sentiment as are the PMN factors. We observe that the FF5-SNT model has a marginally better explanatory power than the FF5 model. Comparing its performance with the social media-based regressions, the news-based model falls short in all categories. First, the FF5 model itself does a better job on news-based sentiment-mimicking portfolios in terms of lower intercepts than for the social-media based portfolios. That means that

fundamental factors from the FF5 framework leave less variance in expected returns unexplained. This points into the direction that those sentiment-tilted stocks behave more like fundamental-driven stocks. Second, the improvement of the FF5-SNT over the FF5 is still marginal, even though the PMN factor is explicitly targeted for those portfolios. Adding sentiment to the pricing model only marginally changes its performance.

One may still be concerned that social media-based sentiment measures fundamental information instead of pure sentiment. In order to purify sentiment, we orthogonalize the social media-based risk factor against the news-based and show the results of the regressions in Table B.5. Compared to Table 9 orthogonalization worsens the results marginally. The outperformance of the sentiment-augmented FF5-SNT over the traditional FF5 model becomes smaller in terms of absolute alphas, GRS scores, and the two measures of dispersion of the intercepts. Only for $Size - B/M - Inv - SNT$ portfolios in Panel F, the variation of alphas is now higher than for the FF5 model. We acknowledge that the social media-based sentiment indicator does capture some notions of the fundamental news flow. However, the differences are marginal and do not motivate further studies. We conclude that there is not sufficient evidence to reject our hypothesis that social media-based sentiment contains more novel information about investors' moods than news-based and is thus more suitable to capture the sentiment risk premium from Shefrin and Belotti (2008).

9 Conclusion

By using the recently developed search-based and bottom-up sentiment indicator from MarketPsych, we construct a novel risk factor from a positive-minus-negative sentiment-portfolio sorts and reveal patterns in average returns related to investors' mood. The sentiment score is beneficial in augmenting the existing fundamental asset pricing model by Fama and French (2015) for the U.S. equity market and add the sentiment premium to the pricing paradigm from Shefrin and Belotti (2008). We can trace back the significant performance improvements for different factor-mimicking and industry portfolios to investors'

activities in social media. The same effect is not found exploiting sentiment extracted from public news channels. Different notions of human mood are expressed in social media and can be used to form a monthly-rebalanced sentiment risk factor which is orthogonal to underlying macroeconomic or business cycle-related developments. It is also unrelated to existing fundamental pricing factors and independent of momentum. Therefore, we believe that this novel risk factor successfully captures and quantifies investor sentiment.

In two sets of portfolio sorts with either 25 or 36 portfolios using 5x5 double-sorting or 2x2x3x3 quadruple-sorting, we show a high average excess return spread between stocks with positive and securities with negative sentiment. This spread is significantly higher than for other style tilts like size, value, profitability or investment. We use those factor-mimicking portfolios in a sentiment-augmented asset pricing model to show that the inclusion of sentiment reduces the intercepts and improves the model performance. We explore two different versions of the sentiment factor, one using the original 2x3 double-sorting from Fama and French (2015) and the other a more granular 2x2x2x2x2 sorting which effectively neutralizes all five factors against each other. While the latter produces cleaner risk premia for the various factors, the more granular approach suffers from less diversified portfolios. This becomes more challenging when the list of factors grows, e.g., by adding a momentum or liquidity factor. However, the performance improvement by sentiment is independent of the sorting approach on the factor.

Our models account for endogeneity, autocorrelation of returns and heteroskedasticity by mapping them to the GMM framework and generating robust standard errors. We show that similar to the five-factor model from Fama and French (2015), the GRS test rejects our sentiment-augmented model to fully explain the cross-section of expected returns for different style or industry portfolios. However, it improves the five-factor model in the cross-section along various dimensions and provides an explanation, at least partially, for asset pricing puzzles of hard-to-value stocks. Our sentiment-augmented models supply an additional component for solving the pricing mosaic for portfolios of small extreme growth and extreme investment stocks. Sentiment increases the negative intercept by 13-17%. For another pricing puzzle of small stocks that invest heavily despite low profitability,

sentiment has an indirect contribution by modifying the slopes of the other factors. We also report a very poor performance of the five-factor model for sentiment-tilted portfolios. Unsurprisingly, the inclusion of the sentiment factor here is particularly beneficial as it is predestined to explain the excess returns of those style portfolios. Judged on regression slopes, both largecaps and smallcaps with negative sentiment seem to behave like stocks with weak profitability. The study's findings contribute to overcoming the research challenges put forward by Fama and French (2015) regarding behavioral stories for stocks that invest heavily despite low profitability. Investor sentiment on stock outlooks expressed in social media channels can add crucial knowledge to answer this question. As such, the wisdom of the crowd seems to matter in U.S. equity pricing.

A Details on Refinitiv-MarketPsych's sentiment construction

MarketPsych's language processing engine goes beyond traditional textual sentiment analysis based on a one-dimensional output of positive or negative sentiment and a notion of neutrality, but exploits a broad range of human emotions. A common classification system of human emotions uses two dimensions known as valence and arousal[19] as psychological research has demonstrated that more than just one dimension has predictable effects on investor behavior (see, e.g., Peterson, 2007; Shu, 2010). Besides positivity or negativity in terms of valence, the level of arousal has been shown to map directly to cognitive performance through an inverse U-curve relationship, the so-called Yerkes-Dodson's Law, capturing both the reduction in complex problem solving skills when stress levels are high and the reduction in attention and reaction times when arousal levels are low (see, e.g., Diamond et al., 2007; Yerkes & Dodson, 1908).

MarketPsych uses this classification system following the affective circumplex model of sentiment by Russell (1980) and constructs RMI indicators spanning the entire multi-dimensional plane of human emotions. Among others, their aggregated sentiment reflects notions of fear, optimism, and joy. As argued by Shen et al. (2017), those are the three most commonly documented emotions in the finance literature. Optimism is generally defined as the tendency to overestimate the future payoffs of a financial asset. Such an overconfidence from investors may result in deviations of asset prices from intrinsic values as observed during extreme bullish or overheated markets. Odean (1998) finds that overconfidence leads to the entry in the market by retail investors, driving up liquidity. Ciccone (2003) reports lower returns for firms characterized by optimistic versus those driven by pessimistic expectations. On the contrary, fear leads to demand shocks, driven by investors' emotional stress, increasing market uncertainty and volatility. Da et al. (2015) establish a daily fear index based on the online searches of U.S. households, predicting

[19] Valence refers to the positive or negative affectivity, while arousal measures the level of calmness or excitement of a statement or news item.

return reversals and volatility. Strongly negative emotions like anger, fear, and gloom, all of which are captured by RMI, bias human decision making and lead to a range of behaviors like herding or panic and affect trading activities, for instance by triggering either under- or over-reactions (see, e.g., Daniel et al., 1998; Lerner & Keltner, 2001; Lerner et al., 2004; Winkielman et al., 2005, 2007). Wright and Bower (1992) found that pleasant emotions like bliss, joy, and optimism affect the subjective probability assessments of uncertain outcomes and therefore influence investors' decision-making as documented by Dolan (2002). Figure A.1 depicts several among the RMI sentiments that are described in detail in Table A.2 devoted to explaining the affective circumplex. Each dot in the figure corresponds to the emotion's location on the circumplex, whereby RMI indicators are themselves hybrids of multiple emotions according to the original framework. The thin grey line connects the positive and negative poles of matching indicators. The RMI sentiment indicator itself spans the entire multi-dimensional plane of the circumplex as described in detail in Table A.1. The table shows that RMI's construction of the sentiment indicator is tilted towards capturing negative statements because, as confirmed in private exchanges, MarketPsych research on business and financial language has found a prevalence of concepts with negative versus positive valence. As a result, the sentiment indicator is usually negative in net terms.

To provide insights about the mechanism of sentiment indicator construction, MarketPsych provides an example on the complex language processing system that reveals how they address some common pitfalls in news and social media sentiment analysis. Figure A.2 evaluates the opinion of a Goldman Sachs' analyst about his expectations of the following day's quarterly call of Apple Inc. of increasing profit margins. MarketPsych is able to differentiate between forward-looking statements and general chatter by breaking down concepts into forecasts (future tense) versus present or past observations. For instance, the PriceForecast category is a future-tense subset of PriceDirection. "The price of Apple rose last week" is a PriceDirection-only reference while "The price of Apple will rise" would be attributed to both PriceDirection and PriceForecast.

In order to have a correct attribution of articles to the right time window MarketPsych

Table A.1: Construction of the Refinitiv-MarketPsych sentiment indicator

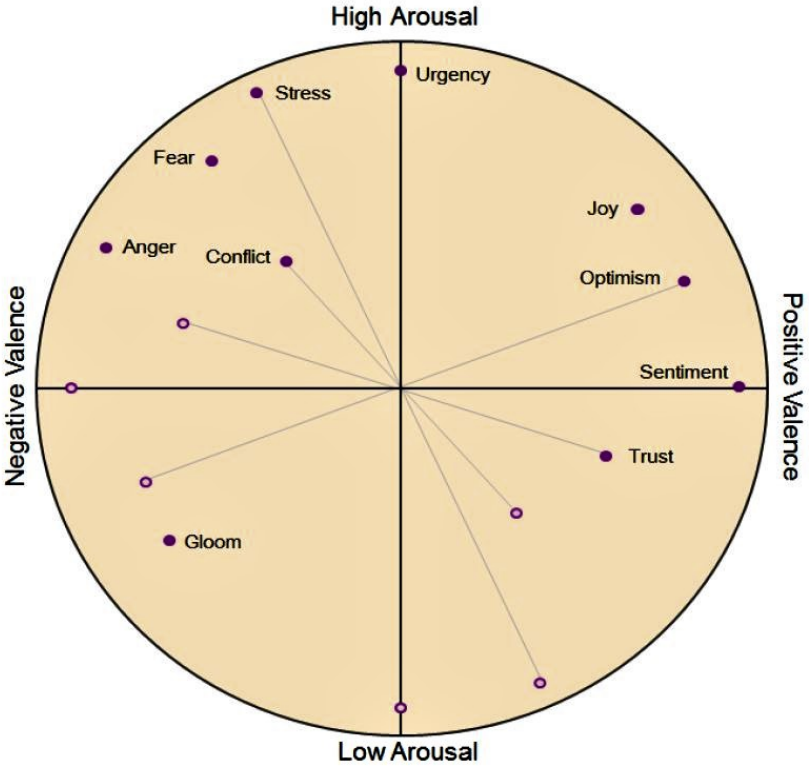
This table provides the construction details for the Refinitiv-MarketPsych (RMI) Sentiment indicator as the net of positive and negative references along the valence-arousal sentiment classification system.

Positive References	Negative References
Positive	Negative
AccountingGood	AccountingBad
Upgrade	Downgrade
EconomicPositive	EconomicNegative
EconomistPositive	EconomistNegative
EconomicActorsPositive	EconomicActorsNegative
ManagementGood	ManagementBad
BullVerbs	BearVerbs
ExcitementPos	
FearDown	FearUp
AngerDown	AngerUp
HappyUp	HappyDown
GloomDown	GloomUp
OptimismUp	OptimismDown
PessimismDown	PessimismUp
LoveUp	LoveDown
HateDown	HateUp
InnovativeUp	InnovativeDown
EarningsSurprisePos	EarningsSurpriseNeg
EarningsUp	EarningsDown
EarningsExpectationsUp	EarningsExpectationsDown
EarningsGuidanceUp	EarningsGuidanceDown
GuidanceUp	GuidanceDown
	ProfitWarning
	CatastropheConcept
	DeclareBankruptcy

Source: MarketPsych

Figure A.1: Sentiment classification system: valence and arousal

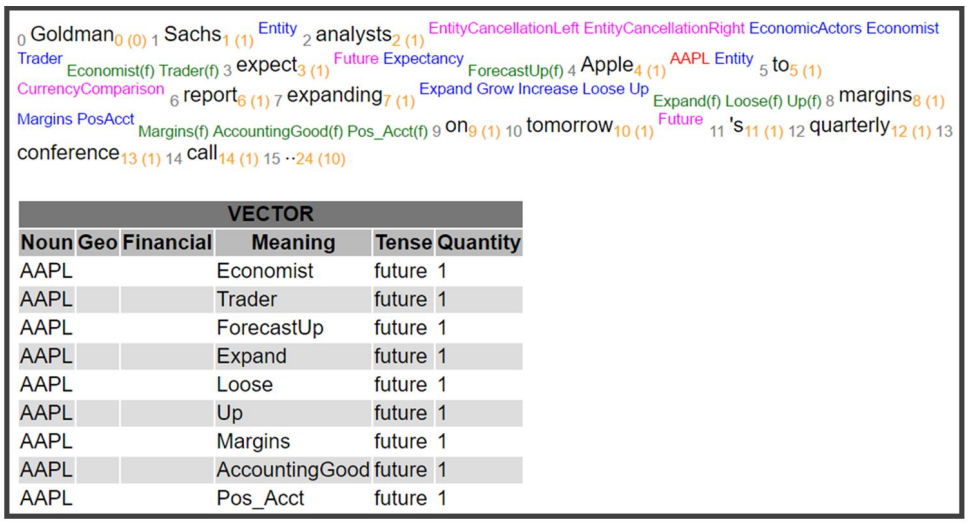
This figure plots a common classification system for human emotions along two dimensions: valence and arousal. MarketPsych uses this classification system following the affective circumplex model of sentiment by Russell (1980) and constructs RMI indicators spanning the plane of human emotions. The figure depicts several of the RMI sentiment indicators on the affective circumplex. Each dot corresponds to the emotion's location on the circumplex, whereby RMI indicators are themselves hybrids of multiple emotions according to the original framework. The grey lines connect the positive and negative poles of matching indicators.



Source: MarketPsych

Figure A.2: Example of MarketPsych's human language processing system

This figure depicts an example of how MarketPsych processes news and evaluates human emotions. Each term is annotated by MarketPsych. Complex meanings such as $AccountingGood_f$ are extracted. This is a forward looking assessment based on the attribute "tomorrow". "Goldman Sachs" is ignored as an irrelevant entity because it relates to the analyst, while "Apple" is correctly recognized as the object of interest. MarketPsych differentiates between value-adding statements as above versus irrelevant terms. Those irrelevant terms are excluded from the score vector and are not used in RMI calculations.



VECTOR					
Noun	Geo	Financial	Meaning	Tense	Quantity
AAPL			Economist	future	1
AAPL			Trader	future	1
AAPL			ForecastUp	future	1
AAPL			Expand	future	1
AAPL			Loose	future	1
AAPL			Up	future	1
AAPL			Margins	future	1
AAPL			AccountingGood	future	1
AAPL			Pos_Acct	future	1

Source: MarketPsych

also limits article consumption to those less than 2,500 words as longer articles usually take longer to write and are unlikely to be timely. In order to avoid the impact of stale news, content that has been published more than 24 hours before a given time t, is excluded. That means all content drops out of the 24 hours averages when it has been more than 24 hours since its publication. Articles that are more than 98% similar to articles recorded in the previous 24 hours are removed from analysis to avoid double-counting.[20]

Various sources, though limited to English, are used to inform the data feed of the language processing system used by MarketPsych. These include news publishers like Refinitiv and Bloomberg, electronic databases like the U.S. Securities and Exchange Commission's Edgar repository of company filings, direct press releases by companies,

[20]In the case of social media for which the concepts of re-tweeting, re-posting, and commenting are defined, MarketPsych employs a rigorous approach to cleanse the data. RMI indicators do not include re-tweets, unless they include additional commentary or remarks about the original tweet. RMI does not include comments with the same title that are repeated multiple times; however, they do include commentary text when it changes from post to post.

Table A.2: Description of Refinitiv-MarketPsych indicators

This table provides a detailed description of Refinitiv-MarketPsych (RMI) indicators to better understand the aggregated sentiment measure.

RMI COMMON NAME	ANTICIPATED MARKET IMPACT
Sentiment	There are several important research findings related to sentiment and price movement. Based on academic research on Thomson Reuters News Analytics sentiment scores, positive and negative sentiment in the news about individual stocks extend price momentum, which is supported by additional evidence that traders collectively under-react to negative sentiment in news reports. Another study finds that market sentiment improves factor weighting in some models. In foreign exchange, news sentiment was found to influence volatility.
Optimism	There is empirical evidence that proxies for optimism correlate with positive price behavior and that bullish comments in financial social media precede higher trading volume. Optimism in earnings press releases was found correlated with future stock price activity.
Fear	Academic researchers who aggregated search terms they deemed reflective of economic fear found short-term mean reversion in prices when fear-related search terms spiked in volume. In experimental markets, fear was found to decrease bid and increase ask prices, leading to less overall trading activity. As a result, we expect wider bid-ask spreads when fear is high.
Joy	Joy is a marker of exuberance. Experimental markets demonstrate higher price peaks and larger collapses during bubble simulations if traders watched a positively exciting movie clip before trading begins.
Trust	Trust was designed specifically for nations and banking and financial groups. Economists have found that national interpersonal Trust levels correlate with future economic growth.
Conflict	The Conflict RMI, which is intended to capture disagreement and dispute, is anticipated to correlate with price volatility. A study of international markets found that global conflicts significantly impact asset prices.
Stress and Urgency	Urgency and Stress are high-arousal indices that vary in valence. Based on evidence that arousal drives cognitive performance in an inverse-U shaped curve, we infer that pricing anomalies are more likely to emerge at low or high arousal values, as seen with both high positive and high negative arousal during research into experimental market bubbles.
Uncertainty	Researchers found that high-uncertainty equities and country indices on average outperform their low-uncertainty peers. 39 In contrast, during speculative bubbles uncertainty amplifies the price momentum of positive sentiment. In emerging fixed income markets, releases of macroeconomic data decrease future volatility.
Gloom	Traders in an experimental market offered lower ask and high bid prices when "sadness" was induced prior to trading, leading to increased transaction volume. If this result transfers into larger market behavior, we expect increased trading volume during periods of high Gloom. Researchers speculate that identified semi-annual variations in country stock index returns - which scale by latitude and reverse from northern to southern hemispheres - may be caused by seasonal changes in affect (the "winter blues") among local traders.
Anger	Traders induced to feel anger in an experimental market decrease both average ask and bid prices. As a result, we speculate that higher RMI Anger readings should lead to increased selling and reduced buying in associated assets, leading to downward pressure on prices during high Anger periods.

Source: MarketPsych

transcripts of conference calls, websites, blogs, and especially posts on social media like Twitter and Yahoo's stock message boards. We use the aggregate measure that reflects activities through all types of channels, news and social media. The indicators are updated at a one-minute frequency and the system works 24/7, continuously scanning all the tracked sources. In order to construct a daily record, 24 hours or 1440 minutes are aggregated into one daily observation. If no records are found for the constituents of a specific equity index, a "N/A" is returned and the observation is not stored. This implies that the retrievable time series of each individual sentiment indicator are not equally spaced over time. In practical terms, if no observation is found, no $Buzz$ is recorded and the time series fails to be updated. Crucially, such a case needs to be differentiated from true "0" values, where positive and negative statements concerning an asset exactly balance each other.[21]

For the purposes of our investigation, we cumulate the RMI index at a lower, monthly frequency thus aggregating the original, higher daily frequency using Equation (11). A monthly frequency appears to strike a reasonable balance between a sufficient granularity of the data and a need to control for the risk of using a noisy estimator of sentiment.[22] There are no missing observations in our monthly sample of international equity indices. Let $Buzz_0$, $Buzz_{-1}$, $Buzz_{-(T-1)}$ and RMI_0, RMI_{-1}, $RMI_{-(T-1)}$ represent the corresponding $Buzz$ RMI data for a given equity market, content source, and timestamp over the past T days. The $Buzz$-weighted average RMI over the trailing T-day window length is then computed as:[23]

$$\frac{(Buzz_0 * RMI_0 + Buzz_{-1} * RMI_{-1} + \ldots + Buzz_{-(T-1)} * RMI_{-(T-1)})}{(Buzz_0 + Buzz_{-1} + \ldots + Buzz_{-(T-1)})}. \tag{11}$$

[21] Positive and negative references that net each other out may still signal increased uncertainty in the market and disagreement between investors and potentially lead to higher trading activity. However, in private exchanges, MarketPsych has confirmed to us that the primary relationship is that sentiment RMI variability rises as the overall $Buzz$ decreases. So $Buzz$ is the primary determinant of sentiment dispersion.

[22] In an unreported exploratory analysis, we checked that sentiment fluctuates massively at daily frequencies, whereas at a monthly frequency it suffers from a loss of valuable information that, however, appears to be manageable. This analysis is available upon request.

[23] This definition ensures comparability of sentiment between different stocks as outlined by MarketPsych in their research guidelines, accessible at `https://old.marketpsych.com/guide/`.

B News-based sentiment

Table B.1: 5x5 portfolios using news-based sentiment

This table shows average monthly excess returns in bps for portfolios formed on *Size* and *Snt* for January 1998 - December 2017. The sorting follows Fama and French (2015) using only NYSE breakpoints: at the end of each June, stocks are allocated to five Size groups (Small to Big). For sentiment we sort the stocks into five sentiment groups (Positive to Negative) based on the previous month's deviation of news-based sentiment from the long-term rolling mean. The intersections produce 25 value-weight $Size - Snt$ portfolios.

	Negative	2	3	4	Positive
Small	180.42	138.86	176.08	184.48	155.55
2	121.87	125.20	119.29	100.75	112.12
3	115.98	106.60	75.35	92.48	106.31
4	66.37	78.95	84.21	89.09	121.99
Big	59.23	36.75	62.10	59.23	64.57

Table B.2: 2x2x3x3 portfolios using news-based sentiment

This table shows average monthly excess returns in bps for portfolios formed on *size* and B/M, OP, Inv, Snt for January 1998 - December 2017. The sorting follows Fama and French (2015) using only NYSE breakpoints: at the end of each June, stocks are allocated to two Size groups (Small and Big), two B/M groups (Low and High), three groups of either profitability or investment and three groups of Snt. For sentiment we sort the stocks based on the previous month's deviation of news-based sentiment from the long-term rolling mean. The intersections produce 36 portfolios.

	Small						Large					
	Low B/M			High B/M			Low B/M			High B/M		
	$s(-)$	$s(0)$	$s(+)$	$s(-)$	$s(0)$	$s(+)$	$s(-)$	$s(0)$	$s(+)$	$s(-)$	$s(0)$	$s(+)$
Panel A: Size-B/M-OP-SNT Portfolios												
Low	159.69	88.39	152.47	110.88	110.39	70.26	18.87	65.67	106.79	30.42	65.52	62.49
Medium	122.76	79.88	96.50	88.05	100.42	111.64	47.79	66.51	71.00	43.86	84.84	78.17
High	159.64	85.86	122.48	96.09	87.35	166.49	55.53	64.91	67.53	89.08	79.23	80.63
Panel B: Size-B/M-Inv-SNT Portfolios												
Low	127.86	166.58	151.54	87.12	92.08	96.22	82.74	59.29	71.55	52.00	76.10	63.80
Medium	138.04	83.01	92.64	106.22	104.89	105.73	59.91	69.00	46.09	59.35	78.14	106.39
High	126.70	65.07	120.24	92.65	97.07	89.79	32.62	64.12	86.91	38.77	53.48	50.64

Table B.3: Summary statistics for factor returns using news-based sentiment

This table shows the summary statistics for monthly factor returns in bps using the 2x3 and 2x2x2x2x2 sorting approach for January 1998 - December 2017. $MktRF$ is the value-weighted return on the market portfolio of all sample stocks minus the one-month Treasury bill rate. For the five Fama French factors SMB, HML, RMW, and CMA we follow their exact methodologly. At the end of each June, stocks are assigned to two $Size$ groups using the NYSE median market cap as the breakpoint. Stocks are also assigned independently to two book-to-market equity (B/M), operating profitability (OP), and investment (Inv) groups, using NYSE medians of B/M, OP, and Inv. Sentiment is constructed as the the news-based sentiment change compared to the rolling mean. As such, this sort is more frequently constructed than those on fundamental factors, given the higher frequency of changes in sentiment. Panel A of the table shows average monthly returns (Mean), the standard deviations of monthly returns (Std dev.) and the t-statistics for the average returns. Panel B shows the correlations between factors and Panel C the corresponding **p**-values. Panel D demonstrates the correlation between the different versions of the same factor from the 2x3 and 2x2x2x2x2x2 sorts.

	2x3						2x2x2x2x2					
	$Mkt-Rf$	SMB	HML	RMW	CMA	PMN	$Mkt-Rf$	SMB	HML	RMW	CMA	PMN
Panel A: Average, standard deviation and one-sample t-statistics for monthly returns												
Mean	55.16	20.27	14.39	33.27	19.38	3.80	55.16	42.84	-9.63	4.98	7.69	6.69
Std. dev	446.16	344.94	338.16	298.61	185.66	216.08	446.16	304.79	278.18	219.74	151.18	163.52
t-Statistic	1.92	0.91	0.66	1.73	1.62	0.27	1.92	2.18	-0.54	0.35	0.79	0.63
Panel B: Correlation between different factors												
$Mkt-Rf$	1.00	0.29	-0.23	-0.47	-0.28	-0.25	1.00	0.23	-0.24	-0.34	-0.23	-0.18
SMB		1.00	-0.34	-0.62	-0.03	0.06		1.00	-0.16	-0.33	0.11	-0.03
HML			1.00	0.58	0.56	-0.11			1.00	0.71	0.24	-0.16
RMW				1.00	0.15	0.00				1.00	0.06	-0.12
CMA					1.00	-0.01					1.00	-0.11
PMN						1.00						1.00
Panel C: P-value of correlations between different factors in Panel B												
$Mkt-Rf$		0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00
SMB			0.00	0.00	0.67	0.36			0.01	0.00	0.09	0.60
HML				0.00	0.00	0.09				0.00	0.00	0.01
RMW					0.02	0.99					0.37	0.06
CMA						0.88						0.10
PMN												
Panel D: Correlation between different versions of the same factor												
2x3							1.00	0.90	0.92	0.81	0.72	0.66

Table B.4: Summary statistics for regressions using news-based sentiment

This table shows the summary statistics for tests of the sentiment-augmented model benchmarked against the FF5 for January 1998 - December 2017. We test the ability of the models to explain monthly excess returns on 25 $Size - B/M$ portfolios (Panel A), 25 $Size - OP$ portfolios (Panel B), 25 $Size - Inv$ portfolios (Panel C), 25 $Size - Snt$ portfolios (Panel D), 36 $Size - B/M - OP - SNT$ Portfolios (Panel E), 36 $Size - B/M - Inv - SNT$ Portfolios (Panel F), 10 Fama-French Industry Portfolios (Panel G), and 30 Fama-French Industry Portfolios. We show the average absolute value of the intercepts, the GRS statistic testing whether the expected values of all intercept estimates are zero, $A|\alpha_i|$, the median absolute value of the intercept over the mean absolute return on portfolio i minus the mean of the cross-sectional portfolio returns $\frac{M|\alpha_i|}{A|\bar{r}_i|}$, and the squared version of the previous ratio $\frac{M|\alpha_i|^2}{A|\bar{r}_i^2|}$ to account for biases due to measurement errors.

	2x3				2x2x2x2x2			
	$A\|\alpha_i\|$	GRS	$\frac{A\|\alpha_i\|}{\bar{r}_i}$	$\frac{A\|\alpha_i\|^2}{\bar{r}_i^2}$	$A\|\alpha_i\|$	GRS	$\frac{A\|\alpha_i\|}{\bar{r}_i}$	$\frac{A\|\alpha_i^2\|}{\bar{r}_i^2}$
Panel A: Size-B/M portfolios								
FF5	10.43	1.96	0.65	0.26	14.58	2.13	0.86	0.44
FF5-SNT	10.72	1.98	0.75	0.34	14.90	2.15	1.03	0.64
Panel B: Size-OP portfolios								
FF5	7.20	1.10	0.39	0.10	15.35	1.54	0.71	0.33
FF5-SNT	6.74	1.07	0.37	0.09	14.87	1.51	0.79	0.41
Panel C: Size-Inv portfolios								
FF5	10.91	1.69	0.54	0.20	14.35	2.21	0.89	0.55
FF5-SNT	10.54	1.65	0.52	0.19	13.87	2.15	0.87	0.53
Panel D: Size-SNT portfolios								
FF5	31.32	2.19	0.68	0.30	26.27	2.45	0.68	0.30
FF5-SNT	30.20	2.16	0.56	0.20	26.21	2.40	0.60	0.23
Panel E: Size-B/M-OP-SNT Portfolios								
FF5	24.51	1.24	0.67	0.26	24.14	1.22	0.89	0.45
FF5-SNT	23.21	1.23	0.64	0.23	24.47	1.18	0.89	0.45
Panel F: Size-B/M-Inv-SNT Portfolios								
FF5	20.71	1.63	0.57	0.21	21.21	1.47	0.70	0.31
FF5-SNT	20.56	1.60	0.68	0.29	21.45	1.42	0.77	0.38
Panel G: 10 Industry Portfolios								
FF5	24.25	4.58	2.27	3.79	32.53	6.35	2.77	4.97
FF5-SNT	23.11	4.74	2.02	2.94	31.56	6.62	2.73	4.84
Panel H: 30 Industry Portfolios								
FF5	23.20	1.66	0.95	0.54	33.89	2.19	1.44	1.25
FF5-SNT	22.47	1.79	1.04	0.65	32.91	2.30	1.51	1.37

Table B.5: Summary statistics for regressions using orthogonalized social media-based sentiment

This table shows the summary statistics for tests of the sentiment-augmented model benchmarked against the FF5 for January 1998 - December 2017. We test the ability of the models to explain monthly excess returns on 25 $Size-B/M$ portfolios (Panel A), 25 $Size-OP$ portfolios (Panel B), 25 $Size-Inv$ portfolios (Panel C), 25 $Size-Snt$ portfolios (Panel D), 36 $Size-B/M-OP-SNT$ Portfolios (Panel E), 36 $Size-B/M-Inv-SNT$ Portfolios (Panel F), 10 Fama-French Industry Portfolios (Panel G), and 30 Fama-French Industry Portfolios. Social media-based sentiment is orthogonalized against the news-based sentiment. We show the average absolute value of the intercepts, the GRS statistic testing whether the expected values of all intercept estimates are zero, $A|\alpha_i|$, the median absolute value of the intercept over the mean absolute return on portfolio i minus the mean of the cross-sectional portfolio returns $\frac{M|\alpha_i|}{A|\bar{r}_i|}$, and the squared version of the previous ratio $\frac{M|\alpha_i|^2}{A|\bar{r}_i^2|}$ to account for biases due to measurement errors.

	2x3				2x2x2x2x2			
	$A\|\alpha_i\|$	GRS	$\frac{A\|\alpha_i\|}{A\|\bar{r}_i\|}$	$\frac{A\|\alpha_i\|^2}{\bar{r}_i^2}$	$A\|\alpha_i\|$	GRS	$\frac{A\|\alpha_i\|}{\bar{r}_i}$	$\frac{A\|\alpha_i^2\|}{A\|\bar{r}_i^2\|}$
Panel A: Size-B/M portfolios								
FF5	10.43	1.96	0.65	0.26	15.33	2.34	1.05	0.66
FF5-SNT	10.30	1.83	0.80	0.39	15.73	2.11	1.00	0.61
Panel B: Size-OP portfolios								
FF5	7.20	1.10	0.39	0.10	15.26	1.68	0.82	0.44
FF5-SNT	6.75	1.09	0.27	0.05	14.55	1.53	0.69	0.31
Panel C: Size-Inv portfolios								
FF5	10.91	1.69	0.54	0.20	15.07	2.26	0.80	0.45
FF5-SNT	9.43	1.47	0.44	0.14	14.44	1.89	0.74	0.38
Panel D: Size-SNT portfolios								
FF5	35.70	2.72	0.52	0.16	26.74	2.86	0.49	0.14
FF5-SNT	30.96	2.45	0.71	0.29	22.46	2.45	0.47	0.13
Panel E: Size-B/M-OP-SNT Portfolios								
FF5	25.63	1.81	0.59	0.24	24.50	1.71	0.85	0.50
FF5-SNT	23.88	1.65	0.62	0.26	23.54	1.47	0.74	0.37
Panel F: Size-B/M-Inv-SNT Portfolios								
FF5	24.92	1.47	0.64	0.27	23.55	1.35	0.64	0.27
FF5-SNT	21.41	1.30	0.48	0.15	21.33	1.10	0.69	0.31
Panel G: 10 Industry Portfolios								
FF5	24.25	4.58	2.27	3.79	32.39	6.28	2.77	4.98
FF5-SNT	23.69	4.91	2.20	3.63	32.16	6.67	2.82	5.15
Panel H: 30 Industry Portfolios								
FF5	23.20	1.66	0.95	0.54	33.20	2.14	1.50	1.35
FF5-SNT	22.73	1.89	1.01	0.61	34.90	2.49	1.62	1.58

C Sentiment, momentum, and liquidity

Table C.1: Summary statistics for factor returns including momentum using social media-based sentiment

This table shows the summary statistics for monthly factor returns in bps using the 2x3 and 2x2x2x2x2 sorting approach for January 1998 - December 2017. $MktRF$ is the value-weighted return on the market portfolio of all sample stocks minus the one-month Treasury bill rate. For the five Fama French factors SMB, HML, RMW, and CMA we follow their exact methodology. At the end of each June, stocks are assigned to two $Size$ groups using the NYSE median market cap as the breakpoint. Stocks are also assigned independently to two book-to-market equity (B/M), operating profitability (OP), and investment (Inv) groups, using NYSE medians of B/M, OP, and Inv. Momentum (UMD) are two groups of ood performers and bad performers, monthly measured as prior (2-12) cumulative returns. Sentiment is constructed as the social media-based sentiment change compared to the rolling mean. As such, this sort is more frequently constructed than those on fundamental factors, given the higher frequency of changes in sentiment. Panel A of the table shows average monthly returns (Mean), the standard deviations of monthly returns (Std dev.) and the t-statistics for the average returns. Panel B shows the correlations between factors and Panel C the corresponding **p**-values. Panel D demonstrates the correlation between the different versions of the same factor from the 2x3 and 2x2x2x2x2x2 sorts.

	2x3							2x2x2x2x2						
	$Mkt-Rf$	SMB	HML	RMW	CMA	UMD	PMN	$Mkt-Rf$	SMB	HML	RMW	CMA	UMD	PMN
Panel A: Average, standard deviation and one-sample t-statistics for monthly returns														
Mean	55.16	20.27	14.39	33.27	19.38	38.02	22.28	55.16	32.15	4.53	11.21	5.06	11.24	22.48
Std. dev	446.16	344.94	338.16	298.61	185.66	526.72	166.49	446.16	280.56	235.30	180.86	128.92	325.62	114.97
t-Statistic	1.92	0.91	0.66	1.73	1.62	1.12	2.07	1.92	1.78	0.30	0.96	0.61	0.53	3.03
Panel B: Correlation between different factors														
$Mkt-Rf$	1.00	0.29	-0.23	-0.47	-0.28	-0.31	-0.29	1.00	0.19	-0.22	-0.35	-0.27	-0.30	-0.23
SMB		1.00	-0.34	-0.62	-0.03	0.06	0.11		1.00	-0.16	-0.35	0.01	0.04	0.13
HML			1.00	0.58	0.56	-0.18	-0.19			1.00	0.75	0.29	-0.22	-0.21
RMW				1.00	0.15	0.04	0.02				1.00	0.18	-0.17	-0.16
CMA					1.00	0.03	-0.03					1.00	-0.04	-0.06
UMD						1.00	0.56						1.00	0.48
PMN							1.00							1.00
Panel C: P-value of correlations between different factors in Panel B														
$Mkt-Rf$		0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.00	0.00	0.00
SMB			0.00	0.00	0.67	0.34	0.08			0.01	0.00	0.88	0.54	0.04
HML				0.00	0.00	0.01	0.00				0.00	0.00	0.00	0.00
RMW					0.02	0.57	0.71					0.01	0.01	0.01
CMA						0.65	0.68						0.50	0.39
UMD							0.00							0.00
PMN														
Panel D: Correlation between different versions of the same factor														
2x3								1.00	0.94	0.94	0.84	0.77	0.96	0.69

Table C.2: Summary statistics for factor returns including liquidity using social media-based sentiment

This table shows the summary statistics for monthly factor returns in bps using the 2x3 and 2x2x2x2x2 sorting approach for January 1998 - December 2017. $MktRF$ is the value-weighted return on the market portfolio of all sample stocks minus the one-month Treasury bill rate. For the five Fama French factors SMB, HML, RMW, and CMA we follow their exact methodology. At the end of each June, stocks are assigned to two $Size$ groups using the NYSE median market cap as the breakpoint. Stocks are also assigned independently to two book-to-market equity (B/M), operating profitability (OP), and investment (Inv) groups, using NYSE medians of B/M, OP, and Inv. Liquidity (LMI) are two groups of liquid and illiquid stocks, measured as monthly traded volume in relation to shares outstanding. Sentiment is constructed as the social media-based sentiment change compared to the rolling mean. As such, this sort is more frequently constructed than those on fundamental factors, given the higher frequency of changes in sentiment. Panel A of the table shows average monthly returns (Mean), the standard deviations of monthly returns (Std dev.) and the t-statistics for the average returns. Panel B shows the correlations between factors and Panel C the corresponding **p**-values. Panel D demonstrates the correlation between the different versions of the same factor from the 2x3 and 2x2x2x2x2 sorts.

	2x3							2x2x2x2x2						
	$Mkt-Rf$	SMB	HML	RMW	CMA	LMI	PMN	$Mkt-Rf$	SMB	HML	RMW	CMA	LMI	PMN
Panel A: Average, standard deviation and one-sample t-statistics for monthly returns														
Mean	55.16	20.27	14.39	33.27	19.38	31.05	22.28	55.16	26.74	1.17	16.51	1.58	22.07	22.36
Std. dev	446.16	344.94	338.16	298.61	185.66	493.71	166.49	446.16	279.36	153.95	158.87	119.43	207.93	113.51
t-Statistic	1.92	0.91	0.66	1.73	1.62	0.97	2.07	1.92	1.48	0.12	1.61	0.20	1.64	3.05
Panel B: Correlation between different factors														
$Mkt-Rf$	1.00	0.29	-0.23	-0.47	-0.28	0.67	-0.29	1.00	0.21	-0.39	-0.33	-0.15	0.66	-0.19
SMB		1.00	-0.34	-0.62	-0.03	0.66	0.11		1.00	-0.11	-0.23	0.13	0.41	0.05
HML			1.00	0.58	0.56	-0.55	-0.19			1.00	0.73	0.23	-0.65	-0.22
RMW				1.00	0.15	-0.78	0.02				1.00	0.00	-0.52	-0.15
CMA					1.00	-0.33	-0.03					1.00	-0.14	-0.16
LMI						1.00	-0.03						1.00	0.02
PMN							1.00							1.00
Panel C: P-value of correlations between different factors in Panel B														
$Mkt-Rf$		0.00	0.00	0.00	0.00	0.00	0.00		0.00	0.00	0.00	0.02	0.00	0.00
SMB			0.00	0.00	0.67	0.00	0.08			0.08	0.00	0.05	0.00	0.44
HML				0.00	0.00	0.00	0.00				0.00	0.00	0.00	0.00
RMW					0.02	0.00	0.71					0.95	0.00	0.02
CMA						0.00	0.68						0.03	0.02
LMI							0.68							0.74
PMN														
Panel D: Correlation between different versions of the same factor														
2x3								1.00	0.93	0.90	0.81	0.77	0.94	0.69

Essay III: Diversifying estimation errors: An efficient averaging rule for portfolio optimization

Diversifying estimation errors: An efficient averaging rule for portfolio optimization

Roland Füss* Christian Koeppel† Felix Miebs‡

Abstract

We propose an averaging rule that combines established minimum-variance strategies to minimize the expected out-of-sample variance. Our rule overcomes the problem of selecting the "best" strategy ex-ante and diversifies remaining estimation errors of the single strategies included in the averaging. Extensive simulations show that the contributions of estimation errors to the out-of-sample variances are uncorrelated between the considered strategies. This implies that averaging over multiple strategies offers sizable diversification benefits. Our rule leverages these benefits and compares favorably to eleven strategies in terms of out-of-sample variance on both simulated and empirical data sets. The Sharpe ratio is across all data sets at least 25% higher than for the 1/N portfolio.

JEL Classification: G11

Keywords: Averaging; diversification; estimation error; portfolio optimization; shrinkage.

*Swiss Institute of Banking and Finance (s/bf), University of St.Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland, Research Associate at the Centre for European Economic Research (ZEW), Mannheim, Germany and a Visiting Professor and Fellow at the Center for Real Estate and Environmental Economics at NTNU Business School, Trondheim, Norway; Phone: +41 (0)71 224-7055, Fax: +41 (0)71 224-7088; Email: roland.fuess@unisg.ch

†University of St.Gallen, Unterer Graben 21, 9000 St. Gallen, Switzerland; Phone: +41 (0)76 38 30 757, Email: christian.koeppel@live.com

‡University of Applied Sciences Cologne, Claudiusstr. 1, 50678 Cologne, Germany; Phone: +49 (0)221 8275 3792; Email: felix.miebs@th-koeln.de.

1 Introduction

The literature on portfolio selection offers a broad range of approaches to alleviate the impact of estimation errors on out-of-sample portfolio performance. Nevertheless, DeMiguel, Garlappi, and Uppal (2009) find that no single portfolio strategy consistently outperforms the 1/N rule. Stimulated by this observation, several authors suggested improved estimates of the input parameters on the covariance (see, e.g., Fan, Liao, & Mincheva, 2013; Ledoit & Wolf, 2017), the inverse covariance (see, e.g., DeMiguel, Martin-Utrera, & Nogales, 2013; Kourtis, Dotsis, & Markellos, 2012; F. Shi, Shu, Yang, & He, 2020) and the portfolio weight level (see, e.g., DeMiguel et al., 2013; Kan & Zhou, 2007; Tu & Zhou, 2011). Further contributions propose extensions of the portfolio optimization problem via norm constraints in order to regularize it (see, e.g., Brodie, Daubechies, De Mol, Giannone, & Loris, 2009; DeMiguel, Garlappi, & Uppal, 2009; Li, 2015; Xing, Hu, & Yang, 2014; Yen, 2016). The plethora of evolving strategies challenges practitioners and academics alike to choose one of the proposed strategies. In addition to the challenge of identifying the "best" strategy ex-ante, the limitation to a single strategy may forego potential diversification benefits. If estimation errors are not perfectly correlated across individual strategies, a combination of different approaches offers the potential to further diversify estimation errors. The resulting portfolio weights are less prone to estimation errors and achieve, on average, an improved out-of-sample performance.

In this paper, we propose an averaging rule that overcomes the strategy selection problem and exploits the potential to diversify estimation errors between individual strategies. Our approach takes the principle of diversification from the portfolio to the estimation level. Following the consensus in the literature, we apply our averaging rule to minimum-variance portfolios because expected returns are notoriously difficult to estimate (see, e.g., Merton, 1980).

Our rule aims to minimize the expected out-of-sample portfolio variance. For this purpose and to reduce the impact of estimation errors, we combine established approaches on three different levels: i) the covariance level, where averaging is applied

before one computes the inverse of the covariance matrix, ii) the inverse covariance level, where averaging is applied on the individual inverses of the various covariance matrices, and iii) the portfolio weight level, where averaging is applied directly on the portfolio weights. Our averaging rule combines the unbiased sample estimator with a multitude of structured estimators.[4] Thus, our approach extends the idea of shrinkage from a single to a multi-target specification.

In extensive simulations, we find that estimation errors between a variety of established minimum-variance strategies are uncorrelated which is undocumented and unexploited so far. This correlation structure offers sizable and persistent diversification benefits. Our averaging rule utilizes this structure to diversify estimation errors, leading to a decrease in the out-of-sample variance as the number of strategies increases. The average out-of-sample variance of our rule over all simulations is lower than the variance of any single strategy. Our averaging rule also achieves higher Sharpe ratios for smaller estimation windows. In addition, we find that our rule compares favorably to the selected benchmarks, not only in terms of lower out-of-sample variance, but also higher out-of-sample Sharpe ratio. This includes the most recent non-linear shrinkage approach of Ledoit and Wolf (2017).

Empirical results on five data sets confirm the findings from the simulation study that our rule achieves a lower out-of-sample variance than any single strategy. The variance reduction in comparison to the non-linear shrinkage strategy of Ledoit and Wolf (2017) is statistically significant on three data sets. Our rule achieves the second highest Sharpe ratio on all data sets in comparison to the single strategies. Relative to the 1/N strategy, the out-of-sample Sharpe ratio of our rule is on all data sets at least 25% higher, being statistically significant in four out of five sets. The turnover and short interest of our averaging rule is modest. Averaging consistently outperforms the single strategies and benchmarks in terms of risk reduction and performance improvement, as well as reduction in turnover and short interest. These properties make our averaging rule both statistically

[4]Throughout the paper, we average over three different levels (covariance, inverse, and weight level), use the Sample and the structured (target) estimators (jointly denoted as strategies) for averaging on the aforementioned levels, and benchmark our rule against existing shrinkage approaches (denoted as benchmarks).

and economically appealing.

Our paper relates to prior contributions proposing combinations of portfolio weights or input parameters to reduce estimation errors. Tu and Zhou (2011) show that combining portfolio weights of established mean-variance strategies with the 1/N rule improves the out-of-sample performance of the respective strategies. Kourtis et al. (2012) find that a linear combination of the sample and two structured estimators on the inverse covariance level improves the out-of-sample performance of minimum-variance portfolios. The results of the aforementioned articles suggest that estimation errors may further be reduced, if just one additional resource for reducing errors is included. We follow this path, exploring whether estimation errors can be diversified more comprehensively if multiple resources for the diversification are taken into account.

We contribute to the existing literature in two ways. First, in a battery of simulations, we find that estimation errors between a variety of established minimum-variance strategies are uncorrelated which is undocumented and unexploited so far. Our findings show that the resulting diversification benefits are sizable and persistent for different estimation windows. Second, we provide a framework for combining multiple strategies to alleviate the impact of estimation errors. Our averaging rule builds on the principle of diversification, is easy to calibrate, and extends to an arbitrary number of strategies, not limited to the ones selected for this paper. On simulated and empirical data our rule compares favorably to selected benchmarks in terms of out-of-sample variance and Sharpe ratio. Thus, averaging over multiple single strategies provides a promising avenue to reduce the impact of estimation errors on out-of-sample portfolio performance.

The paper is organized as follows. Section 2 reviews the considered minimum-variance strategies and introduces our averaging rule. Section 3 describes the simulation procedure and explores the behavior of the averaging rule using simulated data. Section 4 evaluates the out-of-sample performance of our averaging rule on five empirical data sets. Section 5 concludes.

2 Portfolio optimization and estimation errors

2.1 Minimum-variance portfolio optimization

Throughout the paper we focus on the estimation of the minimum-variance portfolio resulting from the following optimization problem:

$$\begin{aligned} \min_{w} \quad & w'\Sigma w \\ \text{s.t.} \quad & w'\mathbf{1}_N = 1, \end{aligned} \tag{12}$$

where Σ is the true but unknown covariance matrix, $\mathbf{1}_N$ is a N-dimensional vector of ones, and N is the number of investable risky assets, respectively. The allocation to the N assets within the minimum-variance portfolio results from solving the problem in Equation (12), with the optimal portfolio weights given by:

$$w = \Sigma^{-1}\mathbf{1}_N(\mathbf{1}'_N\Sigma^{-1}\mathbf{1}_N)^{-1}. \tag{13}$$

Because w depends on the true but unknown covariance matrix, Σ has to be replaced by a suitable estimator, $\hat{\Sigma}$. A natural estimator for the covariance matrix is the sample covariance matrix. Nevertheless, the estimation error associated with the sample covariance matrix can be substantial (see, e.g., DeMiguel, Garlappi, & Uppal, 2009; Ledoit & Wolf, 2003). Any estimation errors in $\hat{\Sigma}$ or its inverse, translate into sub-optimal allocations with an adverse effect on the portfolio performance.

2.2 Alleviating estimation errors: Existing approaches

The literature proposes different approaches to mitigate the impact of estimation errors on minimum-variance portfolio weights and out-of-sample performance. The first branch of literature suggests structured estimators of the covariance matrix. Examples include the constant correlation model (see, e.g., Elton & Gruber, 1973) and (approximate) factor models (see, e.g., Chan, Karceski, & Lakonishok, 1999; Fan et al., 2013). The second

branch proposes shrinkage estimators of the covariance matrix (see, e.g., Ledoit & Wolf, 2003, 2004a, 2004b, 2017), its inverse (see, e.g., DeMiguel et al., 2013; Kourtis et al., 2012; F. Shi et al., 2020), and portfolio weights (see, e.g., DeMiguel et al., 2013; Kan & Zhou, 2007; Tu & Zhou, 2011). The third branch regularizes portfolio weights through the imposition of weight constraints. Frost and Savarino (1986) as well as Jagannathan and Ma (2003), study the effects of short-sale constraint. Brodie et al. (2009), DeMiguel, Garlappi, and Uppal (2009), Xing et al. (2014), Li (2015), and Yen (2016) suggest the imposition of norm constraints on portfolio weights.[5]

Throughout the paper we report results for eleven minimum-variance strategies, covering at least one representative of each aforementioned branch from the existing literature. Each of these can be represented in terms of a covariance matrix estimator. The portfolio weights of each strategy are given by Equation (13), with $\hat{\Sigma}$ being one of the following estimators, clustered into two groups.[6] The first group comprises the sample as well as structured estimates of the covariance matrix and represents the building blocks for our averaging rule:

- Sample: The (unbiased) sample covariance matrix, $\hat{\Sigma}_S$.
- 1F: The covariance matrix implied by the single factor model of Sharpe (1967), $\hat{\Sigma}_{1F}$.
- CC: The covariance matrix recovered from a constant correlation matrix with the average of all pairwise correlations off the main diagonal and the assets' individual standard deviations, $\hat{\Sigma}_{CC}$.
- ID: A diagonal matrix with the average variance of all assets on the main diagonal, $\hat{\Sigma}_{ID}$. The resulting minimum-variance portfolio corresponds to the 1/N portfolio, which represents a benchmark in terms of risk-adjusted performance (DeMiguel, Garlappi, & Uppal, 2009).

[5] We do not include norm-constrained strategies in our paper because the moment-shrinkage representation of norm constraints does not hold for arbitrary levels of the constraint. Thus, it is not possible to use the aforementioned approaches on the three considered averaging levels.

[6] We provide a comprehensive review of the strategies in section A of the Internet Appendix.

- SC: The covariance matrix corresponding to the short-sale constrained minimum-variance portfolio from Jagannathan and Ma (2003), implying non-negative portfolio weights, $\hat{\Sigma}_{SC}$.

- POET: The covariance matrix based on the approximate factor model by Fan et al. (2013) with risk factors defined by the principal components of the sample covariance matrix, $\hat{\Sigma}_{POET}$.

The second group of estimators covers combinations of the sample and another covariance estimator serving as benchmarks for our averaging rule:

- LW1F: A linear combination of $\hat{\Sigma}_S$ and $\hat{\Sigma}_{1F}$. The combination intensity is determined by the linear shrinkage estimator of Ledoit and Wolf (2003).

- LWID: A linear combination of $\hat{\Sigma}_S$ and $\hat{\Sigma}_{ID}$. The combination intensity is determined by the linear shrinkage estimator of Ledoit and Wolf (2004b).

- LWCC: A linear combination of $\hat{\Sigma}_S$ and $\hat{\Sigma}_{CC}$. The combination intensity is determined by the linear shrinkage estimator of Ledoit and Wolf (2004a).

- LWNLS: A non-linear shrinkage estimator, correcting for over-dispersed eigenvalues of the sample covariance matrix following Ledoit and Wolf (2017). This approach represents a benchmark in terms of out-of-sample variance.

- KDM: A linear combination of $\hat{\Sigma}_S$, $\hat{\Sigma}_{1F}$ and $\hat{\Sigma}_{CC}$. The combination intensity is determined following Kourtis et al. (2012).

All approaches aim to determine the single best strategy to alleviate the impact of estimation errors. The wide variety of approaches leaves investors with the question of which suggested strategy to follow. Without any ex-ante knowledge about the "best" approach, investors face the problem of picking a single strategy. In contrast to choosing one single strategy, we advocate that investors should rather diversify over various strategies to mitigate estimation errors.

2.3 Estimation error diversification: An averaging approach

Our approach of estimation error diversification generalizes the idea of shrinkage. We suggest blending the sample with a multitude of structured estimators. If estimation errors can be reduced by combining the sample with one structured estimator, averaging over multiple estimators should offer further diversification potential. In essence, our approach brings the principle of diversification from the portfolio to the estimation level.

We suggest an averaging rule that combines individual estimators with the objective of minimizing the expected out-of-sample variance of the corresponding minimum-variance portfolio.[7] Because there is no consensus in the literature whether estimation errors can most efficiently be diversified on the covariance, inverse or portfolio weight level, we consider averaging over estimators on the three aforementioned levels. We assign the averaging weights to the considered estimators on the respective level such that the expected out-of-sample variance of the corresponding minimum-variance portfolio is minimized:

$$\min_{\lambda_{Level}} \quad E\left(\hat{w}_{Level}\left(\lambda_{Level}\right)' \Sigma \hat{w}_{Level}\left(\lambda_{Level}\right)\right) \tag{14}$$

$$\text{s.t.} \quad \lambda_{Level}' \mathbf{1}_M = 1 \tag{15}$$

$$\lambda_{Level} \geq 0, \tag{16}$$

where λ_{Level} are the averaging weights on the respective averaging level, $\hat{w}_{Level}$ are the corresponding portfolio weight estimates, $\mathbf{1}_M$ is a M-dimensional vector of ones, and M is the number of estimators. Equations (14) - (16) correspond to a minimum-variance portfolio optimization problem with non-negativity restrictions on the averaging weights. We impose the non-negativity constraint because Jagannathan and Ma (2003) show that it has a shrinkage-like effect and improves the portfolio performance. The literature on linear regression model averaging also finds that non-negativity constraints enhance the

[7] We do not claim that other, potentially more sophisticated averaging rules may deliver improved results. We rather advocate in a comparatively simple setting that averaging over individual established approaches provides a promising avenue to alleviate the impact of estimation errors on portfolio performance.

quality of the combination (see, e.g., Hansen, 2008; Timmermann, 2006).[8]

If averaging is conducted on the covariance or inverse level, we combine the M estimators of the covariance matrix or its inverse in a first step and compute the corresponding minimum-variance portfolio weights in a second step. Following the objective in Equation (14), we combine the individual estimators such that the expected out-of-sample variance of the corresponding minimum-variance portfolio is minimized. We approximate the expected out-of-sample variance based on jackknifed returns, considering a one-month holding period throughout the paper (see, e.g., DeMiguel, Garlappi, Nogales, & Uppal, 2009; Li, 2015; Xing et al., 2014; Yen, 2016). We drop the excess returns falling into the i-th month from our in-sample period and compute the covariance or its inverse, respectively, using the remaining excess returns within the in-sample period. We denote the m-th estimator of the covariance or its inverse, by $\hat{\Sigma}_{-i,m}$ and $\hat{\Sigma}^{-1}_{-i,m}$, respectively. The minimum-variance portfolio weights corresponding to the combination of the M estimators of the covariance matrix, $\hat{w}_{-i,AV-Cov}$, and its inverse, $\hat{w}_{-i,AV-Inv}$, are given by:

$$\hat{w}_{-i,AV-Cov} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_{-i,m}\right)^{-1} \mathbf{1}_N \left(\mathbf{1}'_N \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_{-i,m}\right)^{-1} \mathbf{1}_N\right)^{-1},$$

$$\hat{w}_{-i,AV-Inv} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}^{-1}_{-i,m}\right) \mathbf{1}_N \left(\mathbf{1}'_N \sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}^{-1}_{-i,m} \mathbf{1}_N\right)^{-1},$$

where $\hat{\lambda}_{Cov}$ and $\hat{\lambda}_{Inv}$ are our estimates of the averaging weights. We then compute the jackknife excess returns for the i-th month, $r^{JK}_{i,Cov}$ and $r^{JK}_{i,Inv}$, representing the excess returns that would have been achieved from holding $\hat{w}_{-i,Cov}$ and $\hat{w}_{-i,Inv}$, respectively, in the hold-out month i. Repeating these steps for all months within our in-sample period yields the time series of jackknifed portfolio excess returns. We use the sample variance of this time series as an estimate of the objective in Equation (14). The portfolio weights

[8]Britten-Jones (1999) shows that the mean-variance portfolio optimization problem can be formulated as a linear regression problem.

resulting from this averaging procedure are given by:

$$\hat{w}_{AV-Cov} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_m\right)^{-1} \mathbf{1}_N \left(\mathbf{1}_N' \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Cov} \cdot \hat{\Sigma}_m\right)^{-1} \mathbf{1}_N\right)^{-1}, \tag{17}$$

$$\hat{w}_{AV-Inv} = \left(\sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}_m^{-1}\right) \mathbf{1}_N \left(\mathbf{1}_N' \sum_{m=1}^{M} \hat{\lambda}_{m,Inv} \cdot \hat{\Sigma}_m^{-1} \mathbf{1}_N\right)^{-1}, \tag{18}$$

where $\hat{\Sigma}_m$ and $\hat{\Sigma}_m^{-1}$ denote the m-th estimator of the covariance or its inverse, based on all in-sample observations, respectively.

If averaging is conducted on the portfolio weight level, we combine the M estimators of the minimum-variance portfolio weights such that the expected out-of-sample variance in Equation (14) is minimized. We follow the outlined jackknife procedure and drop the excess returns falling into the i-th month. Thereafter, we compute the minimum-variance portfolio weights $\hat{w}_{-i,m}$ for each of the M estimators. The minimum-variance portfolio weights of the combination, $\hat{w}_{-i,AV-Wgt}$, in the hold-out month i are given by:

$$\hat{w}_{-i,AV-Wgt} = \sum_{m=1}^{M} \hat{\lambda}_{m,Wgt} \cdot \hat{w}_{-i,m},$$

where $\hat{\lambda}_{Wgt}$ are the estimated averaging weights and $r_{i,AV-Wgt}^{JK}$ is the jackknife return for the i-th month. Repeating the procedure for all months and computing the sample variance of the time series of jackknife excess returns again gives the estimate of the expected out-of-sample variance in Equation (14). The portfolio weights corresponding to averaging on the weight level are given by:

$$\hat{w}_{AV-Wgt} = \sum_{m=1}^{M} \hat{\lambda}_{m,Wgt} \cdot \hat{w}_m. \tag{19}$$

Table 3 provides an overview of our proposed averaging rule on the three levels in Panel A and the single strategies in Panel B. Panel C lists the five selected benchmarks of which LWNLS is the latest and most sophisticated, and thus our target to beat in terms of out-of-sample variance.

Throughout the paper we consider averaging over the sample and the 1F, CC, ID, SC, and POET estimates on the three aforementioned levels. The selected set of estimators is, as any other set, an ad-hoc selection for which the motivation is twofold. First, we seek to build our averaging rule on an established set of simple estimators. Second, we want to include a sufficiently large number of estimators to explore the trade-off between a higher potential to diversify estimation errors and the newly arising estimation problem. Including more estimators in the averaging rule offers a greater potential to diversify estimation errors. Yet, including more strategies requires the estimation of more averaging weights, giving rise to a new source of estimation errors. We explore this trade-off in the following simulation study.

3 Simulation study

3.1 Simulation procedure

The data generating process of our simulation study follows Tu and Zhou (2011). We consider $N = 25$ assets with the simulated excess return of security j at time t, $r_{j,t}$, coming from the Fama-French three-factor model with mispricing:

$$r_{j,t} = \alpha_j + \beta_{j,MKT} r_{t,MKT} + \beta_{j,SMB} r_{t,SMB} + \beta_{j,HML} r_{t,HML} + u_{j,t}, \tag{20}$$

where $\beta_{j,MKT}$, $\beta_{j,SMB}$, and $\beta_{j,HML}$ are the factor loadings, α_j is the mispricing factor, and $u_{j,t}$ is the residual at time t. The returns of the market $r_{t,MKT}$, size $r_{t,SMB}$, and value factor $r_{t,HML}$ at t follow a multivariate normal distribution. Table 4 shows the means and standard deviations of the factors as well as the correlation matrix between the factors.

The assets' factor loadings are randomly paired and evenly spread between 0.9 and 1.2 for β_{MKT}, -0.3 and 1.4 for β_{SMB}, -0.5 and 0.9 for β_{HML}, and -2.0 and 2.0% for the annualized mispricing factor α, in each simulation. The error term of each asset at t, $u_{j,t}$, comes from a multivariate normal distribution with $N \sim (0, \Sigma_u)$. Again, we follow Tu and Zhou (2011) and assume that the covariance matrix Σ_u is diagonal. The residuals'

Table 3: List of averaging rule, single strategies, and benchmarks

This table lists the proposed averaging rule on the considered levels in Panel A, the single strategies in Panel B, and the benchmarks, representing sophisticated estimates of the minimum-variance portfolio, in Panel C. The last column of the table shows the abbreviation that we use to refer to the respective averaging rule or strategy. We do not include the 1/N strategy from DeMiguel, Garlappi, and Uppal (2009) separately because ID corresponds to the 1/N strategy.

#	Description	Abbreviation
Panel A: Averaging rule across three levels		
1	Averaging rule on the portfolio weight level	AV-Wgt
2	Averaging rule on the inverse covariance level	AV-Inv
3	Averaging rule on the covariance level	AV-Cov
Panel B: Single strategies		
4	Minimum-variance portfolio without constraints based on the sample covariance matrix	Sample
5	Minimum-variance portfolio with the market as a single factor (Sharpe, 1963)	1F
6	Minimum-variance portfolio with the constant correlation model implied covariance matrix (Elton & Gruber, 1973)	CC
7	Minimum-variance portfolio with a scalar multiple of the identity covariance matrix	ID
8	Minimum-variance portfolio with short-sale constraints (Jagannathan & Ma, 2003)	SC
9	Minimum-variance portfolio based on the principal components of the sample covariance matrix (Fan et al., 2013)	POET
Panel C: Benchmarks		
10	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and the single factor covariance matrix (Ledoit & Wolf, 2003)	LW1F
11	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and a scalar multiple of the identity matrix (Ledoit & Wolf, 2004b)	LWID
12	Minimum-variance portfolio with the covariance matrix as weighted average between the sample covariance and a constant correlation covariance matrix (Ledoit & Wolf, 2004a)	LWCC
13	Minimum-variance portfolio with the non-linear shrinkage implied covariance matrix (Ledoit & Wolf, 2017)	LWNLS
14	Three-fund strategy (Kourtis et al., 2012)	KDM

Table 4: Summary statistics of simulation parameters

This table reports the monthly means, standard deviations, and cross-correlations between the three Fama-French factors market (MKT), size (SMB), and value (HML) in the simulation. The reported values correspond to the empirical values over the period from 06/1963 - 08/2007. The time horizon for the calibration of the parameters corresponds to Tu and Zhou (2011).

	MKT	SMB	HML
Mean	0.478	0.236	0.450
Standard Deviation	4.379	3.215	2.809
Cross-correlations			
MKT	1.000	0.287	-0.390
SMB		1.000	-0.260
HML			1.000

annualized volatility is drawn from a uniform distribution with a lower bound of 10% and an upper bound of 30%, such that the average idiosyncratic volatility in the cross-section equals 20%.

We run 10,000 simulations. In each run k we draw $\tau = \{60, 120, 240, 480, 960\}$ monthly observations from the outlined data generating process and save the population mean vector, μ_k, and the covariance matrix, Σ_k. These parameters result from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. We then estimate the portfolio weights of the m-th strategy, $\hat{w}_{k,m}$, based on the τ simulated excess returns and evaluate the estimated portfolio weights on the out-of-sample variance, $\hat{\sigma}_{k,m}$, and Sharpe ratio, $\hat{\Psi}_{k,m}$, of this simulation run:

$$\hat{\sigma}^2_{k,m} = \hat{w}'_{k,m}\Sigma_k\hat{w}_{k,m}, \tag{21}$$

$$\hat{\Psi}_{k,m} = \frac{\hat{w}'_{k,m}\mu_k}{\sqrt{\hat{w}_{k,m}\Sigma_k\hat{w}_{k,m}}}. \tag{22}$$

3.2 Simulation results

In a first step, we investigate the estimation errors of the individual strategies. We quantify the contribution of the estimation errors to the out-of-sample variances of the m-th strategy in the k-th simulation run as the following loss:

$$L_{k,m} = (\hat{w}_{k,m} - w_k)' \Sigma_k (\hat{w}_{k,m} - w_k), \tag{23}$$

where w_k are the true minimum-variance portfolio weights, which we compute using the population covariance matrix of the respective run, Σ_k. Table 5 reports the means, standard deviations, and correlation coefficients between the losses of the single strategies over the 10,000 simulation runs. The means of all strategies decrease for increasing estimation windows in Panels A - E, except for the ID strategy. As expected, this decrease is most pronounced for the unbiased Sample strategy. The ID strategy corresponds to the 1/N portfolio, such that the means and standard deviations are constant over the estimation windows. The standard deviations of the other strategies decrease with larger sample sizes, being highest for the Sample for $\tau = 60$ and smallest for $\tau = 960$. Most importantly, the correlations are small in absolute terms, offering sizable diversification opportunities. The average pairwise correlation reduces from an already low level of 0.173 to 0.046 going from $\tau = 60$ to $\tau = 960$. The minimum and maximum pairwise correlation between the considered strategies range from -0.345 to 0.745 and -0.392 to 0.394 for $\tau = 60$ and $\tau = 960$, respectively. We observe the highest correlation between the two (approximate) factor model strategies POET and 1F, highlighting the need for heterogeneous strategies to lever the diversification potential. In contrast to existing linear shrinkage approaches that combine the Sample with one or two other strategies (see, e.g., Kourtis et al., 2012; Ledoit & Wolf, 2003, 2004a, 2004b), our results suggest that combining a multitude of strategies offers sizable diversification benefits. Yet, averaging over multiple strategies gives rise to a new estimation problem, resulting in the following trade-off: on the one hand, increasing the number of strategies lowers the portfolio variance; on the other hand, it increases the number of averaging weights that need to be estimated,

creating a new source of estimation errors.

We evaluate this trade-off in Table 6 by applying our averaging rule to the sample covariance matrix and an increasing number of strategies from Panel B of Table 3. Importantly, we consider all possible permutations that result from combining the sample covariance matrix with the respective subset of strategies.[9] For each permutation in each simulation run we compute the averaging weights according to our averaging rule and obtain the resulting portfolio weights. We then evaluate the estimated portfolio weights based on their out-of-sample variance according to Equation (21).

Table 6 shows the average out-of-sample variances over all simulations and permutations for the corresponding number of strategies. Panels A-E present the results for the different estimation windows $\tau = \{60, 120, 240, 480, 960\}$, when averaging is conducted on the portfolio weight (AV-Wgt), the inverse (AV-Inv), or the covariance (AV-Cov) level. Our results show that the average out-of-sample variance decreases as the number of strategies increases.[10] This result holds irrespective of the estimation window and the averaging level, demonstrating that the diversification benefits outweigh the newly created estimation error problem.

Investigating the relevance of the single strategies within our averaging rule, we turn to Figure 1. This figure shows the average weights of the considered strategies over all simulations when averaging is conducted over $M = 6$ strategies. Panel A displays the averaging weights for AV-Wgt, Panels B and C show the respective weights for AV-Inv and AV-Cov. We observe similar patterns for the three averaging levels. Our rule utilizes, on average, the multitude of strategies, corroborating their relevance for an efficient diversification of estimation errors. Turning to the weighting of the individual strategies, we observe that the weight of the Sample is low when the estimation window is small and

[9] When averaging over the sample and another strategy, we have $M = 2$, which gives rise to five permutations. The number of permutations for $M = 3$ and $M = 4$ amounts to ten, respectively, for $M = 5$ to five, and for $M = 6$ to one.

[10] We evaluate the distribution of the out-of-sample variances over all simulation runs and permutations for AV-Wgt, AV-Inv, and AV-Cov using kernel densities in the left column of Figures B.3, B.4, and B.5 of the Internet Appendix. We observe that the distributions for $M = 6$ center around smaller out-of-sample variances and collapse more tightly around this value than the distributions for smaller values of M. This result holds for all averaging levels and supports the finding that combining all six strategies is beneficial when compared to averaging over a subset.

Table 5: Descriptive statistics of estimation errors across estimation windows

This table shows the means, standard deviations (Std), and Bravais-Pearson correlation coefficients of the estimation errors between the single strategies in Panel B of Table 3. The estimation error for the m-th portfolio strategy in the k-th simulation run is defined as the following loss: $L_{k,m} = (\hat{w}_{k,m} - w_k)' \Sigma_k (\hat{w}_{k,m} - w_k)$, where $\hat{w}_{i,k}$ are the estimated minimum-variance portfolio weights and w_k are the true minimum-variance portfolio weights, which are computed using the population covariance matrix of the respective run, Σ_k. The reported values show the correlation over 10,000 simulation runs between the losses of the respective estimators. Panels A - E report the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The reported means and standard deviations are scaled by 10e+03, respectively.

	Sample	1F	CC	ID	SC	POET
Panel A: Estimation window $\tau = 60$						
Mean	1.326	0.440	0.537	0.940	0.441	0.548
Std	0.524	0.148	0.193	0.118	0.150	0.169
Cross-correlations						
Sample	1.000	0.365	0.224	-0.171	0.259	0.434
1F		1.000	0.357	-0.174	0.139	0.745
CC			1.000	-0.345	0.349	0.413
ID				1.000	-0.086	-0.257
SC					1.000	0.347
POET						1.000
Panel B: Estimation window $\tau = 120$						
Mean	0.481	0.394	0.480	0.940	0.306	0.402
Std	0.159	0.128	0.166	0.118	0.092	0.115
Cross-correlations						
Sample	1.000	0.216	0.180	-0.209	0.347	0.418
1F		1.000	0.106	-0.024	-0.043	0.609
CC			1.000	-0.370	0.327	0.232
ID				1.000	-0.060	-0.189
SC					1.000	0.191
POET						1.000
Panel C: Estimation window $\tau = 240$						
Mean	0.209	0.376	0.453	0.940	0.228	0.291
Std	0.066	0.120	0.151	0.118	0.066	0.088
Cross-correlations						
Sample	1.000	0.103	0.140	-0.220	0.282	0.334
1F		1.000	-0.025	0.106	-0.112	0.492
CC			1.000	-0.388	0.344	0.097
ID				1.000	0.013	-0.048
SC					1.000	0.053
POET						1.000

Table continues on next page

Table 5 continues here

	Sample	1F	CC	ID	SC	POET
Panel D: Estimation window $\tau = 480$						
Mean	0.099	0.369	0.438	0.940	0.185	0.212
Std	0.030	0.113	0.145	0.118	0.056	0.072
Cross-correlations						
Sample	1.000	0.024	0.117	-0.210	0.162	0.225
1F		1.000	-0.058	0.194	-0.101	0.391
CC			1.000	-0.380	0.323	0.014
ID				1.000	0.109	0.090
SC					1.000	-0.023
POET						1.000
Panel E: Estimation window $\tau = 960$						
Mean	0.049	0.366	0.432	0.940	0.162	0.164
Std	0.015	0.109	0.140	0.118	0.053	0.059
Cross-correlations						
Sample	1.000	-0.026	0.114	-0.238	0.077	0.141
1F		1.000	-0.057	0.226	-0.113	0.349
CC			1.000	-0.392	0.308	-0.027
ID				1.000	0.158	0.180
SC					1.000	-0.007
POET						1.000

Table 6: Simulation-based average out-of-sample variances for a varying number of strategies

This table reports the average out-of-sample variances of our proposed averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level over 10,000 simulation runs for a varying number of single strategies, M. The out-of-sample variance in the k-th simulation run is computed as $\hat{\sigma}_k^2 = \hat{w}_{k,m}' \Sigma_k \hat{w}_{k,m}$, where Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the portfolio weights of the m-th strategy based on the τ simulated returns. Panels A-E show the respective weights for the estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for $M = 2$ up to $M = 6$ represent averages over all possible permutations from combining the Sample with the five remaining strategies from Panel B in Table 3. The reported average out-of-sample variances are scaled by 10e+03.

	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6$
Panel A: Estimation window $\tau = 60$					
AV-Wgt	2.400	2.281	2.232	2.191	2.190
AV-Inv	2.399	2.279	2.232	2.190	2.189
AV-Cov	2.347	2.289	2.261	2.240	2.237
Panel B: Estimation window $\tau = 120$					
AV-Wgt	2.200	2.134	2.097	2.063	2.062
AV-Inv	2.201	2.134	2.097	2.064	2.063
AV-Cov	2.170	2.131	2.107	2.091	2.089
Panel C: Estimation window $\tau = 240$					
AV-Wgt	2.052	2.030	2.013	1.995	1.994
AV-Inv	2.052	2.030	2.014	1.995	1.994
AV-Cov	2.044	2.028	2.016	2.007	2.006
Panel D: Estimation window $\tau = 480$					
AV-Wgt	1.967	1.960	1.955	1.949	1.948
AV-Inv	1.967	1.960	1.955	1.949	1.948
AV-Cov	1.969	1.963	1.960	1.956	1.956
Panel E: Estimation window $\tau = 960$					
AV-Wgt	1.925	1.923	1.922	1.920	1.920
AV-Inv	1.925	1.923	1.922	1.920	1.920
AV-Cov	1.925	1.923	1.922	1.921	1.921

vice versa. The average loss of the Sample is high (low) for smaller (higher) estimation windows, offering higher (lower) diversification gains. These higher (lower) gains are leveraged by assigning higher (lower) weights to the remaining strategies. This is in line with Table 5 where the average loss of the Sample decreases as the estimation window increases.

In the context of our proposed rule, the previous findings demonstrate that averaging over a greater number of strategies is superior to averaging over fewer. This result holds irrespective of the averaging level as well as the estimation window. We further find that averaging over more strategies is not only beneficial with respect to the out-of-sample variance, but also in terms of the out-of-sample Sharpe ratio.[11] Thus, for the rest of our paper we apply our rule to the Sample and the remaining five strategies in Panel B of Table 3, setting $M = 6$.

We compare the out-of-sample variances of our averaging rule against the considered established strategies in Table 7. We observe that our rule achieves, on average, lower out-of-sample variances than the single strategies over all estimation windows. This finding again holds for all averaging levels. We note that the out-of-sample variances of our rule are, on average, lower than any of the sophisticated benchmarks in Panel C.[12]

We report the risk-adjusted performance of our rule in comparison to the established strategies in Table 8. We observe that AV-Wgt delivers, on average, slightly higher Sharpe ratios than AV-Inv and AV-Cov. The Sharpe ratios of AV-Wgt for $\tau = 60$ and 120 are

[11] We report the results for the average out-of-sample Sharpe ratio in Table B.15 of the Internet Appendix. We find that the average Sharpe ratio increases with the number of strategies. Investigating the distribution of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov using kernel densities, we turn to the right column of Figures B.3, B.4, and B.5 of the Internet Appendix. We find that the distributions for $M = 6$ are centered around a higher mean value and collapse for estimation windows up to $\tau = 120$ more tightly around the mean value. We observe that this effect diminishes as the estimation window increases.

[12] We investigate the out-of-sample variances of our rule in comparison to the existing strategies in more detail, exploring the kernel densities for AV-Wgt, AV-Inv, and AV-Cov in Figures B.6, B.7, and B.8 of the Internet Appendix. The left column in each figure shows the densities in comparison to the single strategies. We find that the distributions of AV-Wgt, AV-Inv, and AV-Cov are centered around a smaller mean value and collapse more tightly relative to the aforementioned strategies for estimation windows up to $\tau = 480$. The kernel densities of each averaging level of our rule also compare favorably to the benchmarks in the right column of the figures. We find that the distributions of AV-Wgt, AV-Inv, and AV-Cov collapse as tightly as the competing benchmarks, but around a smaller mean value. This observation holds for estimation windows of up to $\tau = 480$.

Figure 1: Average model weights in the simulated data sets

This figure plots the average weight over all 10,000 simulations that our averaging rule assigns to one of the considered estimators for estimation window sizes of $\tau = \{60, 120, 240, 480, 960\}$. Panel A shows the mean averaging weights when averaging is conducted on the portfolio weight level (AV-Wgt), Panel B when averaging is conducted on the the inverse (AV-Inv), and Panel C when averaging is conducted on the portfolio weight level (AV-Wgt). The abbreviations for the six single strategies are reported in Panel B of Table 3.

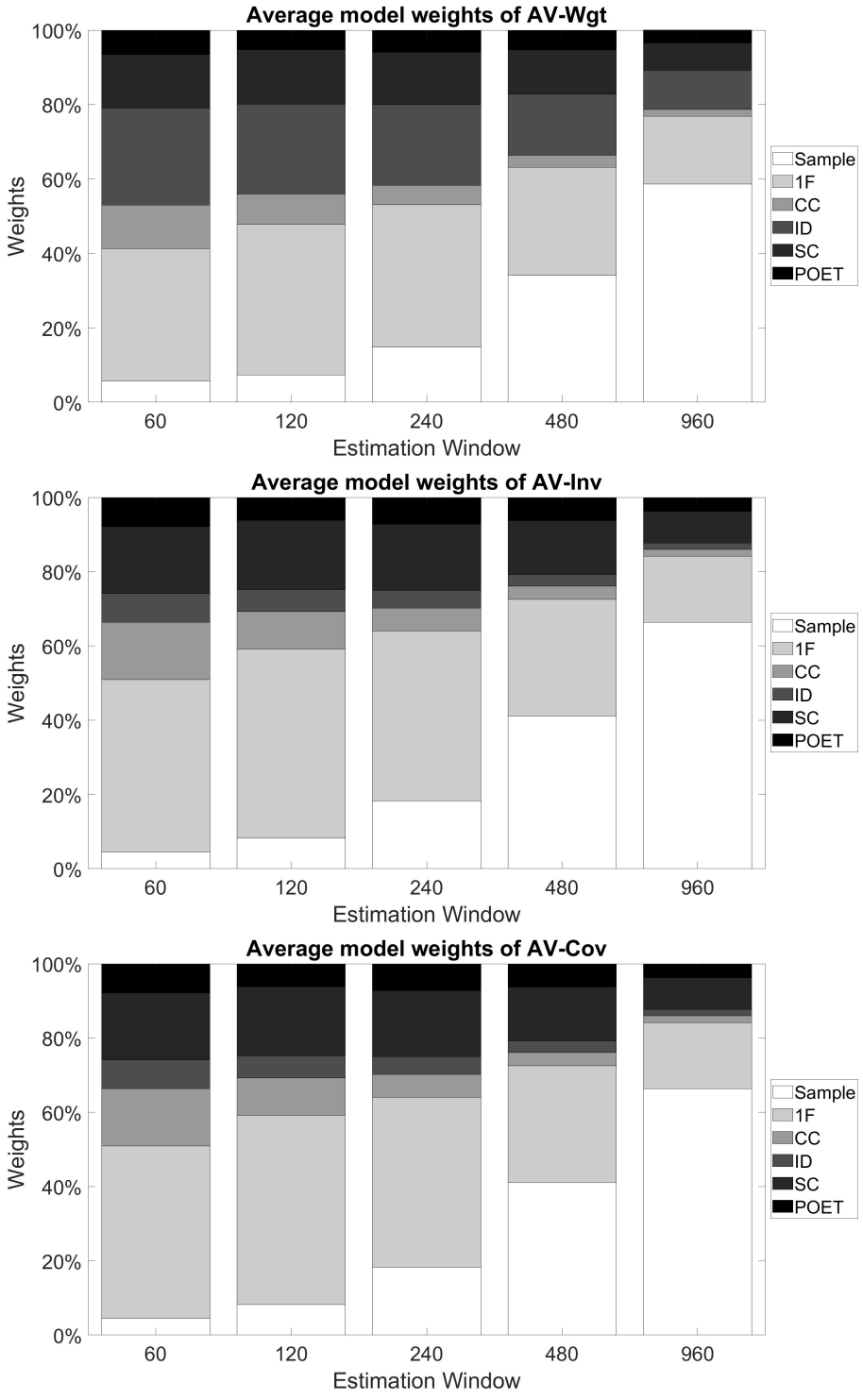


Table 7: Simulation-based average out-of-sample variances for the averaging rule, strategies, and benchmarks

This table reports the average out-of-sample variances of our proposed averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level over 10,000 simulation runs in comparison to established strategies. The out-of-sample variance in the k-th simulation run is computed as $\hat{\sigma}_k^2 = \hat{w}_{k,m}' \Sigma_k \hat{w}_{k,m}$, where Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the portfolio weights of the m-th strategy based on the τ simulated returns. The reported variances for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$ are scaled by 10e+03. The results for AV-Wgt, AV-Inv, and AV-Cov are based on averaging over all $M = 6$ strategies from Panel B in Table 3.

	$\tau = 60$	$\tau = 120$	$\tau = 240$	$\tau = 480$	$\tau = 960$
Panel A: Averaging rule across three levels					
AV-Wgt	2.190	2.062	1.994	1.948	1.920
AV-Inv	2.189	2.063	1.994	1.948	1.920
AV-Cov	2.237	2.089	2.006	1.956	1.921
Panel B: Single strategies					
Sample	3.203	2.357	2.085	1.975	1.927
1F	2.316	2.271	2.251	2.244	2.244
CC	2.413	2.357	2.329	2.313	2.311
ID	2.816	2.816	2.816	2.816	2.816
SC	2.317	2.183	2.103	2.060	2.040
POET	2.424	2.290	2.167	2.088	2.042
Panel C: Benchmarks					
LW1F	2.276	2.141	2.034	1.963	1.924
LWID	2.474	2.225	2.057	1.968	1.925
LWCC	2.315	2.165	2.044	1.966	1.925
LWNLS	2.338	2.166	2.041	1.965	1.926
KDM	2.316	2.271	2.251	2.244	2.244

Table 8: Simulation-based average out-of-sample Sharpe ratios for the averaging rule, strategies, and benchmarks

This table reports the average out-of-sample Sharpe ratios of our proposed averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level over 10,000 simulation runs in comparison to established portfolio strategies. The out-of-sample Sharpe ratio in the k-th simulation run is computed as $\hat{\Psi}_k = \left(\hat{w}'_{k,m}\mu_k\right)\left(\hat{w}_{k,m}\Sigma_k\hat{w}_{k,m}\right)^{-1/2}$, where μ_k is the population mean vector, and Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the estimated portfolio weights of the m-th strategy based on the τ simulated returns. The reported Sharpe ratios for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$ are scaled by 10e+03. The results for AV-Wgt, AV-Inv, and AV-Cov are based on averaging over all $M = 6$ strategies from Panel B in Table 3.

	$\tau = 60$	$\tau = 120$	$\tau = 240$	$\tau = 480$	$\tau = 960$
Panel A: Averaging rule across three levels					
AV-Wgt	140.320	142.616	143.620	143.577	143.478
AV-Inv	140.341	142.571	143.580	143.547	143.460
AV-Cov	139.006	141.766	143.051	143.410	143.230
Panel B: Single Portfolio Strategies					
Sample	110.089	127.257	135.407	138.937	140.939
1F	128.728	127.730	126.888	126.411	126.456
CC	134.878	135.983	136.395	136.773	136.936
ID	136.063	136.063	136.063	136.063	136.063
SC	139.594	142.490	144.272	145.157	145.734
POET	129.419	132.886	136.204	138.706	140.071
Panel C: Benchmarks					
LW1F	131.854	134.464	137.395	139.498	141.087
LWID	129.352	133.675	137.726	139.910	141.370
LWCC	134.884	136.993	139.235	140.677	141.751
LWNLS	130.462	132.675	134.965	136.282	137.310
KDM	128.728	127.730	126.888	126.411	126.456

higher than the Sharpe ratios of any strategy in Panel B. We find that only SC delivers, on average, a higher Sharpe ratio for the remaining estimation windows. The average Sharpe ratios of AV-Wgt in comparison to the benchmarks in Panel C, are higher than those of the considered strategies across all estimation windows. These results suggest that our rule compares favorably to the existing strategies not only with respect to the variance, but also in terms of the Sharpe ratio.[13]

Our simulation study has three key results: First, the contributions of the estimation errors to the out-of-sample variances are uncorrelated between the considered strategies, offering sizable diversification benefits. This suggests that averaging over a multitude of strategies reduces the overall impact of estimation errors. Second, combining more strategies yields a lower variance and a higher risk-adjusted performance. The performance across the averaging levels is similar, but slightly in favor of the portfolio weight level. We thus use AV-Wgt as the reference strategy in the empirical section of this paper. Third, we find that our rule compares favorably to the set of established strategies, suggesting that the combination of existing strategies represents an efficient way to alleviate the impact of estimation errors on out-of-sample performance.

4 Empirical study

4.1 Data

We evaluate the out-of-sample performance of our averaging rule in comparison to the eleven competing strategies on five empirical data sets based on daily excess returns.[14]

[13] The kernel densities of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov in Figures B.9, B.10, and B.11 of the Internet Appendix, corroborate the favorable comparison relative to the existing strategies. The left column of the aforementioned figures presents the results in comparison to the single strategies. We observe that the distribution of the out-of-sample Sharpe ratios for all averaging levels of our rule collapses more tightly around a higher mean value than for the aforementioned strategies, except SC. The right column of the figures shows the results in comparison to the benchmarks. We find that our rule centers for all estimation windows around a higher mean value and collapses more tightly around the respective mean for estimation windows up to $\tau = 120$.

[14] Our choice of a daily data frequency follows Ledoit and Wolf (2017) and is also in-line with industry standards. We report empirical results on a monthly data frequency in Tables C.17 - C.20 of the Internet Appendix. We find that the results based on monthly excess returns are qualitatively similar to the ones presented in this section.

Table 9 provides an overview of the data sets and their source. We use the Fama-French six- and 25-factor as well as the ten- and 30-industry portfolios from Kenneth French's homepage. The construction of the STOCK500 data set follows Chan et al. (1999), Jagannathan and Ma (2003), and DeMiguel, Garlappi, and Uppal (2009). In July of each year, we select the 500 largest stocks in terms of market capitalization from all NYSE, AMEX, and NASDAQ stocks in the Center for Research in Security Prices (CRSP) database that fulfill the following specific criteria. We filter out stocks with a price of less than \$5, as well as stocks with missing excess returns in the 756 trading days preceding, or the 252 days subsequent to the selection date.[15] The selected stocks constitute the investment universe for one year, which is then revised based on the outlined criteria.

We compute the out-of-sample excess returns for each portfolio strategy from Table 3 using the common rolling-sample procedure, setting $\tau = 756$ observations. Starting in June 1973, we estimate the portfolio weights of each portfolio strategy using only the information in the estimation window comprising the most recent τ excess returns. We hold the estimated portfolio weights constant for 21 trading days, representing a one-month holding period, and save the corresponding out-of-sample excess returns. We then move the estimation window forward by 21 trading days and repeat the aforementioned procedure over the entire sample period.

4.2 Performance evaluation

We compare our averaging rule with each competing portfolio strategy in terms of variance (s^2) and Sharpe ratio (SR) of out-of-sample excess returns, the average portfolio turnover (TRN), and the average short interest (SI). The respective metrics for the i-th portfolio strategy are the same as used by DeMiguel et al. (2013) and given by:

$$s_i^2 = \frac{1}{T-\tau-1}\sum_{t=\tau}^{T-1}\left(\hat{w}_{i,t}' r_{t+1} - \bar{r}_i\right)^2, \text{ with } \bar{r}_i = \frac{1}{T-\tau}\sum_{t=\tau}^{T-1}\hat{w}_{i,t}' r_{t+1}, \tag{24}$$

[15] We use the common convention that a calendar month comprises 21 trading days throughout the empirical section of this paper. The 252 and 756 trading days correspond to one and three years of data.

Table 9: List of the empirical data sets for daily data

This table lists the data sets for the empirical evaluation of our proposed averaging rule in comparison to existing portfolio strategies, their abbreviations, the number of assets in each data set, and the data sources. All data sets span the period 06/1973 - 06/2019, constitute daily data, and apply in the case of portfolio data the value weighting scheme to the respective constituents. Data from Kenneth French is taken from his website (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) and represents different cuts of the U.S. stock market. The STOCK500 data set contains the 500 largest single stocks in terms of market capitalization in July of every year after filtering out stocks that have a price of less than \$5, or exhibit missing returns in the preceding 756 and subsequent 252 trading days to the selection date. All stock prices are taken from the Center of Research in Security Prices (CRSP).

#	Data set	Abbreviation	N	Source
1	6 Fama and French portfolios of firms sorted by size and book-to-market	6FF	6	K. French
2	25 Fama and French portfolios of firms sorted by size and book-to-market	25FF	25	K. French
3	10 industry portfolios representing the U.S. stock market	10Ind	10	K. French
4	30 industry portfolios representing the U.S. stock market	30Ind	30	K. French
5	500 Stocks with the largest market capitalization	STOCK500	500	CRSP

$$SR_i = \frac{\bar{r}_i}{s_i}, \tag{25}$$

$$TRN_i = \frac{1}{T-\tau-1} \sum_{t=1}^{(T-21)/21} \left\| \hat{w}_{i,\tau+t\cdot 21+1} - \hat{w}_{i,(\tau+t\cdot 21)^+} \right\|_1, \tag{26}$$

$$SI_i = \frac{1}{T-\tau} \sum_{t=\tau}^{T} \frac{\|\hat{w}_{i,t}\|_1 - 1}{2}, \tag{27}$$

where $\hat{w}_{i,t}$ are the portfolio weights of strategy i at time t, $\hat{w}_{i,t^+}$ are the weights of rule i at the end of t but before rebalancing, $\hat{w}_{i,t+1}$ are the congruent weights after rebalancing at time $t+1$, and $\|\hat{w}_t\|_1$ is the 1-norm of the portfolio weights at time t. We further denote the vector of excess returns at $t+1$ by r_{t+1}.[16]

We measure the statistical significance of the differences between our averaging rule on the portfolio weight level, AV-Wgt, and LWNLS when investigating the out-of-sample variance using robust inference proposed by Ledoit and Wolf (2011). We compare our rule with ID when evaluating the out-of-sample Sharpe ratio, using the robust procedure from Ledoit and Wolf (2008).[17] We choose LWNLS as a benchmark for variance testing because it represents the most sophisticated strategy in terms of portfolio variance minimization. The ID portfolio serves as a benchmark for the Sharpe ratio because DeMiguel, Garlappi, and Uppal (2009) find that no sophisticated portfolio strategy outperforms the 1/N portfolio.[18]

4.3 Out-of-sample results

Table 10 reports the out-of-sample variances of our proposed rule on the three averaging levels, the single strategies, and the benchmarks. For all considered data sets we observe that our rule achieves comparable variances across all averaging levels.[19] The variance of AV-Wgt is on all averaging levels and across all data sets lower than the variances of the

[16] We use the 30-day T-bill rate from Kenneth French's website to proxy the risk free rate.

[17] We confine the statistical inference to the aforementioned pairwise tests to avoid multiple-testing problems in the inference (see Ledoit and Wolf (2017) and references therein).

[18] Note that the minimum-variance portfolio based on a scalar multiple of the identity matrix corresponds to the 1/N portfolio.

[19] For the remaining performance metrics we also find similar results on the different averaging levels, and thus, confine the discussion to AV-Wgt vis-a-vis the single strategies and benchmarks for the sake of brevity.

single strategies, except the Sample strategy on the 25FF data set.

Contrasting AV-Wgt to the benchmark strategies in Panel C, we find that AV-Wgt achieves on four out of five data sets lower variances than LWNLS, being significant on three out of five. In comparison to the linear shrinkage strategies LW1F and LWID, we find that AV-Wgt delivers on three, respectively four, data sets lower variances. The variance of AV-Wgt is on all levels and across all data sets lower than LWCC and KDM. Our results show that averaging over multiple strategies leads to a more efficient diversification and compares favorably to existing shrinkage approaches in terms of out-of-sample variance.

To study the relevance of each individual strategy within our AV-Wgt rule, we plot the estimated averaging weights for all data sets in Figure 2. We observe that AV-Wgt combines at least two estimators at any point in time across all data sets. The maximum number of strategies with positive weights amounts to four for the 6FF, five for the 30Ind and STOCK500, and all six strategies for the remaining data sets. The results confirm the relevance of each strategy within our averaging rule. Our rule is hereby not confined to combining (always) the sample with just one strategy.

Table 11 shows the out-of-sample Sharpe ratios for our averaging rule and the various strategies. In Panel B, we observe that our rule delivers the second highest Sharpe ratio on the Fama-French and industry portfolios compared to the single strategies. The Sharpe ratio of AV-Wgt on the STOCK500 data set is higher than for any strategy in Panel B. We observe that AV-Wgt delivers on four out of five data sets a statistically significant Sharpe ratio increase relative to the ID strategy. The smallest increase in Sharpe ratio amounts to 25% for the 30Ind portfolio, being economically, but not statistically significant. Thus, we conclude that our averaging rule meets the challenge set forth by DeMiguel, Garlappi, and Uppal (2009) to deliver significantly higher risk-adjusted excess returns.

Contrasting AV-Wgt to the Ledoit-Wolf shrinkage strategies LW1F, LWID, LWCC, and LWNLS, we observe a comparable performance across all data sets.[20] AV-Wgt performs somewhat better on the 6FF and 30Ind data set, while the shrinkage strategies

[20]We also observe that the benchmarks in Panel C generate higher Sharpe ratios than the ID portfolio, which is in line with the results of Ledoit and Wolf (2017).

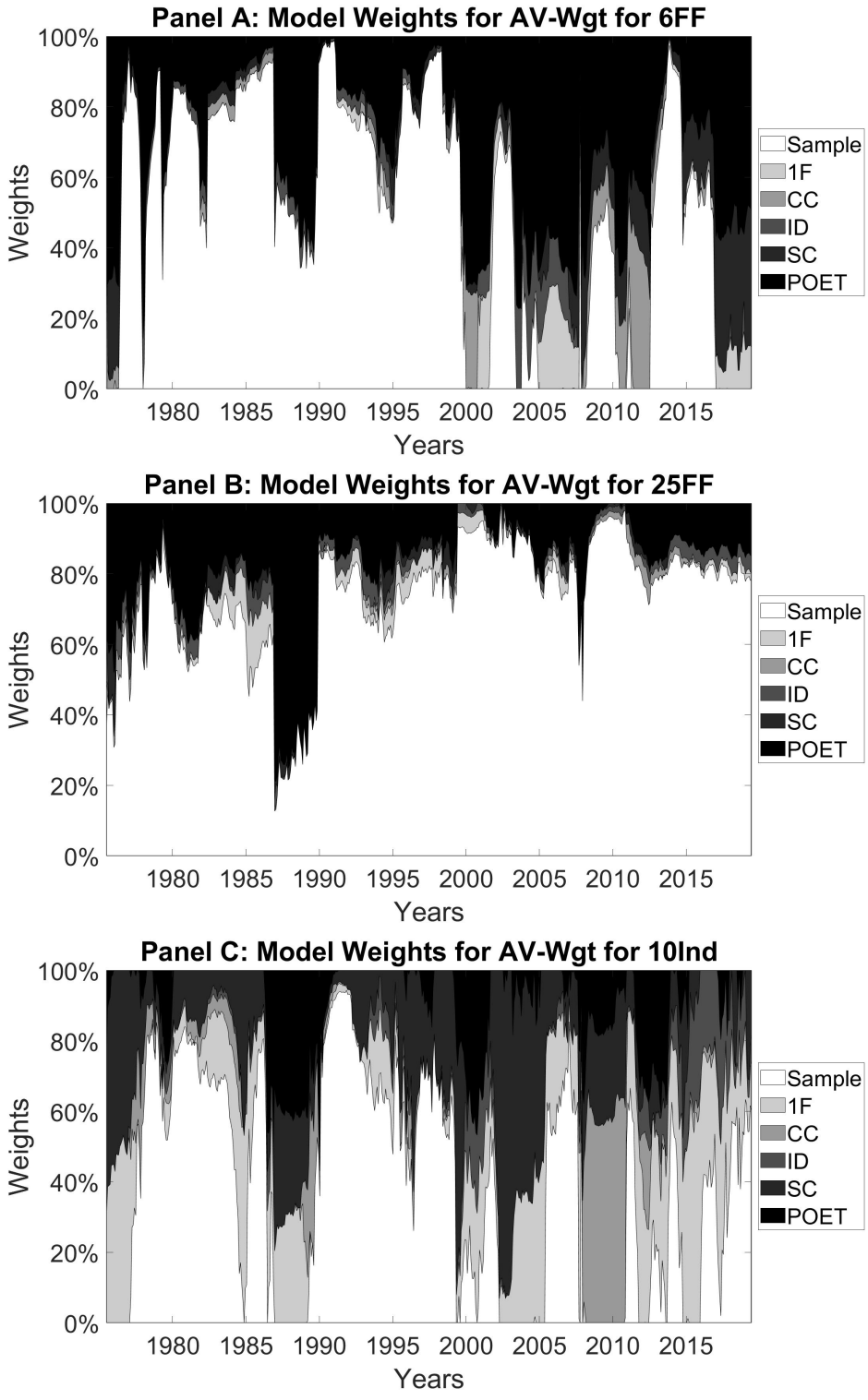
Table 10: Daily out-of-sample variances in empirical data sets

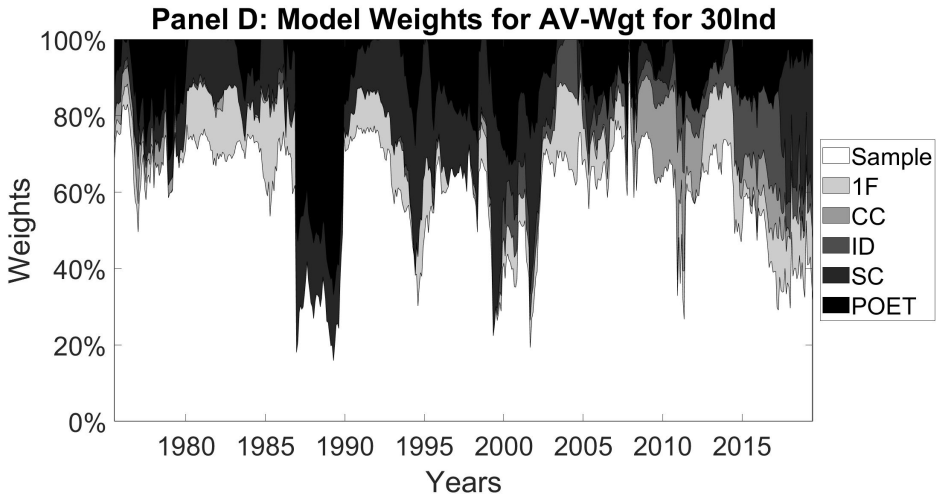
This table reports the daily out-of-sample variances of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The reported variances are scaled by 10e+03. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. We report statistical significance for the null hypothesis wherein the log out-of-sample variance of AV-Wgt is greater than that of the non-linear shrinkage approach of Ledoit and Wolf (2017), LWNLS: $H_0 : log(\hat{\sigma}^2_{AV-Wgt}) - log(\hat{\sigma}^2_{LWNLS}) \leq 0$. We follow Ledoit and Wolf (2011) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across three levels					
AV-Wgt	0.5732	0.3802	0.5233	0.4489	0.2887
AV-Inv	0.5737	0.3806	0.5237	0.4491	0.2897
AV-Cov	0.5724	0.3789	0.5233	0.4517	0.2714
Panel B: Single strategies					
Sample	0.5742	0.3773	0.5305	0.4554	0.5313
1F	0.7787	0.7313	0.5742	0.5654	0.5593
CC	0.7340	0.7903	0.5755	0.6424	0.8001
ID	1.0892	1.0908	0.9686	1.0537	1.1287
SC	0.7700	0.7072	0.5812	0.5415	0.4192
POET	0.5937	0.4056	0.5463	0.5032	0.3247
Panel C: Benchmarks					
LW1F	0.5720	0.3769	0.5297	0.4536	0.2910
LWID	0.5820	0.3766	0.5302	0.4535	0.3648
LWCC	0.5874	0.4046	0.5321	0.4530	0.3219
LWNLS	0.5745	0.3758	0.5289***	0.4530***	0.3328***
KDM	0.7787	0.7313	0.5742	0.5654	0.5593

Figure 2: Average strategy weights of AV-Wgt for portfolios and stocks

This figure plots the averaging weights of the averaging rule on the portfolio weight level (AV-Wgt) over the out-of-sample period from 06/1976 - 06/2019. The weights represent the allocation of the respective estimator within the averaging rule over the holding period, which comprises 21 trading days. Panel A shows the plot for the 6FF, Panel B for the 25FF, Panel C for the 10Ind, Panel D for the 30Ind, and Panel E for the STOCK500 data set. The abbreviations for the portfolio strategies and data sets are reported in Tables 3 and 9, respectively.




Panel D: Model Weights for AV-Wgt for 30Ind
100%
80%
60%
40%
20%
0%
Weights
1980
1985
1990
1995
2000
2005
2010
2015
Years
Sample
1F
CC
ID
SC
POET

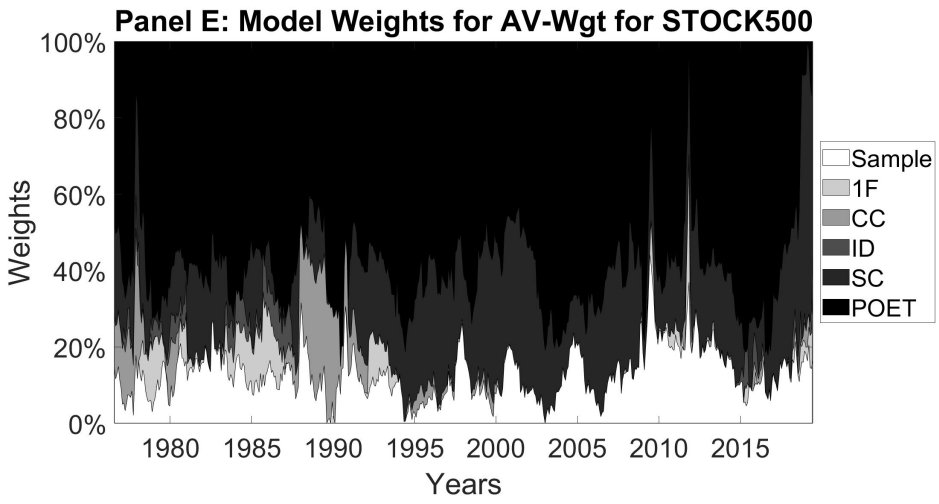

Panel E: Model Weights for AV-Wgt for STOCK500
100%
80%
60%
40%
20%
0%
Weights
1980
1985
1990
1995
2000
2005
2010
2015
Years
Sample
1F
CC
ID
SC
POET

Table 11: Daily out-of-sample Sharpe ratios in empirical data sets

This table reports the daily out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for to the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. We use the 30-day T-bill rate as the risk-free rate. We report statistical significance for the null hypothesis wherein the Sharpe ratio of AV-Wgt is greater than that of the 1/N portfolio, corresponding to the ID strategy: $H_0 : SR_{AV-Wgt} - SR_{ID} \leq 0$. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5 and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across three levels					
AV-Wgt	0.0770	0.0614	0.0471	0.0401	0.0509
AV-Inv	0.0769	0.0615	0.0471	0.0399	0.0499
AV-Cov	0.0760	0.0610	0.0468	0.0399	0.0555
Panel B: Single strategies					
Sample	0.0771	0.0641	0.0464	0.0376	0.0422
1F	0.0606	0.0452	0.0498	0.0398	0.0428
CC	0.0624	0.0460	0.0451	0.0391	0.0331
ID	0.0343***	0.0189***	0.0332*	0.0320	0.0363*
SC	0.0507	0.0300	0.0418	0.0404	0.0503
POET	0.0806	0.0546	0.0444	0.0373	0.0421
Panel C: Benchmarks					
LW1F	0.0756	0.0631	0.0467	0.0379	0.0581
LWID	0.0737	0.0614	0.0482	0.0394	0.0503
LWCC	0.0712	0.0550	0.0496	0.0380	0.0540
LWNLS	0.0763	0.0625	0.0465	0.0382	0.0545
KDM	0.0606	0.0452	0.0498	0.0398	0.0428

seem to deliver slightly higher Sharpe ratios on the 10Ind and STOCK500 data sets. The performance of AV-Wgt and the shrinkage strategies on the 10Ind data set is almost identical.

We assess the practicality of our averaging rule against the established strategies in terms of turnover and short interest. The results in Table 12 report that the average monthly turnover of AV-Wgt is modest across all data sets and ranges from 4.721% on the 10Ind to 16.520% on the 25FF data set. The turnover of AV-Wgt is similar to the Ledoit-Wolf shrinkage strategies on the Fama-French and industry portfolios, but only half the turnover of the aforementioned strategies on the STOCK500 data set.[21] The turnover of AV-Wgt is generally lower for all data sets in comparison to the single strategies allowing for short sales, i.e. Sample, 1F, CC and POET.

As shown in Table 13, the short interest of our averaging rule ranges between 45.509% on the 10Ind and 201.652% on the 25FF data set. On average, the reported short interest of AV-Wgt is lower than that of the Ledoit-Wolf shrinkage strategies. The favorable comparison is mainly due to the STOCK500 data set, where AV-Wgt compared to LWNLS and LWID realizes a short interest reduction between 10 and 60%. The short interest of our rule is, on average, slightly higher than for the strategies that allow for short sales, except for the Sample and POET strategy.

To further assess the practicality of our averaging rule compared to the existing portfolio strategies, we account for transaction costs in the empirical analysis by correcting the out-of-sample performance for the implied cost of rebalancing the portfolio. The portfolio return, net of transaction costs, is given by:

$$r_i^{tc} = (1 + r_i)(1 - \kappa \sum_{t=1}^{(T-21)/21} \left\| \hat{w}_{i,\tau+t\cdot 21+1} - \hat{w}_{i,(\tau+t\cdot 21)^+} \right\|_1) - 1 \tag{28}$$

where $\hat{w}_{i,t}$ are the portfolio weights of strategy i at time t, $\hat{w}_{i,t^+}$ are the weights of rule i at

[21] The Fama-French and industry portfolios are constructed at the end of each June and its composition is kept stable for the next 12 months until revision of the applicable investment universe. Next June the composition changes, stocks no longer matching the criteria are dropped, and new stocks are included in the portfolios. This cannot be accounted for during the turnover computation and, consequently, we apply the same logic to the CRSP stock data set, i.e. the replacement of a stock in the investment universe is not considered for the turnover.

Table 12: Average monthly turnover in empirical data sets based on daily observations

This table reports the average monthly turnover of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. Turnover is measured as the average percentage of total wealth traded in each period. The reported numbers are in percentages.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across three levels					
AV-Wgt	11.35	16.52	4.72	8.51	11.36
AV-Inv	11.49	16.62	4.76	8.59	11.27
AV-Cov	11.32	16.31	4.79	8.56	13.03
Panel B: Single strategies					
Sample	12.07	17.67	5.78	10.49	49.57
1F	9.39	16.22	7.61	11.00	12.27
CC	7.40	15.21	6.29	9.76	16.49
ID	1.48	1.68	2.32	2.87	5.45
SC	0.46	0.94	1.10	1.86	3.65
POET	15.96	18.25	6.62	11.28	17.75
Panel C: Benchmarks					
LW1F	11.70	17.24	5.74	10.11	22.64
LWID	9.05	15.58	5.00	9.63	33.38
LWCC	8.64	15.18	5.44	9.58	26.65
LWNLS	11.82	17.19	5.70	10.01	18.70
KDM	9.39	16.22	7.60	11.00	12.27

Table 13: Average daily short interest in empirical data sets

This table reports the average daily short interest of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. Short interest is measured by the average amount of wealth that is held in short positions. The reported numbers are in percentages.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across three levels					
FAWgt	143.50	201.65	45.52	75.14	97.75
FAInv	145.02	202.89	45.87	76.00	96.67
FACov	143.57	197.73	45.78	75.51	115.48
Panel B: Single strategies					
Sample	151.70	215.21	57.83	97.21	366.94
1F	102.92	156.50	68.81	88.75	62.30
CC	87.44	146.64	55.29	74.97	85.18
ID	0.00	0.00	0.00	0.00	0.00
SC	0.00	0.00	0.00	0.00	0.00
POET	191.03	210.65	67.52	97.94	113.66
Panel C: Benchmarks					
LW1F	147.10	209.67	57.28	92.65	156.13
LWID	114.82	189.64	48.04	87.77	240.22
LWCC	107.85	180.44	52.12	88.00	192.04
LWNLS	149.13	209.27	56.74	91.52	113.18
KDM	102.92	156.50	68.81	88.75	62.30

the end of t but before rebalancing, $\hat{w}_{i,t+1}$ are the congruent weights after rebalancing at time $t+1$, and $\|\hat{w}_t\|_1$ is the 1-norm of the portfolio weights at time t. κ are the transaction costs of 30bps for rebalancing the portfolio.

In Table 14, we follow DeMiguel et al. (2013) and compare the out-of-sample Sharpe ratios after adjustment for proportional transaction costs of 50bps. Balduzzi and Lynch (1999)argue that 50bps are a good estimate of transaction costs for investors who trade in individual stocks. Comparing our rule to the benchmarks, AV-Wgt performs somewhat better on the 6FF, 30Ind and STOCK500 data set, while the shrinkage strategies seem to deliver slightly higher Sharpe ratios on the 25FF and 10Ind data sets. The cost-adjusted Sharpe ratios of AV-Wgt are higher than for the ID strategy across all data sets, being significantly higher in three out of five. The Sharpe ratio of AV-Wgt is after the adjustment for transaction costs no longer significantly higher than ID on the stock data set due to the higher turnover.

Concluding our empirical analysis, we find that our averaging rule delivers significantly lower out-of-sample variances than established estimators of the global minimum-variance portfolio. This result is important because the out-of-sample variance is the primary performance criterion when estimating a minimum-variance portfolio (see Ledoit & Wolf, 2017). The risk-adjusted performance of our averaging rule is significantly higher in comparison to the 1/N portfolio and is akin to the Ledoit-Wolf shrinkage strategies, including the non-linear estimator of Ledoit and Wolf (2017). The modest turnover, especially on the STOCK500 data set, compares favorably to existing strategies and supports the practicality of our rule.[22] This still holds after adjustment for transaction costs.

[22]In robustness checks in Appendix D, we further test the practicality of our rule by imposing short-sale constraints for all strategies considered for averaging. The results for the out-of-sample variances, Sharpe ratios, and turnovers are qualitatively similar. The variance of AV-Wgt compared to LWNLS is smaller on four out of five data sets, being significantly smaller for 30Ind and STOCK500. The Sharpe ratio of our rule is higher than for the ID strategy on all data sets, being significant in four out of five. The turnover for all strategies is reduced by 50 to 97%.

Table 14: Daily out-of-sample Sharpe ratios adjusted for transaction costs in empirical data sets

This table reports the daily out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C, after adjustment for transaction costs. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. We use the 30-day T-bill rate as the risk-free rate. The out-of-sample Sharpe ratio after adjustments for transaction costs is computed by: $SR_i^{tc} = (\bar{r}_i^{tc})(s_i^{tc})^{-1}$, where $\bar{r}_i^{tc}$ is the average portfolio excess return, net of transaction costs, i.e., $r_i^{tc} = (1+r_i)(1-\kappa \sum_{t=1}^{(T-21)/21} \left\| \hat{w}_{i,\tau+t\cdot 21+1} - \hat{w}_{i,(\tau+t\cdot 21)^+} \right\|_1) - 1$ and s_i^{tc} is the corresponding standard deviation. $\hat{w}_{i,t}$ are the portfolio weights of strategy i at time t, $\hat{w}_{i,t^+}$ are the weights of rule i at the end of t but before rebalancing, $\hat{w}_{i,t+1}$ are the congruent weights after rebalancing at time $t+1$, and $\|\hat{w}_t\|_1$ is the 1-norm of the portfolio weights at time t. κ are the transaction costs of 50bps for rebalancing the portfolio. We report statistical significance for the null hypothesis wherein the Sharpe ratio of AV-Wgt is greater than that of the 1/N portfolio, corresponding to the ID strategy: $H_0 : SR_{AV-Wgt} - SR_{ID} \leq 0$. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5 and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across three levels					
AV-Wgt	0.0734	0.0550	0.0455	0.0370	0.0459
AV-Inv	0.0733	0.0550	0.0455	0.0368	0.0449
AV-Cov	0.0724	0.0546	0.0452	0.0369	0.0496
Panel B: Single strategies					
Sample	0.0733	0.0572	0.0445	0.0339	0.0261
1F	0.0581	0.0407	0.0474	0.0363	0.0389
CC	0.0604	0.0419	0.0432	0.0362	0.0287
ID	0.0339***	0.0185***	0.0326*	0.0313	0.0351
SC	0.0506	0.0298	0.0415	0.0398	0.0490
POET	0.0756	0.0477	0.0423	0.0335	0.0348
Panel C: Benchmarks					
LW1F	0.0719	0.0563	0.0448	0.0343	0.0482
LWID	0.0709	0.0553	0.0466	0.0360	0.0372
LWCC	0.0685	0.0493	0.0478	0.0346	0.0428
LWNLS	0.0726	0.0558	0.0446	0.0346	0.0469
KDM	0.0581	0.0407	0.0474	0.0363	0.0389

5 Conclusion

In this paper, we propose a novel approach to alleviate the impact of estimation errors on out-of-sample portfolio performance. Our approach builds on the idea of shrinkage and consists of combining the Sample with a multitude of structured estimators to minimize the expected out-of-sample variance of the corresponding minimum-variance portfolio. The motivation behind this idea is to bring the principle of diversification from the portfolio to the estimation level. If errors in the individual estimators are not perfectly correlated, the combination of the latter offers diversification benefits. Consequently, the resulting estimates of the minimum-variance portfolio weights from our averaging procedure should be less error prone and, on average, deliver an improved out-of-sample performance.

We evaluate our rule through extensive simulations and find that averaging over a greater number of estimators is superior to averaging over fewer in terms of out-of-sample variance and Sharpe ratio. This observation holds for all possible averaging levels, i.e. the portfolio weight, the inverse, and the covariance level. We attribute the performance gains to the leveraged benefits of estimation error diversification between the considered estimators. This diversification potential exists for large and small estimation windows. Our rule achieves, on average, lower out-of-sample variances than any of the competing eleven minimum-variance strategies. The results for the risk-adjusted performance are similar for small estimation windows, suggesting that our averaging rule alleviates the impact of estimation errors on out-of-sample portfolio performance.

Our empirical results are in line with the findings from the simulation study. Our proposed averaging rule achieves a lower out-of-sample portfolio variance than any single strategy and compares favorably to the sophisticated non-linear shrinkage approach of Ledoit and Wolf (2017). In terms of out-of-sample Sharpe ratio, we find that our averaging rule outperforms the 1/N portfolio on all data sets, with significant increases on four out of five. In terms of turnover and short interest, the portfolio weights of our rule are smooth over time and do not contain extreme positions.

Our results suggest that the diversification of estimation errors through averaging

over the sample and multiple structured estimators provides a fruitful way to reduce the impact of estimation errors on out-of-sample portfolio performance. Future research may elaborate on alternative approaches for determining averaging weights, or may extend the set of estimators used in the averaging procedure to explore potential limits of the suggested estimation error diversification.

A Description of estimators

In this section, we describe the considered covariance matrix estimators introduced in section 2.2.

Sample estimator

The Sample estimator is denoted as $\hat{\Sigma}_S$ and is based on the unbiased estimation of the vector of means of the in-sample asset excess returns in the estimation window, i.e. the sample size τ:

$$\hat{\Sigma}_S = (1/(\tau - 1))\Sigma_{t=1}^{\tau}(r_{i,t} - \mu_S)(r_{i,t} - \mu_S)', \tag{29}$$

where $\mu_S = (1/\tau)\sum_{t=1}^{\tau} r_{i,t}$ is the unbiased sample vector of means and $r_{i,t}$ is the excess return of stock i at time t. Similar to existing shrinkage strategies, our averaging rule combines the Sample estimator with one or more of the following single strategies.

Single-factor model

The return of stock i at time t, $r_{i,t}$, is in the single-factor model of Sharpe (1967) described as:

$$r_{i,t} = \alpha_i + \beta_i r_{MKT,t} + \epsilon_{i,t}, \tag{30}$$

where α_i is the mispricing of stock i, β_i is the beta factor of stock i, $r_{MKT,t}$ is the excess return of the market over the risk-free asset at time t, and $\epsilon_{i,t}$ is the residual return of stock i at time t. Following Chan et al. (1999), the covariance matrix estimator $\hat{\Sigma}_{1F}$ of the single-factor model (1F) is given by:

$$\hat{\Sigma}_{1F} = \hat{\Sigma}^2_{MKT}\hat{\beta}\hat{\beta}' + \hat{\Delta}, \tag{31}$$

where $\hat{\Sigma}^2_{MKT}$, $\hat{\beta}$, and $\hat{\Delta}$ are estimates of market variance, asset betas, and a diagonal matrix with the residual variances along the main diagonal, respectively.

Constant correlation model

Elton and Gruber (1973) suggest a structured estimator of the covariance matrix by assuming that each pair of assets has the identical correlation coefficient, while each of the N assets has an individual standard deviation. The resulting constant correlation covariance matrix estimator is given by:

$$\hat{\Sigma}_{CC} = \Lambda C \Lambda', \tag{32}$$

where Λ denotes a diagonal matrix containing the standard deviations $\hat{\sigma}_{i,i}$ of the assets i, j, and C is the constant correlation matrix with ones along the main diagonal and the average of all pairwise correlations $\hat{\sigma}_{i,j}$ off the main diagonal.

Multiple of the identity matrix

This estimator implies that all assets are uncorrelated and exhibit equal variances. Following Ledoit and Wolf (2004b), the elements along the main diagonal are set to the average variance of all assets in the investment universe. The covariance matrix estimator for the scaled identity matrix model ID is given by:

$$\hat{\Sigma}_{ID} = I\mu_{\hat{s}}, \text{ with } \mu_{\hat{s}} = \frac{1}{N}\sum_{i=1}^{N} \hat{s}_i^2, \tag{33}$$

where $\hat{s}_i^2$ is the sample variance of asset i. The resulting minimum-variance portfolio corresponds to the 1/N portfolio.

Principal components

Fan et al. (2013) introduce the principal orthogonal complement thresholding method, POET, which uses an approximate factor model to define the risk factors based on principal components of the sample covariance matrix and impose a threshold on the remaining principal components after taking out the first K. The covariance matrix estimator $\hat{\Sigma}_{POET}$ of the POET model is given by:

$$\hat{\Sigma}_{POET} = \sum_{i=1}^{K} \hat{\gamma}_{\tau,i} \hat{\xi}_i \hat{\xi}_i' + \hat{R}_K^T, \tag{34}$$

where $\hat{\gamma}_{\tau,1} \geq \hat{\gamma}_{\tau,2} \geq \cdots \geq \hat{\gamma}_{\tau,p}$ are the ordered eigenvalues of the sample covariance matrix $\hat{\Sigma_S}$ over the estimation window τ, $(\hat{\xi}_i)_{i=1}^p$ are the corresponding eigenvectors, $\hat{R}_K^T$ is the principal orthogonal complement of K diverging eigenvalues of $\hat{\Sigma}_S$.

Linear shrinkage estimators of the covariance matrix

The linear shrinkage estimators of Ledoit and Wolf (2003, 2004a, 2004b) have the following general form:

$$\hat{\Sigma}_{LW} = \phi \hat{F} + (1 - \phi) \hat{\Sigma}_S. \tag{35}$$

The shrunken covariance matrix $\hat{\Sigma}_{LW}$ is a convex combination of the estimated sample covariance matrix $\hat{\Sigma}_S$ and a shrinkage target $\hat{F}$, with the shrinkage intensity, ϕ, taking values between zero and one. The optimal shrinkage intensity is determined by minimizing the mean-squared error of $\hat{\Sigma}_{LW}$. The authors consider three different candidates for $\hat{F}$: the single-factor model covariance matrix $\hat{\Sigma}_{LW1F}$ by Ledoit and Wolf (2003), a multiple of the identity matrix $\hat{\Sigma}_{LWID}$ as in Ledoit and Wolf (2004b), and the constant correlation model implied covariance matrix $\hat{\Sigma}_{LWCC}$ by Ledoit and Wolf (2004a).

Shrinkage estimators for the inverse covariance matrix

Kourtis et al. (2012) propose a shrinkage estimator of the inverse covariance matrix, which takes the general form for shrinking the inverse covariance matrix $\hat{\Sigma}'$ as given by:

$$\hat{\Sigma}^{-1} = \phi \hat{F}^{-1} + (1 - \phi)\hat{\Sigma}_S^{-1}, \tag{36}$$

where $\hat{\Sigma}^{-1}$ is a convex combination of the estimated inverse sample covariance matrix $\hat{\Sigma}_S^{-1}$ and an inverse shrinkage target $\hat{F}^{-1}$, with the shrinkage intensity ϕ.

Kourtis et al. (2012) use three particular targets: the identity, the inverse covariance matrix generated by the single-factor model of Sharpe (1963), and a weighted sum of these two matrices. The resulting strategies offer significant improvements in out-of-sample portfolio performance in terms of both variance and risk-adjusted excess returns compared to the GMVP, the 1/N portfolio, and benchmarks based on shrinkage estimators of the covariance matrix. This can be explained by the more immediate impact on portfolio weights by the inverse of the covariance matrix. They also highlight that more than one target further reduces estimation errors. We denote the three-fund strategy of Kourtis et al. (2012) as KDM.

Non-linear shrinkage estimators for the inverse covariance matrix

Ledoit and Wolf (2017) show that a non-linear shrinkage estimator (NLS) based on Stein (1975, 1986) outperforms linear shrinkage both on the covariance matrix and its inverse by correcting the over-dispersed eigenvalues of the sample covariance matrix. This approach is asymptotically optimal for portfolio selection when the number of assets N is of the same magnitude as the sample size τ. The non-linear shrinkage estimator for the NLS model is given by:

$$\hat{\Sigma}_{NLS} \quad := U_\tau \hat{D}_\tau U_\tau' \text{ where } \hat{D}_\tau := Diag(\hat{d}_\tau(\gamma_{\tau,1}), \ldots, \hat{d}_\tau(\gamma_{\tau,N})) \text{ and} \tag{37}$$

$$\hat{d}_\tau(\gamma_{\tau,i}) := \begin{cases} \frac{1}{\gamma_{\tau,i}|\hat{c}(\gamma_{\tau,i})|^2} \text{ if } \gamma_{\tau,i} > 0, \\ \frac{1}{(\frac{N}{\tau}-1)\hat{c}(0)} \text{ if } N > \tau \text{ and } \gamma_{\tau,i} = 0, \end{cases} \tag{38}$$

where U is an orthogonal matrix from a spectral decomposition of the sample covariance matrix $\hat{\Sigma}_S$, $\hat{c}$ is the complex-valued Stieltjes transformation, and $\gamma_{\tau,i}$ are the sample eigenvalues, sorted in increasing order.

Short-sale constraints

Jagannathan and Ma (2003) provide a moment shrinkage interpretation of short-sale constraints in the minimum-variance portfolio optimization problem. The additional weight constraint is formulated as:

$$\hat{w}_i \geq 0, \text{ } \textbf{for } i = 1, 2, \ldots, N. \tag{39}$$

Jagannathan and Ma (2003) show that the asset weights of the short-sale constrained minimum-variance portfolio correspond to those of the GMVP based on the following covariance estimator:

$$\hat{\Sigma}_{SC} = \hat{\Sigma}_S - (\delta 1' + 1\delta'), \tag{40}$$

where $\delta = (\delta_1, \delta_2, \ldots, \delta_N)$ is the vector of Lagrange multipliers for the non-negativity constraints in Equation (39).

B Additional simulation results

This section contains the supporting figures for our simulation study. Complementing the results for the out-of-sample variance in Table 6, we report the results for the average out-of-sample Sharpe ratio in Table B.15. Similar to the variance, we find that the average Sharpe ratio increases with the number of strategies.

In the left column of Figures B.3, B.4, and B.5, we evaluate the distribution of the out-of-sample variances for AV-Wgt, AV-Inv, and AV-Cov with the kernel densities over all simulation runs and permutations. We observe that the distributions for $M = 6$ center around smaller out-of-sample variances and collapse more tightly around this value than the distributions for smaller values of M. The result holds for all averaging levels and supports the finding that combining all six strategies is beneficial when compared to averaging over a subset. Investigating the distribution of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov, we turn to the right column of Figures B.3, B.4, and B.5. We find that the distributions for $M = 6$ are centered around a higher mean value and collapse more tightly around the mean value for estimation windows up to $\tau = 120$. We observe that this effect diminishes as the estimation window increases.

In Figures B.6, B.7, and B.8, we investigate the out-of-sample variances of our averaging rule in relation to the existing strategies in more detail, exploring the kernel densities for AV-Wgt, AV-Inv, and AV-Cov, respectively. The left column in each figure shows the densities in comparison to the single strategies. We find that the distributions of AV-Wgt, AV-Inv, and AV-Cov are centered around a smaller mean value and collapse more tightly for estimation windows up to $\tau = 480$ in comparison to the aforementioned strategies. The kernel densities of each averaging level of our rule also compare favorably to the benchmarks. We find that the distributions of AV-Wgt, AV-Inv, and AV-Cov collapse as tightly as the competing strategies, but around a smaller mean value. This observation holds for estimation windows of up to $\tau = 480$.

The kernel density plots of the out-of-sample Sharpe ratios for AV-Wgt, AV-Inv, and AV-Cov in Figures B.9, B.10, and B.11 corroborate the favorable comparison relative

to the existing strategies. The left column of the aforementioned figures presents the results in comparison to the single strategies. We observe that the distribution of the out-of-sample Sharpe ratios for all averaging levels of our rule collapses more tightly around a higher mean value than the aforementioned strategies, except for SC. The right column of the aforementioned figures shows the results in comparison to the benchmarks. We find that our rule centers around a higher mean value for all estimation windows and collapses more tightly around the respective mean for estimation windows up to $\tau = 120$.

Table B.15: Sharpe ratio comparison of averaging strategies in terms of averaged strategies

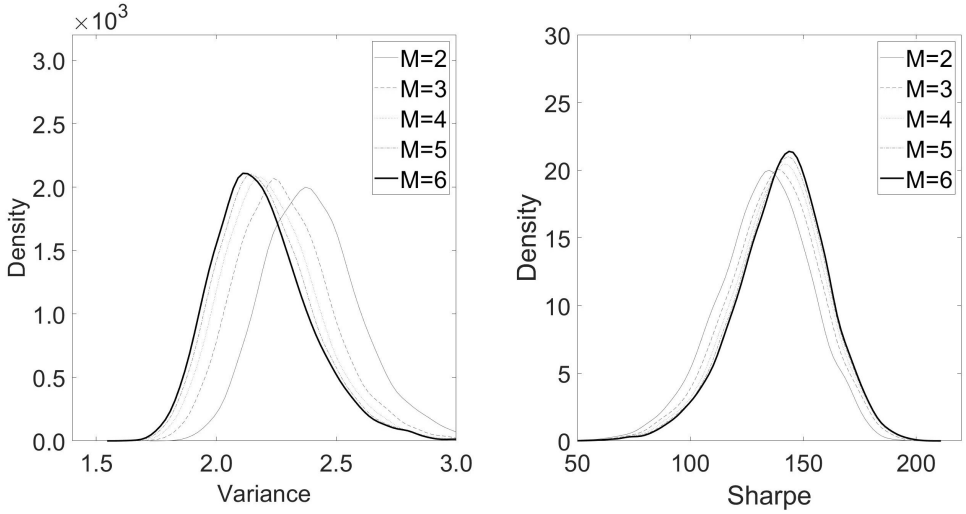
This table reports the average out-of-sample Sharpe ratios of our proposed averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance matrix (AV-Cov) level over 10,000 simulation runs for a varying number of single strategies, M. The out-of-sample Sharpe ratio in the k-th simulation run is computed as $\hat{\Psi}_k = \left(\hat{w}'_{k,m}\mu_k\right)\left(\hat{w}_{k,m}\Sigma_k\hat{w}_{k,m}\right)^{-1/2}$, where μ_k is the population mean vector, and Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the estimated portfolio weights of the m-th strategy based on the τ simulated returns. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The results for $M = 2$ up to $M = 6$ represent averages over all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 3. The reported average out-of-sample Sharpe ratios are scaled by 10e+03.

	$M = 2$	$M = 3$	$M = 4$	$M = 5$	$M = 6$
Panel A: Estimation window $\tau = 60$					
AV-Wgt	132.293	136.889	138.788	140.256	140.320
AV-Inv	132.421	136.980	138.817	140.278	140.341
AV-Cov	133.496	136.412	138.077	138.896	139.006
Panel B: Estimation window $\tau = 120$					
AV-Wgt	135.646	139.653	141.141	142.527	142.616
AV-Inv	135.625	139.636	141.107	142.480	142.571
AV-Cov	136.772	139.582	140.992	141.674	141.766
Panel C: Estimation window $\tau = 240$					
AV-Wgt	138.562	141.202	142.315	143.496	143.620
AV-Inv	138.542	141.176	142.283	143.452	143.579
AV-Cov	139.199	141.309	142.414	142.960	143.051
Panel D: Estimation window $\tau = 480$					
AV-Wgt	140.270	141.819	142.535	143.435	143.577
AV-Inv	140.277	141.795	142.510	143.399	143.547
AV-Cov	140.769	142.133	142.939	143.319	143.410
Panel E: Estimation window $\tau = 960$					
AV-Wgt	141.527	142.359	142.764	143.373	143.478
AV-Inv	141.541	142.346	142.747	143.348	143.460
AV-Cov	141.639	142.413	142.915	143.161	143.230

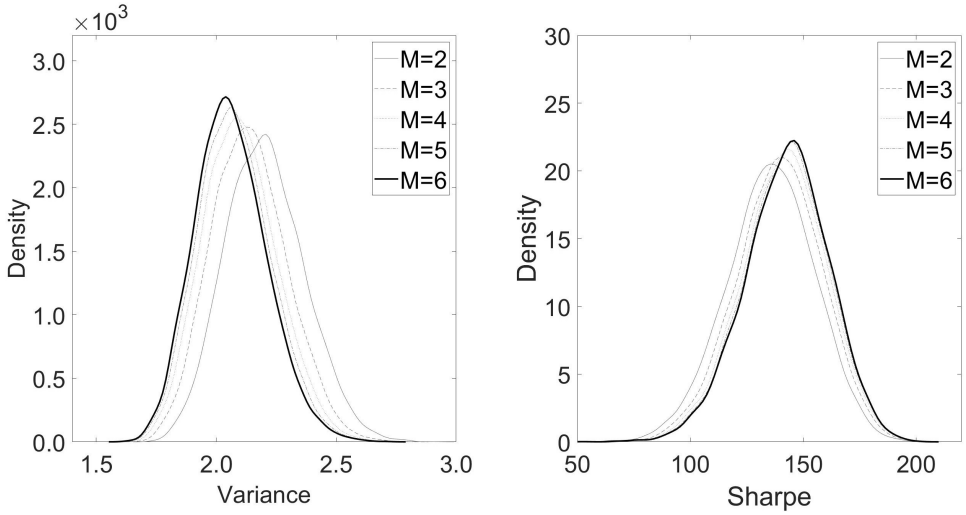
Figure B.3: Kernel densities of the out-of-sample variances and Sharpe ratios of AV-Wgt

This figure plots the kernel densities of the out-of-sample variances and Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt) level over 10,000 simulation runs for a varying number of single strategies, M. The densities for $M = 2$ up to $M = 6$ are based on all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 3. The out-of-sample variance in the k-th simulation run is computed as $\hat{\sigma}_k^2 = \hat{w}_{k,m}' \Sigma_k \hat{w}_{k,m}$, where Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the estimated portfolio weights of the m-th strategy based on the τ simulated returns. The out-of-sample Sharpe ratio in the k-th simulation run is computed as $\hat{\Psi}_k = \left(\hat{w}_{k,m}' \mu_k\right) \left(\hat{w}_{k,m} \Sigma_k \hat{w}_{k,m}\right)^{-1/2}$, where μ_k is the population mean vector. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The reported variances and Sharpe ratios are scaled by 10e+03.

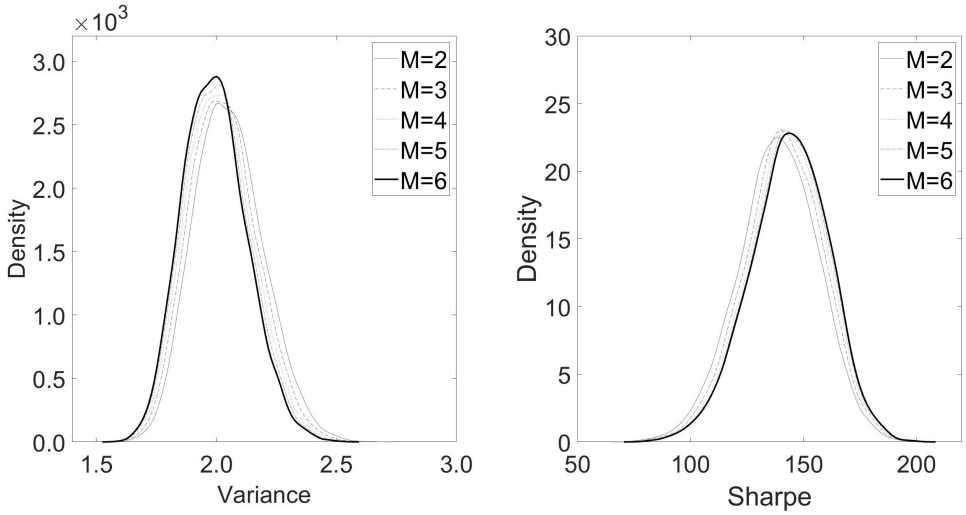
Panel A: Estimation window = 60



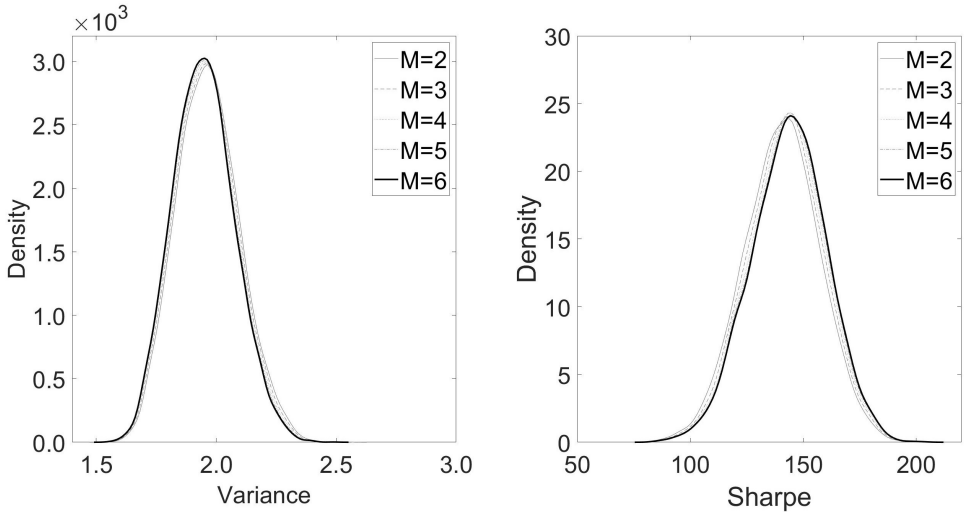
Panel B: Estimation window = 120



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

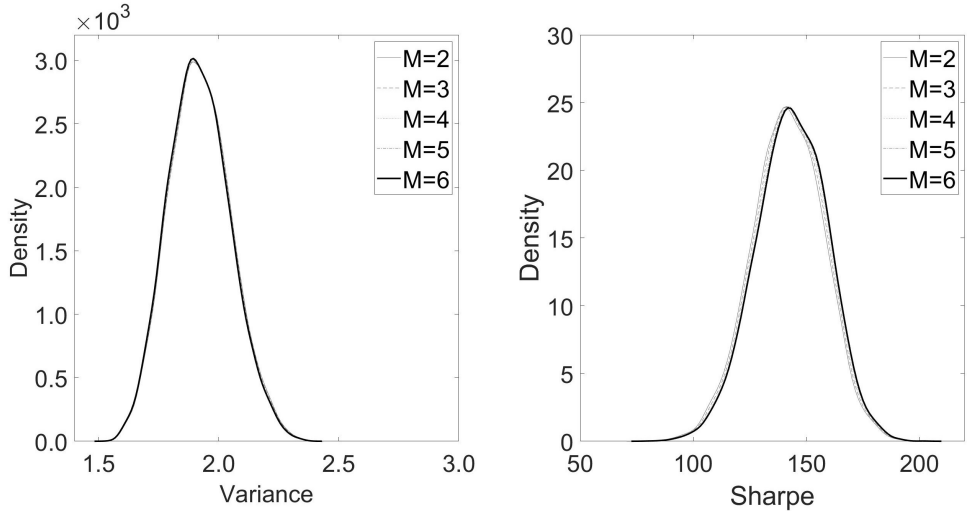
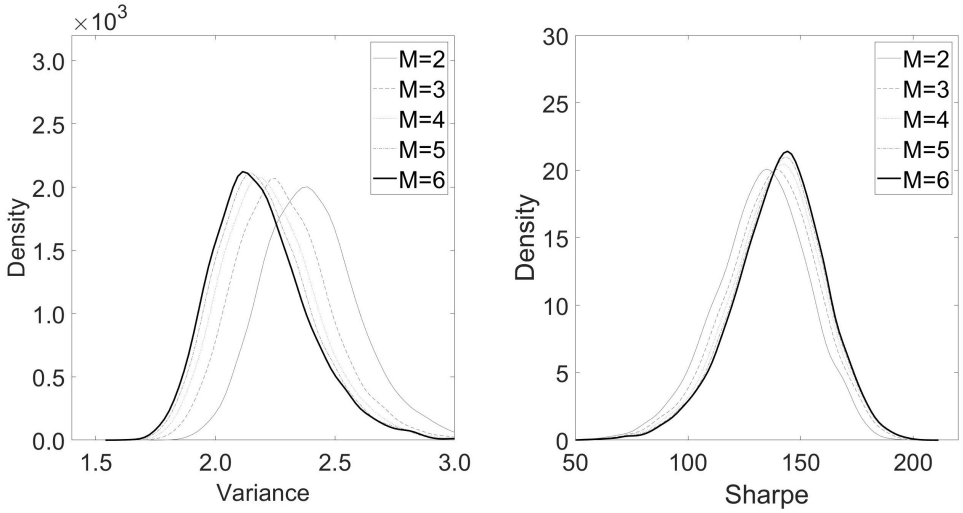


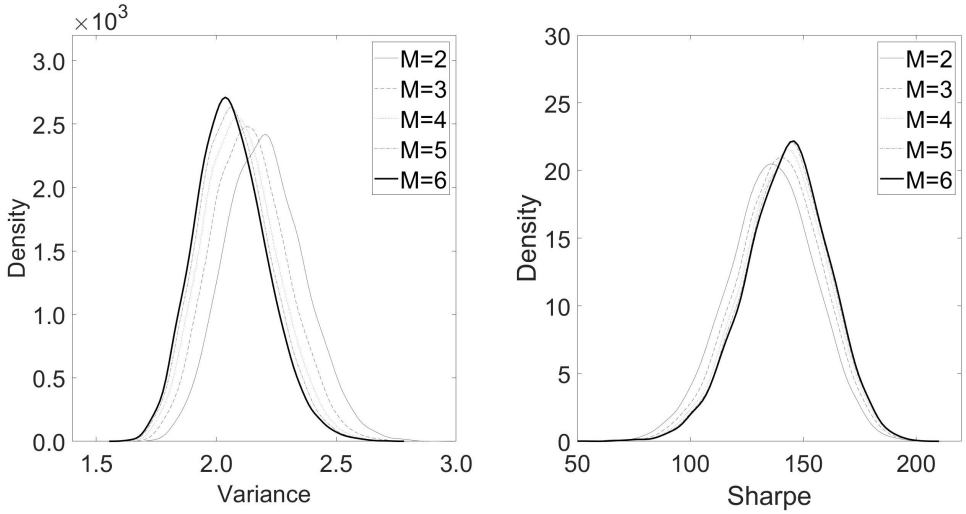
Figure B.4: Kernel densities of the out-of-sample variances and Sharpe ratios of AV-Inv

This figure plots the kernel densities of the out-of-sample variances and Sharpe ratios of our averaging rule on the inverse covariance matrix (AV-Inv) level over 10,000 simulation runs for a varying number of single strategies, M. The densities for $M = 2$ up to $M = 6$ are based on all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 3. The out-of-sample variance in the k-th simulation run is computed as $\hat{\sigma}_k^2 = \hat{w}_{k,m}' \Sigma_k \hat{w}_{k,m}$, where Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the estimated portfolio weights of the m-th strategy based on the τ simulated returns. The out-of-sample Sharpe ratio in the k-th simulation run is computed as $\hat{\Psi}_k = \left(\hat{w}_{k,m}' \mu_k\right) \left(\hat{w}_{k,m} \Sigma_k \hat{w}_{k,m}\right)^{-1/2}$, where μ_k is the population mean vector. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The reported variances and Sharpe ratios are scaled by 10e+03.

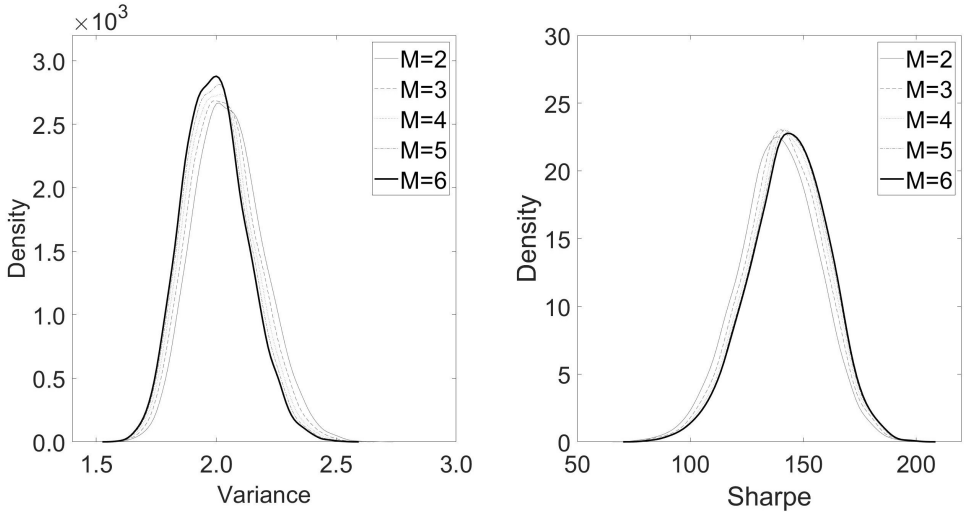
Panel A: Estimation window = 60



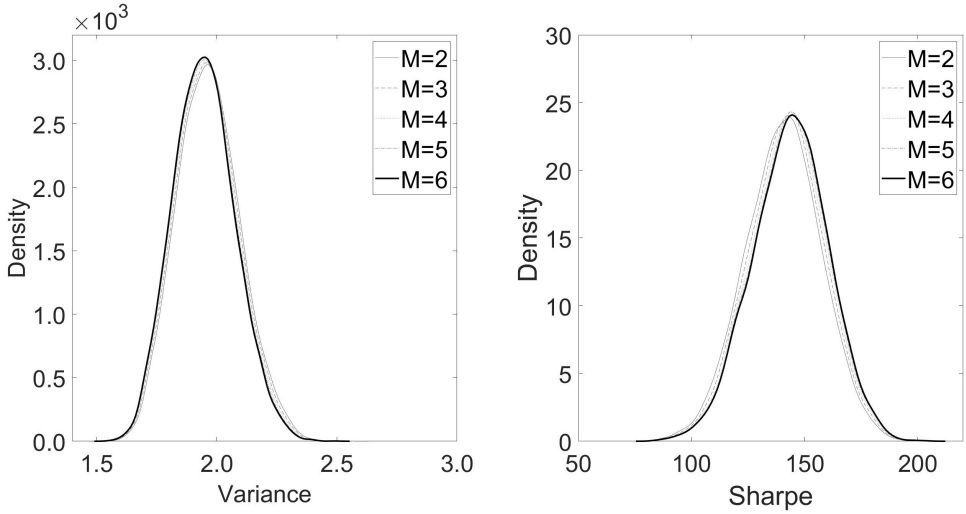
Panel B: Estimation window = 120



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

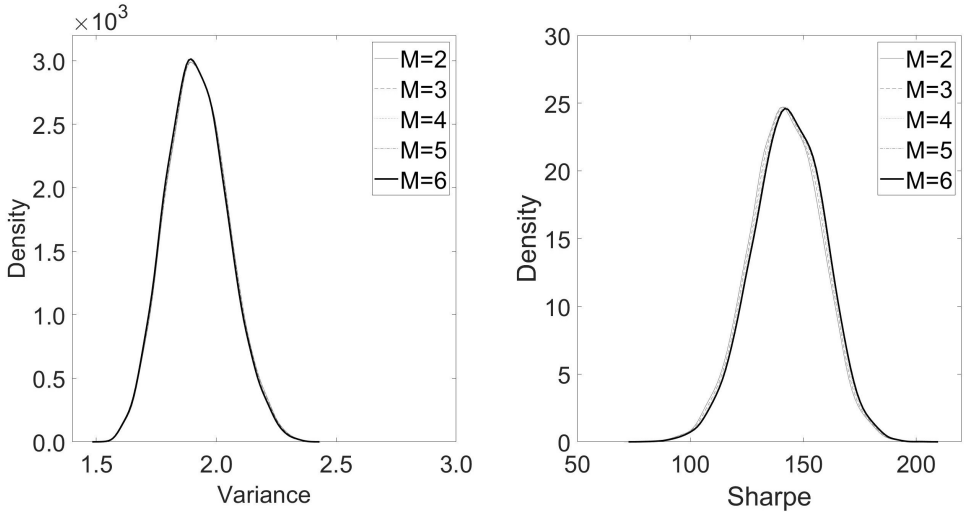
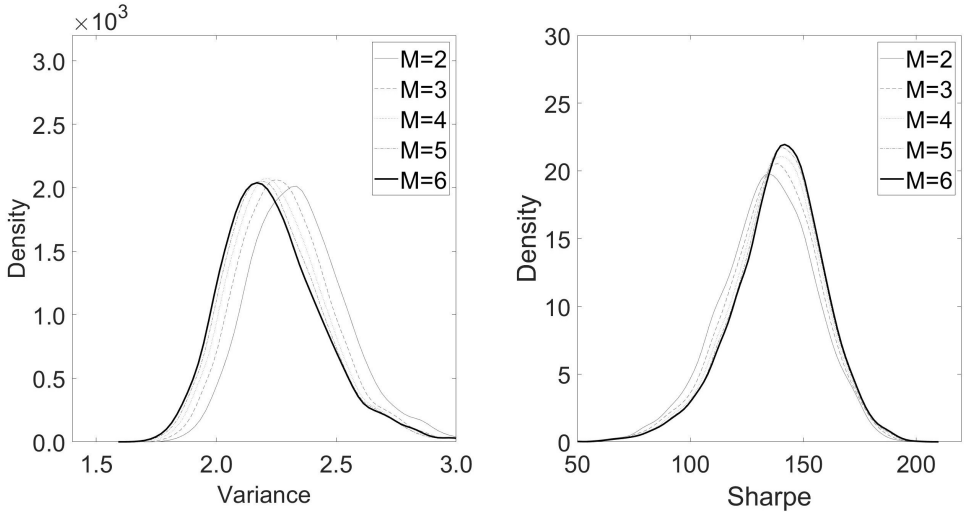


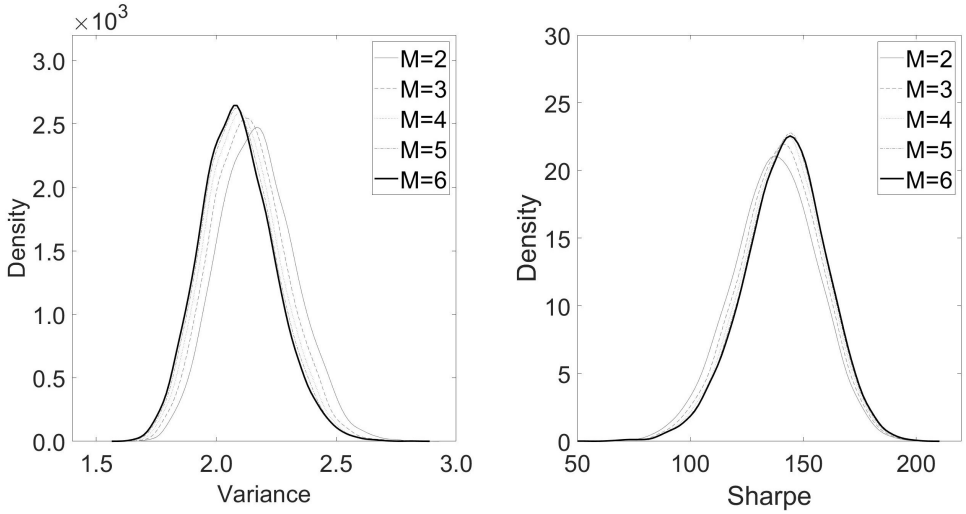
Figure B.5: Kernel densities of the out-of-sample variances and Sharpe ratios of AV-Cov

This figure plots the kernel densities of the out-of-sample variance and Sharpe ratio of our averaging rule on the covariance matrix (AV-Cov) level over 10,000 simulation runs for a varying number of single strategies, M. The densities for $M = 2$ up to $M = 6$ are based on all possible permutations from combining the Sample with the remaining five strategies from Panel B in Table 3. The out-of-sample variance in the k-th simulation run is computed as $\hat{\sigma}_k^2 = \hat{w}_{k,m}' \Sigma_k \hat{w}_{k,m}$, where Σ_k is the population covariance matrix resulting from the random pairing of the assets' factor loadings, their mispricing, as well as their idiosyncratic volatility. $\hat{w}_{k,m}$ are the estimated portfolio weights of the m-th strategy based on the τ simulated returns. The out-of-sample Sharpe ratio in the k-th simulation run is computed as $\hat{\Psi}_k = \left(\hat{w}_{k,m}' \mu_k\right) \left(\hat{w}_{k,m} \Sigma_k \hat{w}_{k,m}\right)^{-1/2}$, where μ_k is the population mean vector. Panels A - E show the respective results for estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The reported variances and Sharpe ratios are scaled by 10e+03.

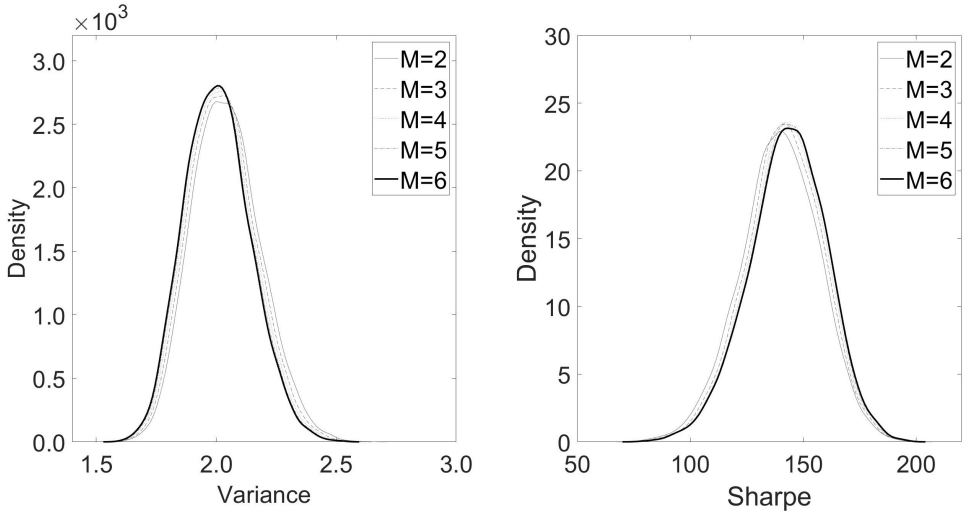
Panel A: Estimation window = 60



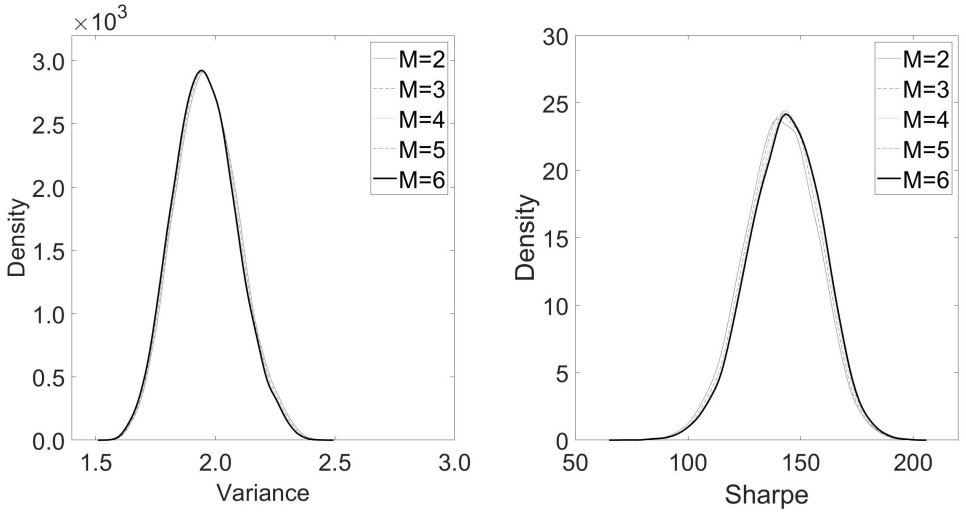
Panel B: Estimation window = 120



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

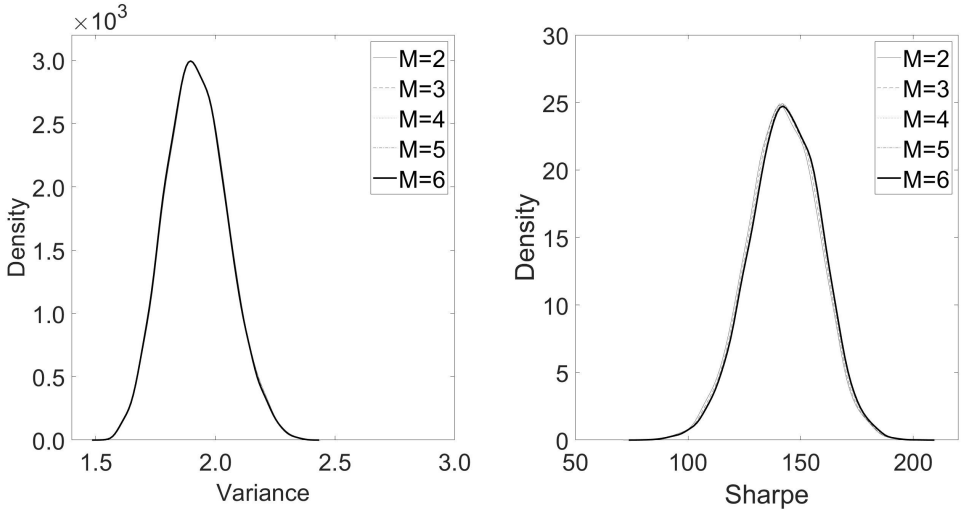
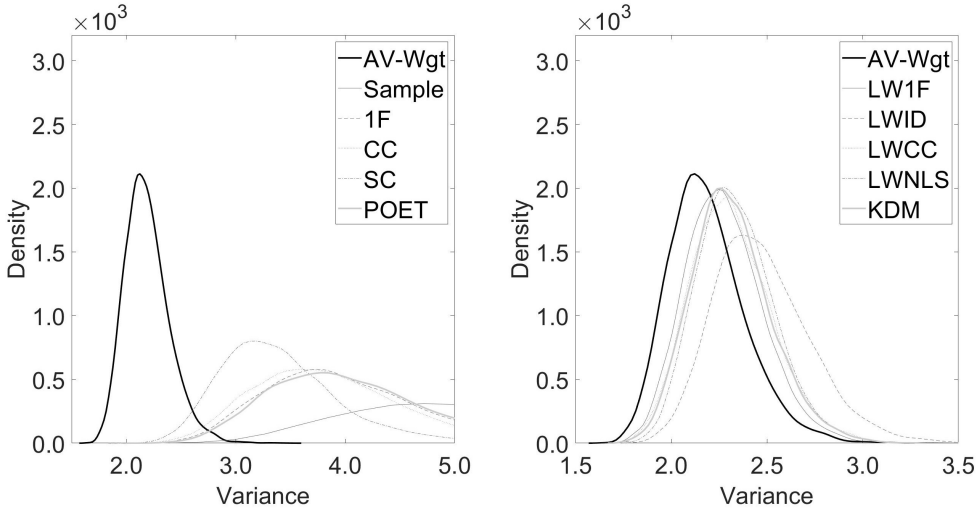


Figure B.6: Kernel densities of the out-of-sample variances of AV-Wgt compared to established strategies and benchmarks

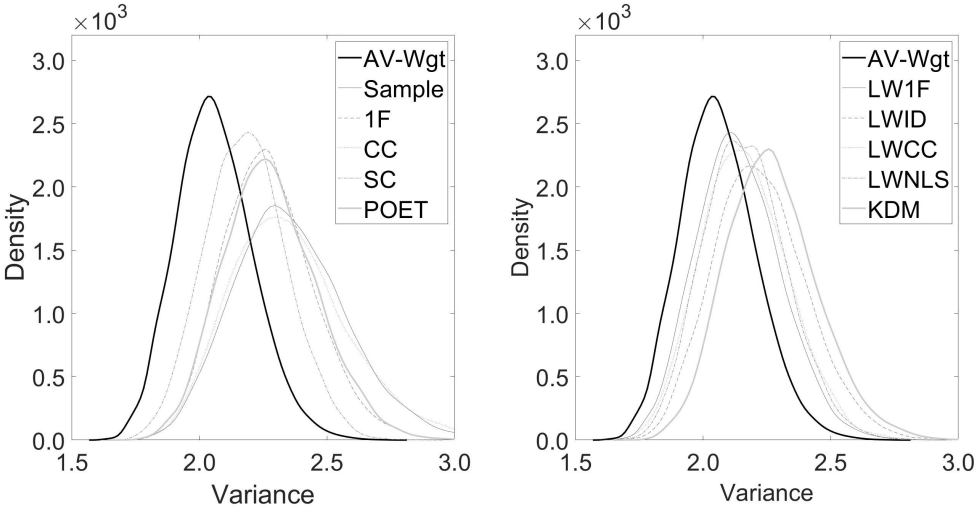
.

This figure plots the kernel densities of the out-of-sample variances of our averaging rule on the portfolio weight (AV-Wgt) level over 10,000 simulation runs in comparison to the single strategies (left column) and the benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The out-of-sample variances are scaled by 10e+03. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 3. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

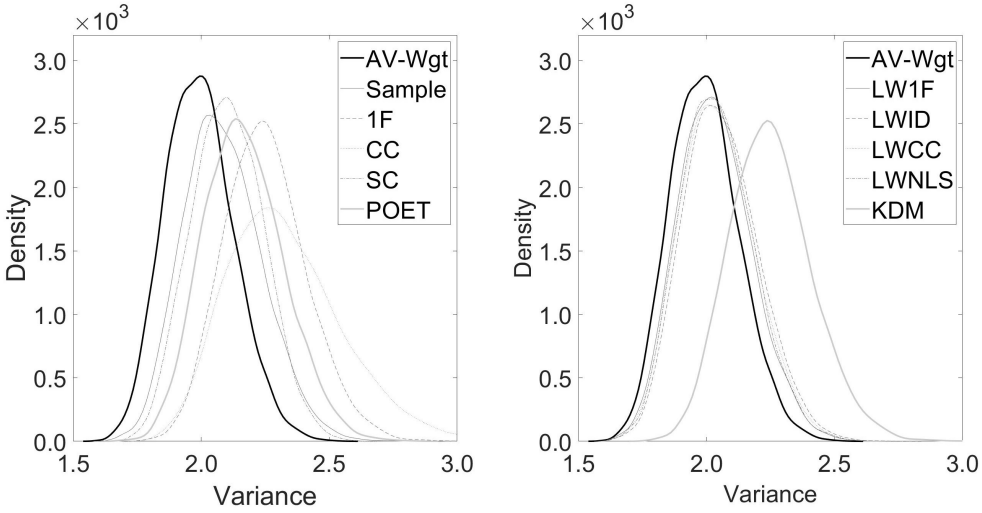
Panel A: Estimation window = 60



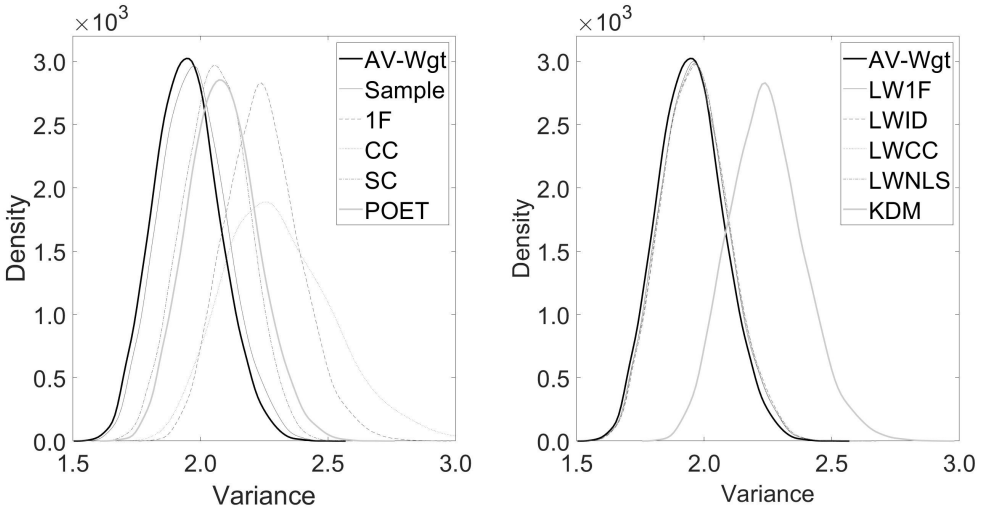
Panel B: Estimation window = 120



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

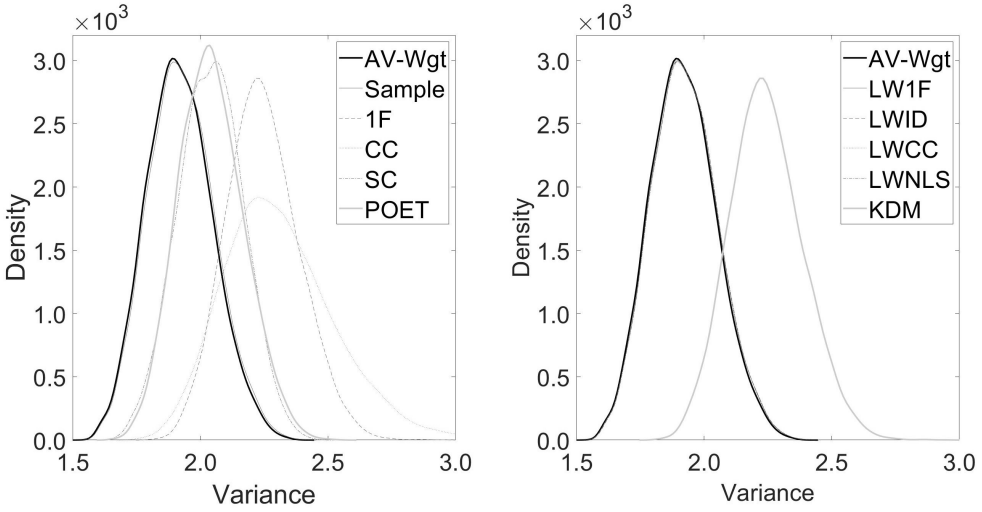
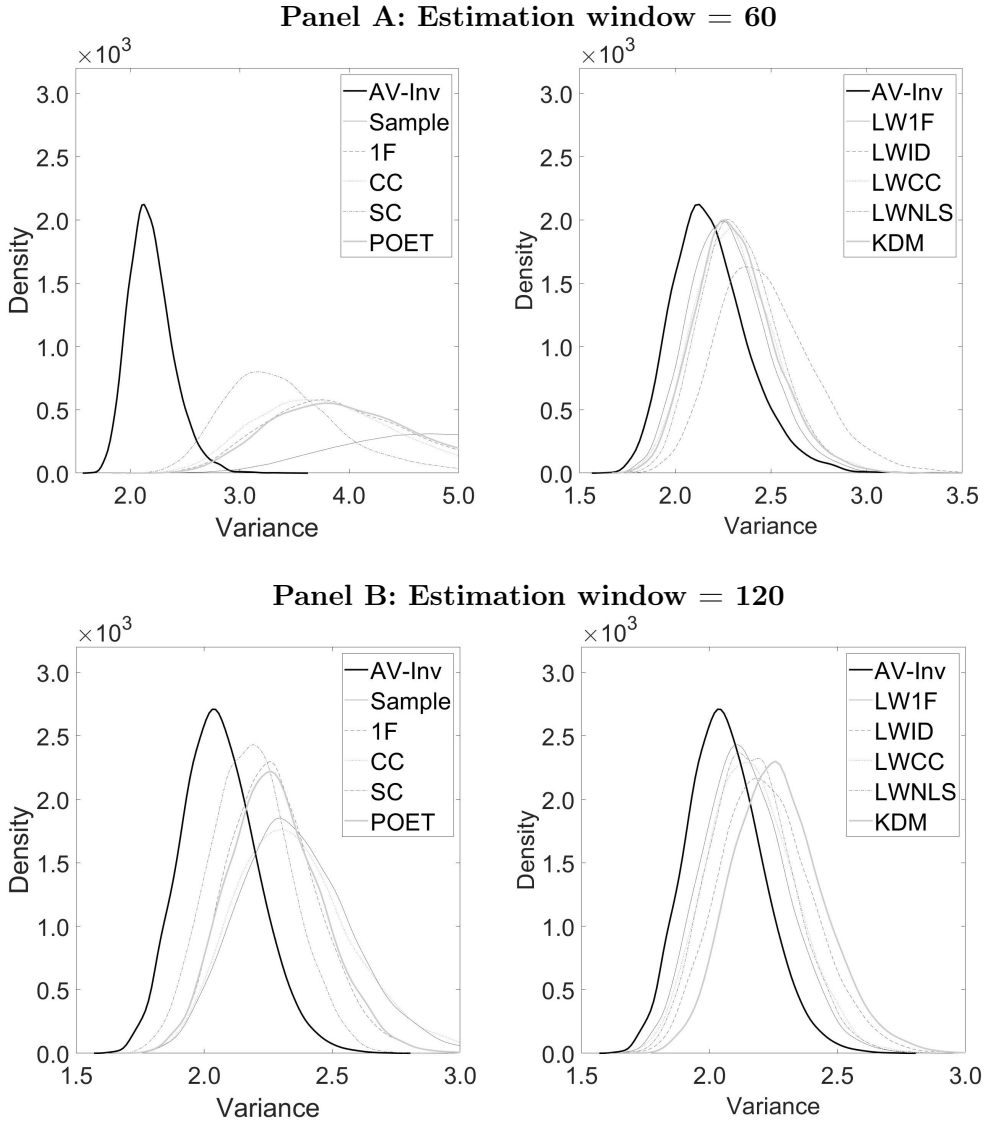
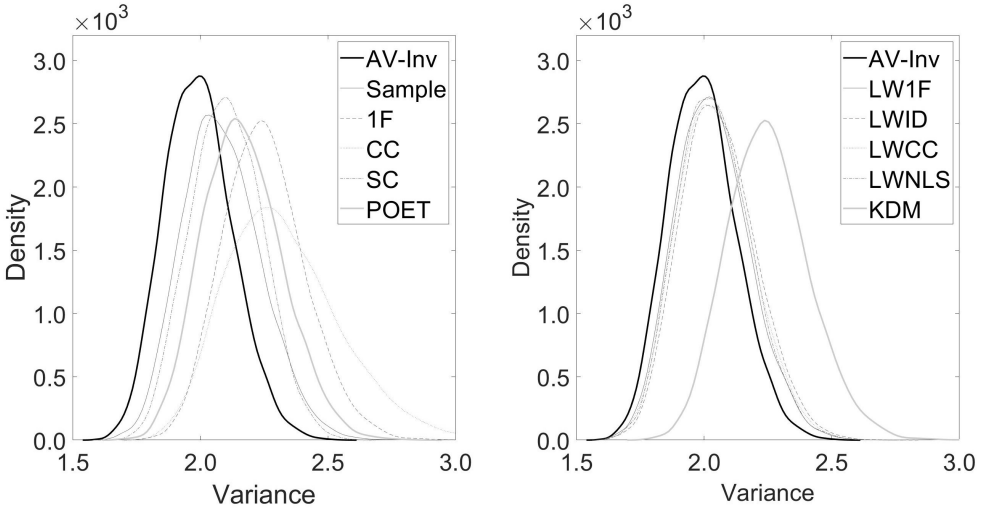


Figure B.7: Kernel densities of the out-of-sample variances of AV-Inv compared to established strategies and benchmarks

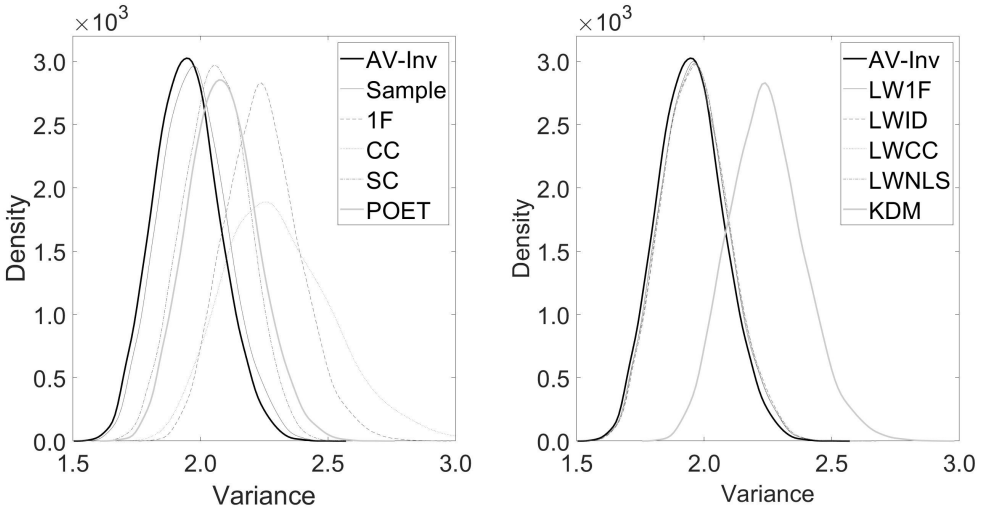
This figure plots the kernel densities of the out-of-sample variance of our averaging rule on the inverse covariance matrix (AV-Inv) level over 10,000 simulation runs in comparison to the single strategies (left column) and the benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The out-of-sample variances are scaled by 10e+03. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 3. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

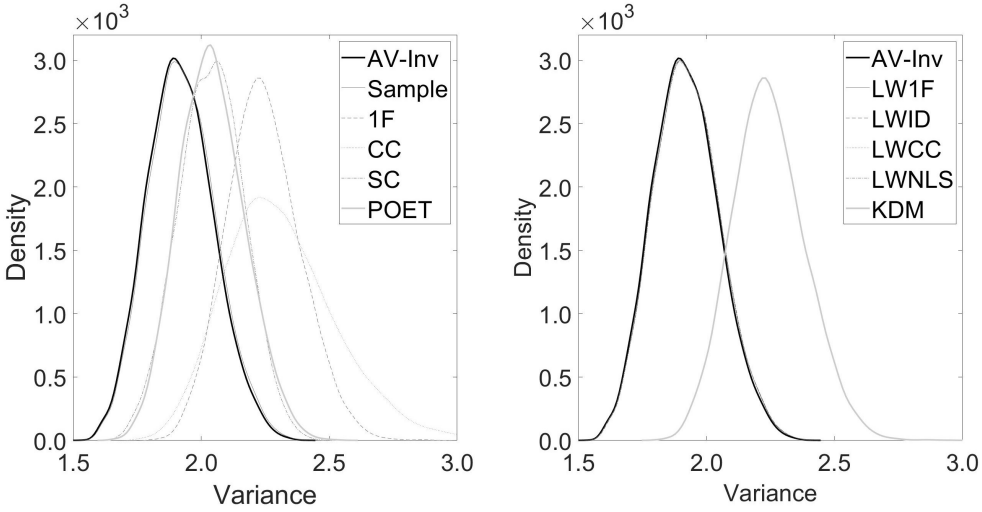
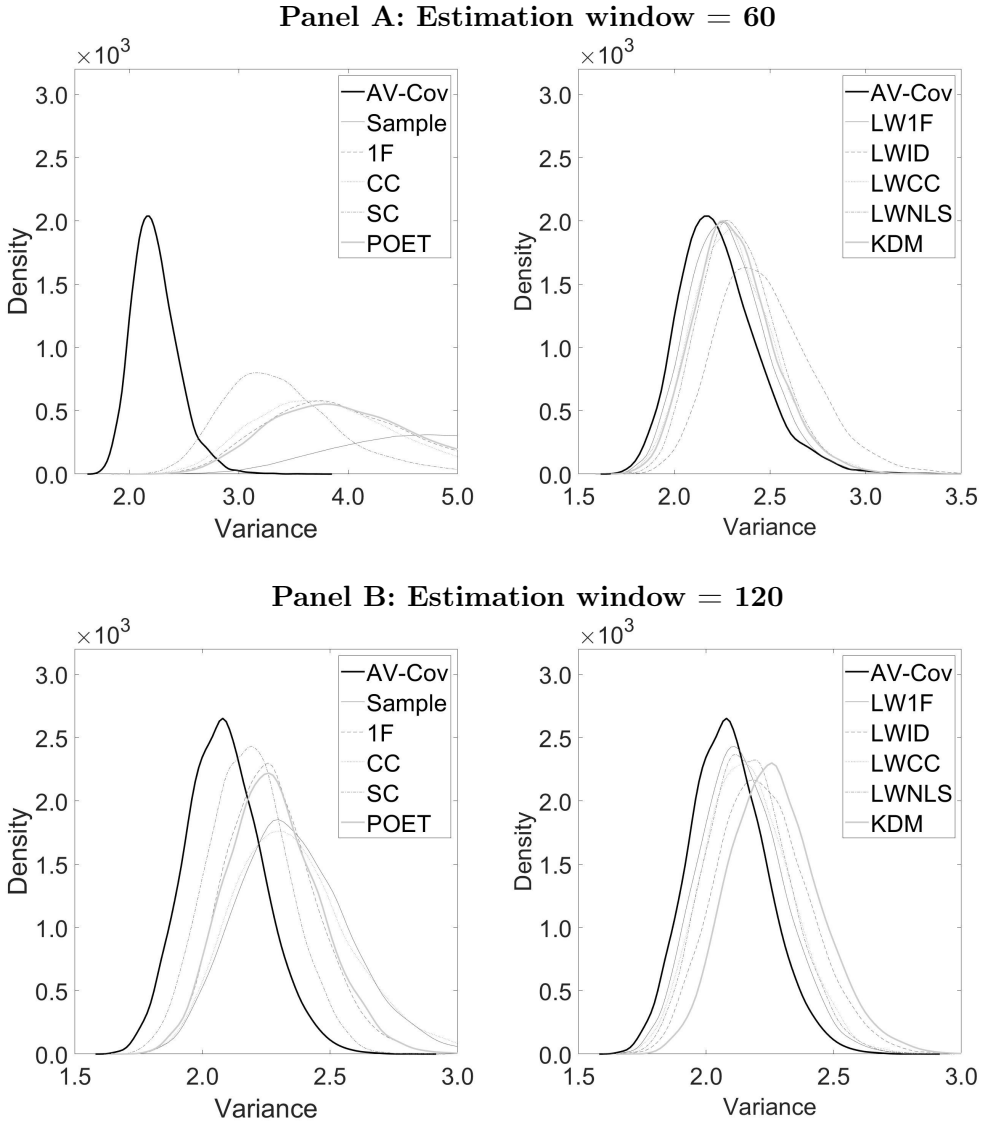
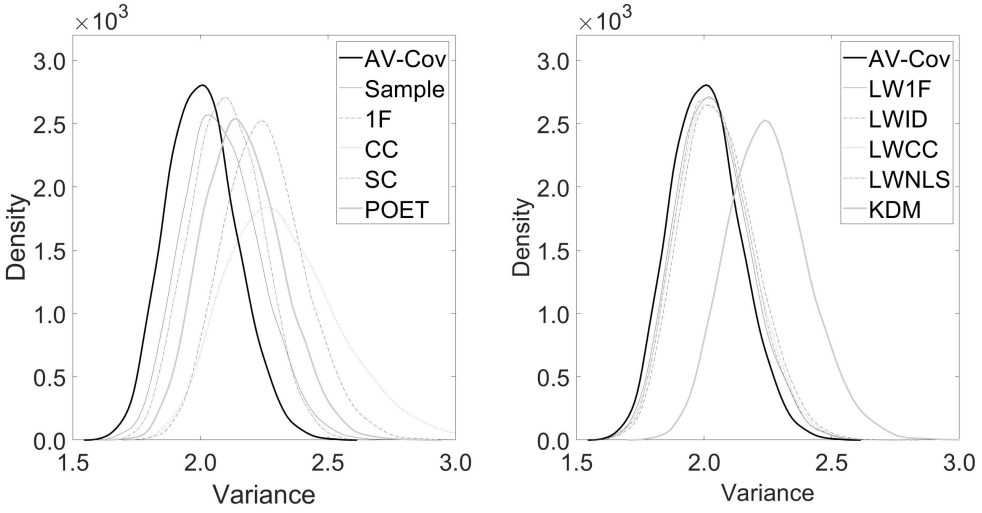


Figure B.8: Kernel densities of the out-of-sample variances of AV-Cov compared to established strategies and benchmarks

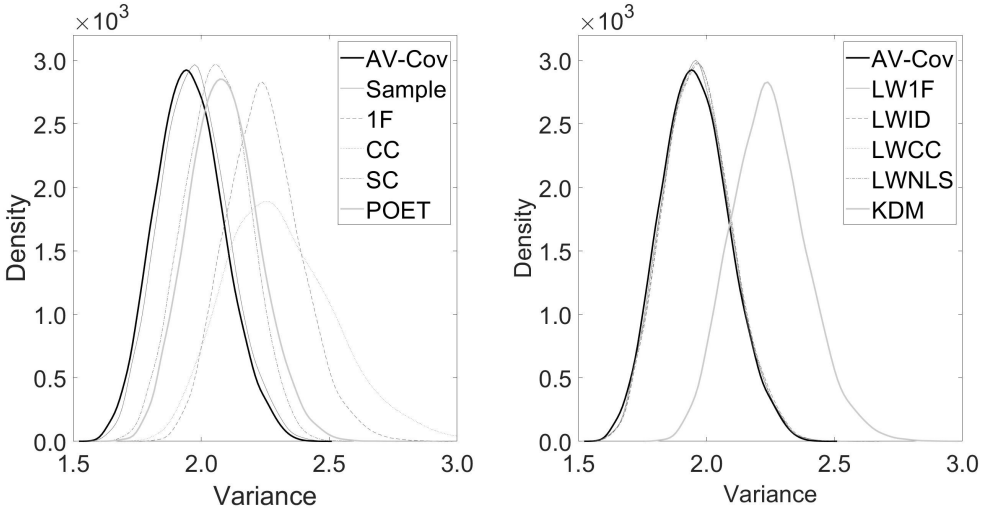
This figure plots the kernel densities of the out-of-sample variance of our averaging rule on the covariance matrix (AV-Cov) level over 10,000 simulation runs in comparison to the single strategies (left column) and the benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The out-of-sample variances are scaled by 10e+03. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 3. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.



Panel C: Estimation window = 2460



Panel D: Estimation window = 480



Panel E: Estimation window = 960

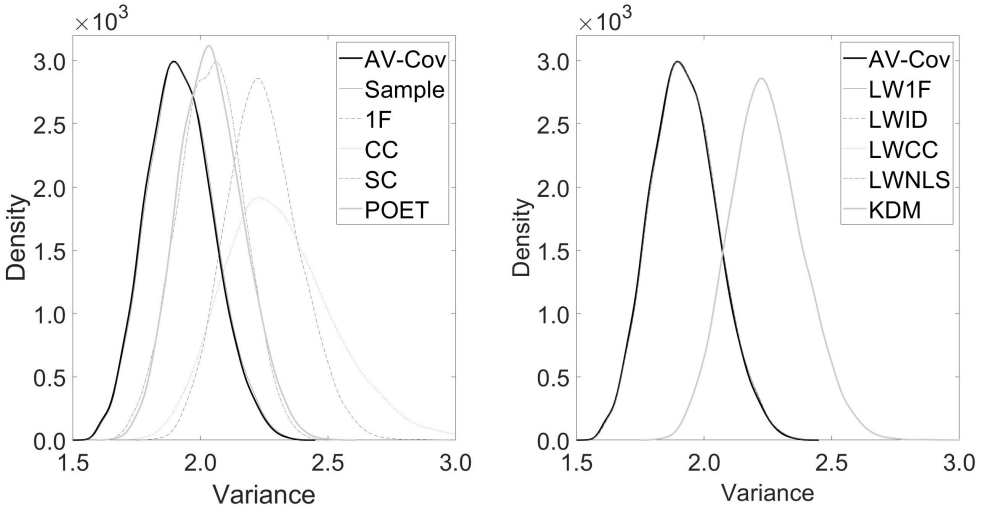
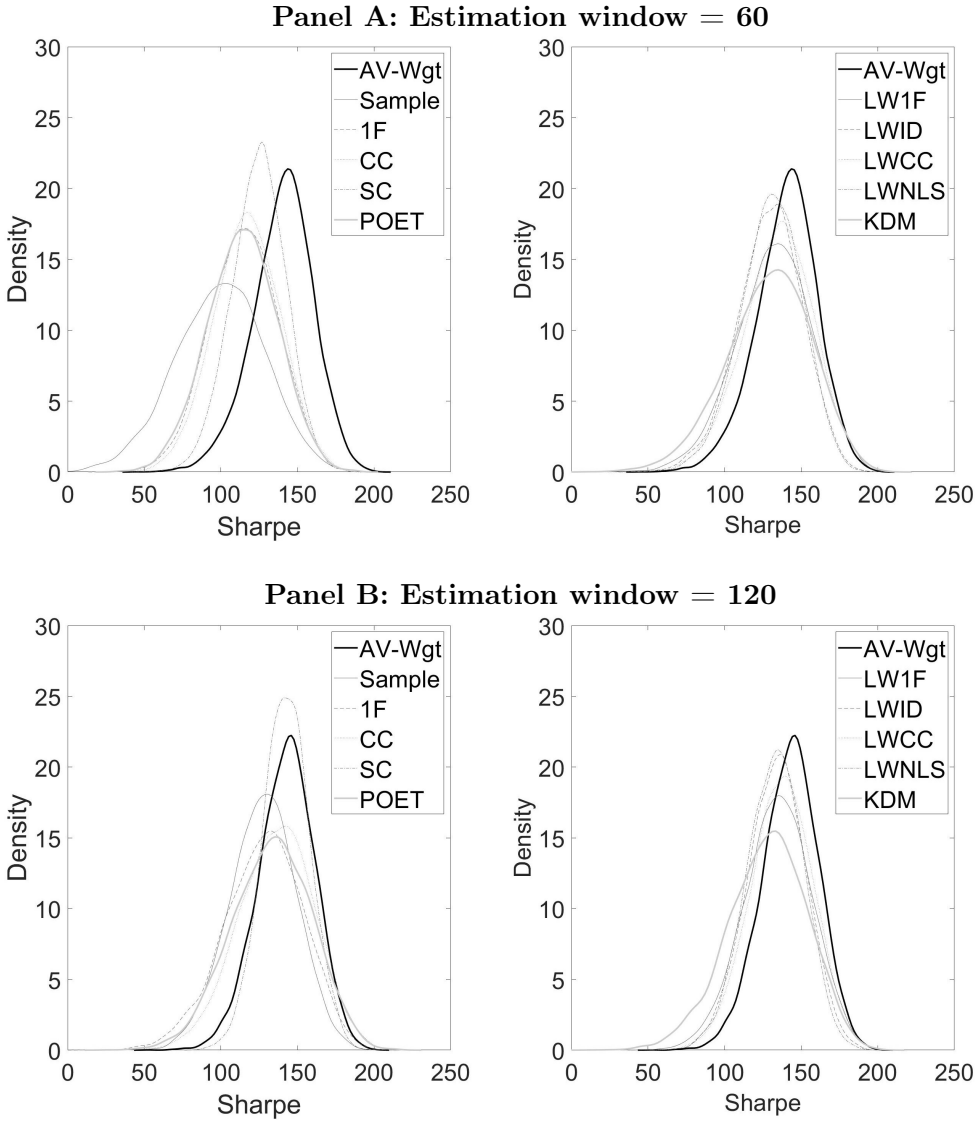
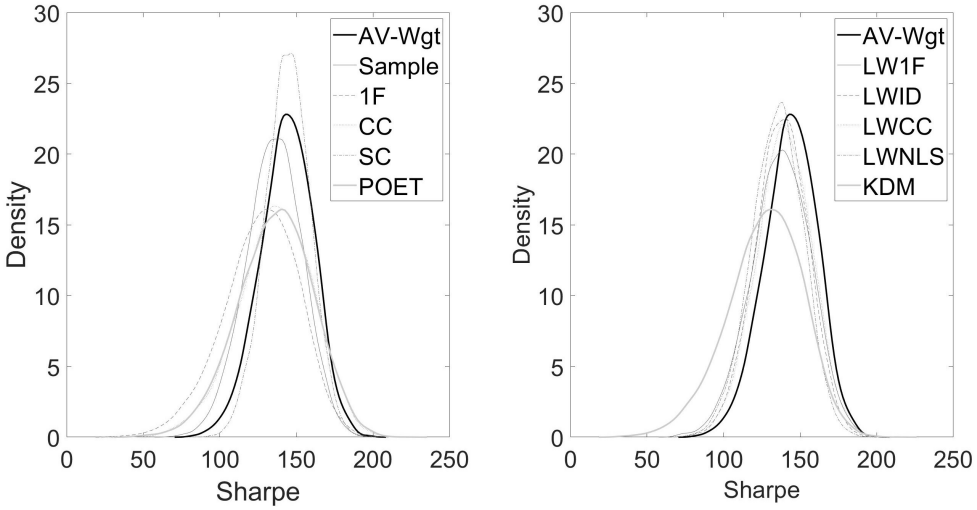


Figure B.9: Kernel densities of the out-of-sample Sharpe ratio of AV-Wgt compared to established strategies and benchmarks

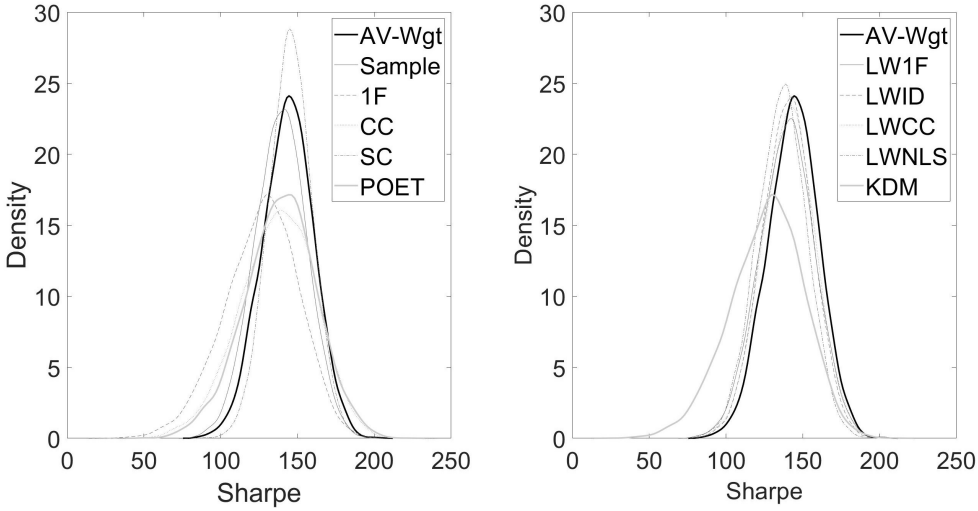
This figure plots the kernel densities of the out-of-sample Sharpe ratio of our averaging rule on the portfolio weight (AV-Wgt) level over 10,000 simulation runs in comparison to the single strategies (left column) and the benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The out-of-sample Sharpe ratios are scaled by 10e+03. The results for AV-Wgt are based on averaging over the $M = 6$ strategies from Panel B in Table 3. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

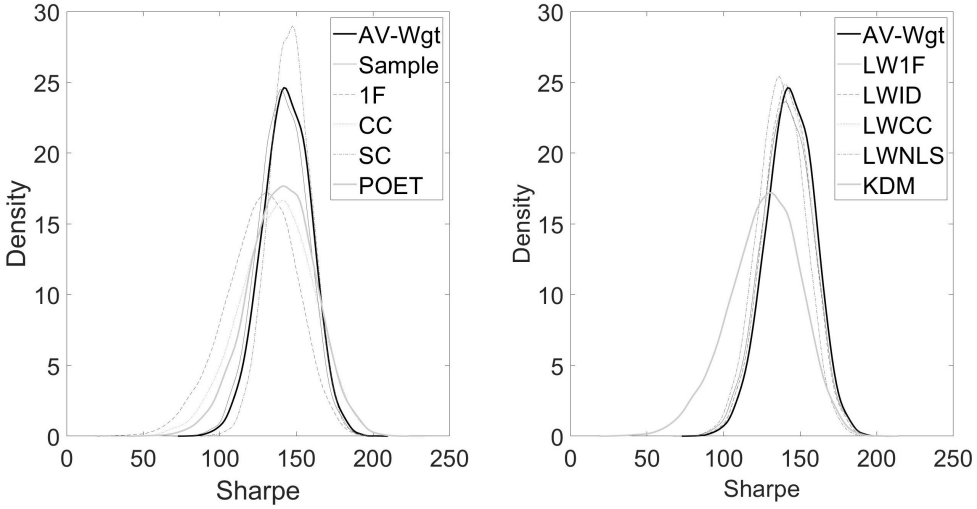
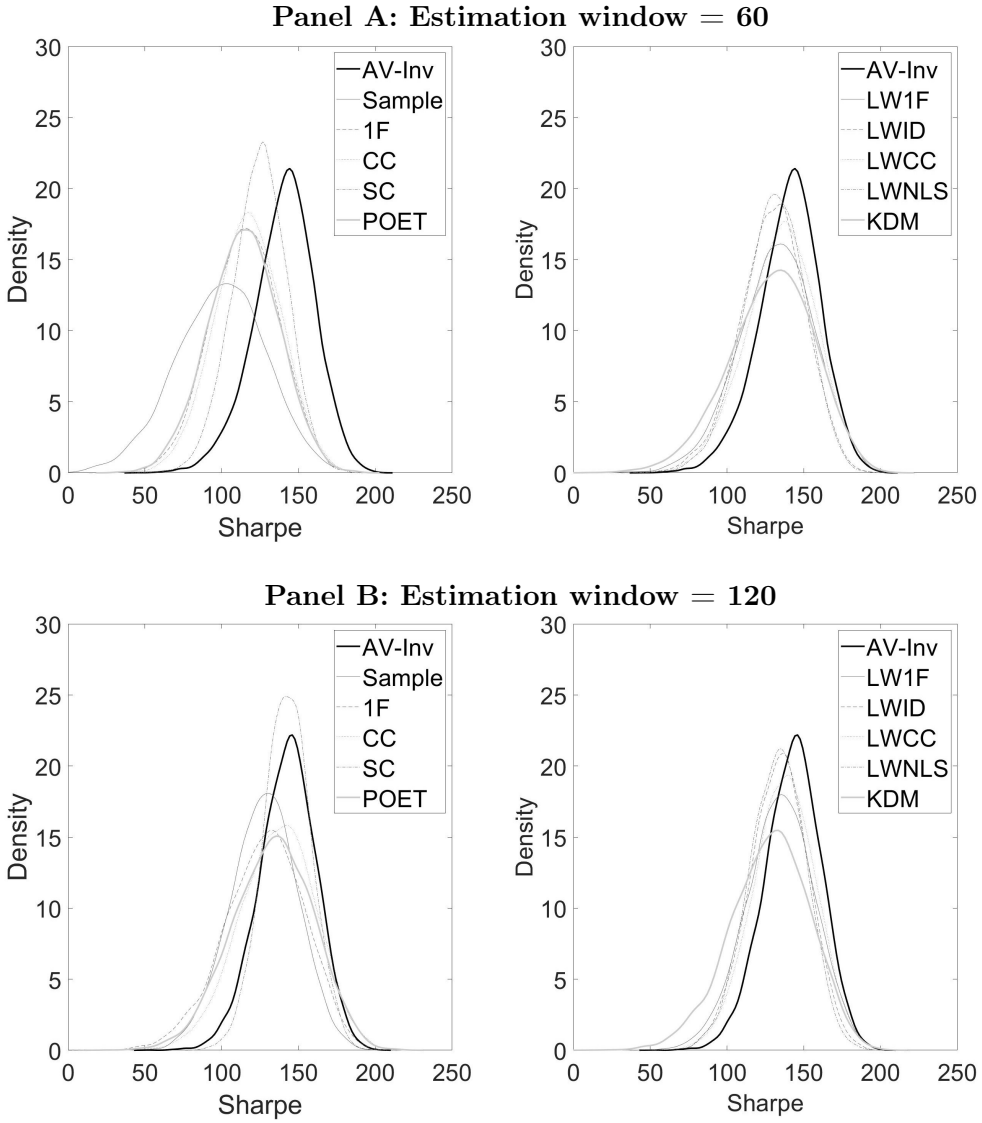
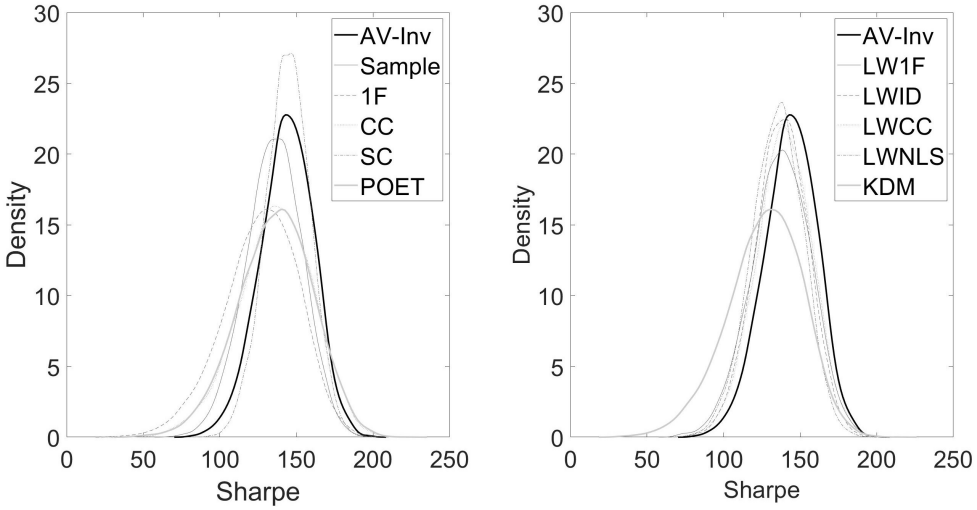


Figure B.10: Kernel densities of the out-of-sample Sharpe ratio of AV-Inv compared to established strategies and benchmarks

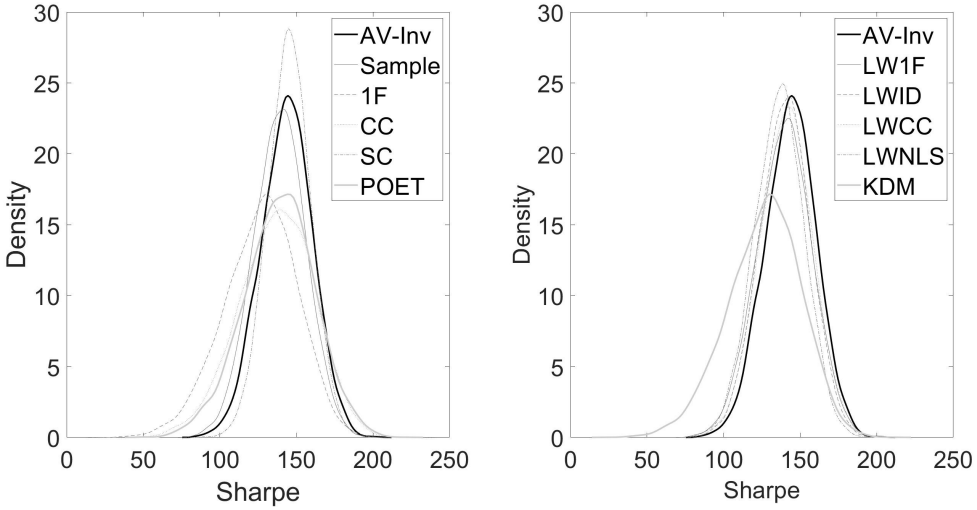
This figure plots the kernel densities of the out-of-sample Sharpe ratio of our averaging rule on the inverse covariance matrix (AV-Inv) level over 10,000 simulation runs in comparison to the single strategies (left column) and the benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The out-of-sample Sharpe ratios are scaled by 10e+03. The results for AV-Inv are based on averaging over the $M = 6$ strategies from Panel B in Table 3. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.



Panel C: Estimation window = 240



Panel D: Estimation window = 480



Panel E: Estimation window = 960

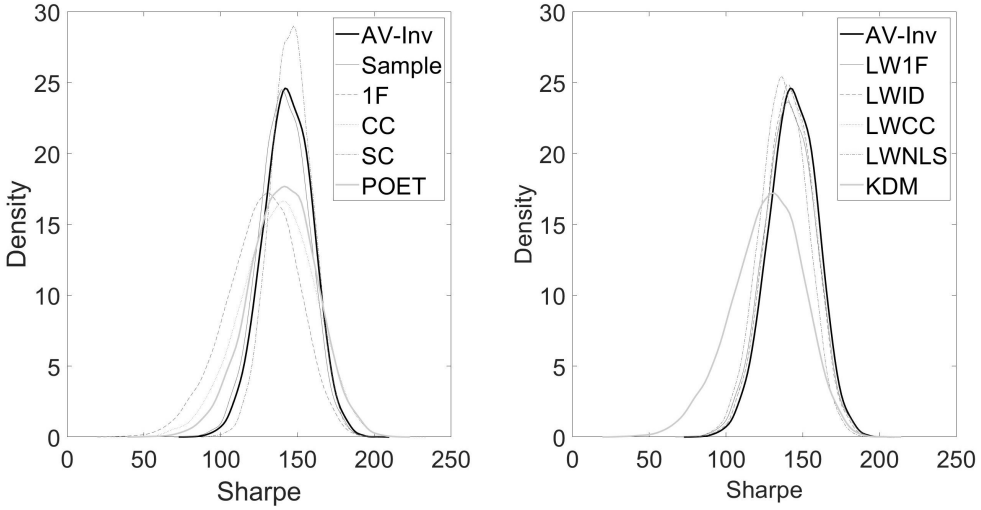
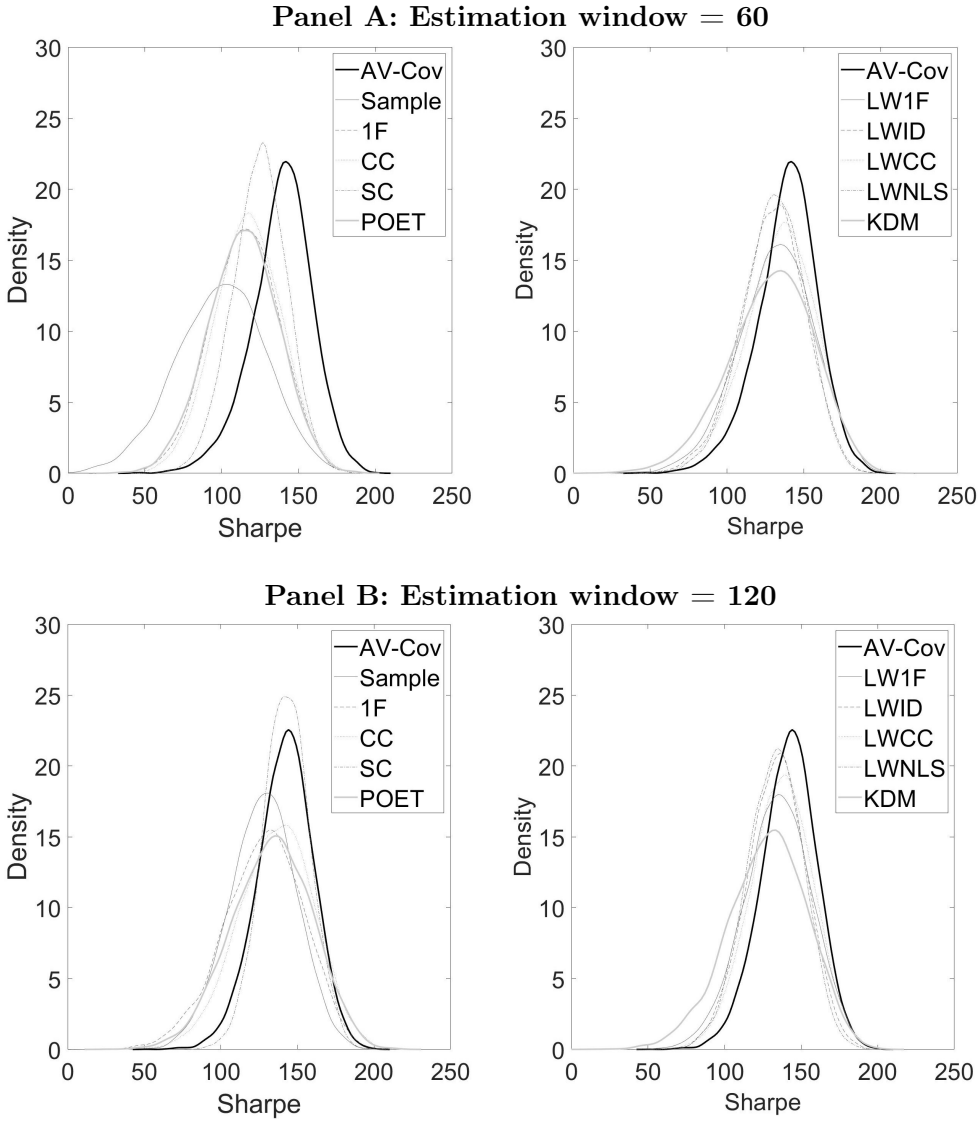
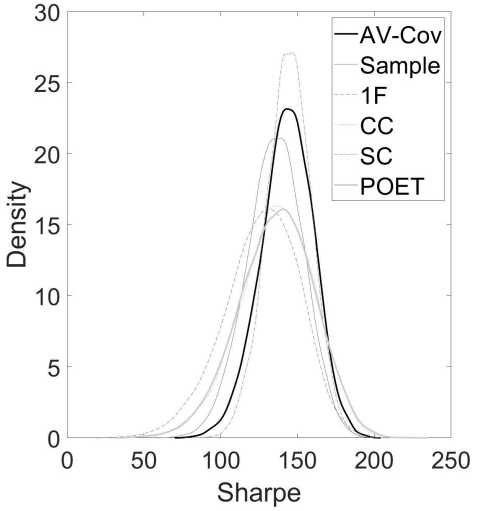


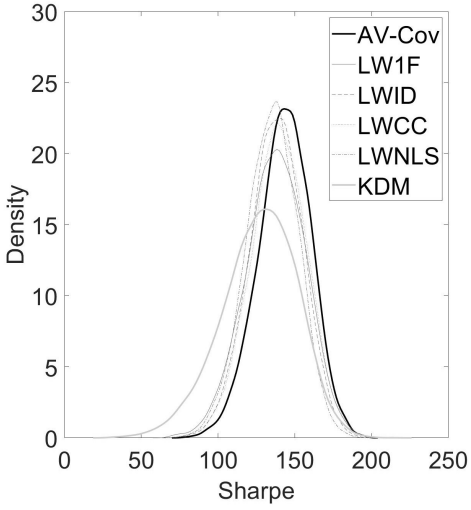
Figure B.11: Kernel densities of the out-of-sample Sharpe ratio of AV-Cov compared to established strategies and benchmarks

This figure plots the kernel densities of the out-of-sample Sharpe ratio of our averaging rule on the covariance matrix (AV-Cov) level over 10,000 simulation runs in comparison to the single strategies (left column) and the benchmarks (right column) for the considered estimation windows of $\tau = \{60, 120, 240, 480, 960\}$. The out-of-sample Sharpe ratios are scaled by 10e+03. The results for AV-Cov are based on averaging over the $M = 6$ strategies from Panel B in Table 3. For illustration purpose only, we exclude the ID strategy because it distorts the visualization of the remaining strategies.

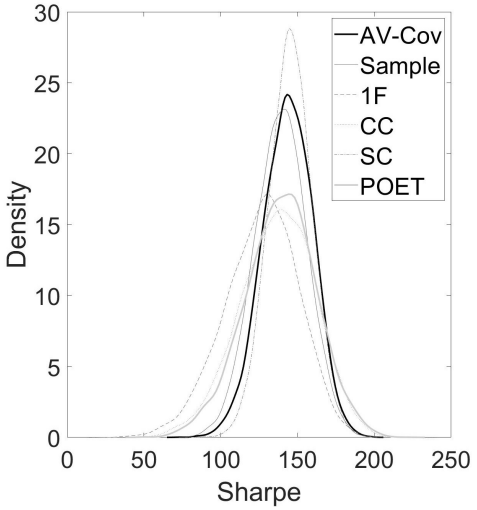


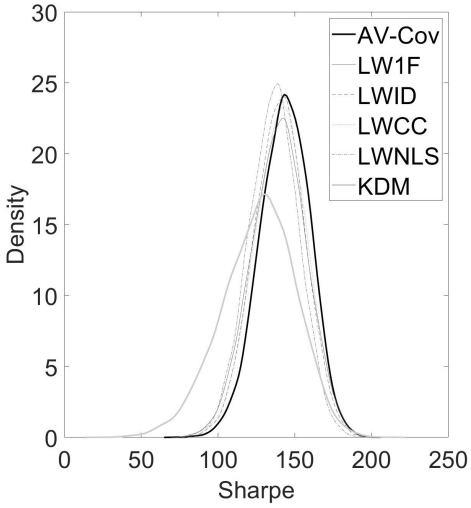
Panel C: Estimation window = 240



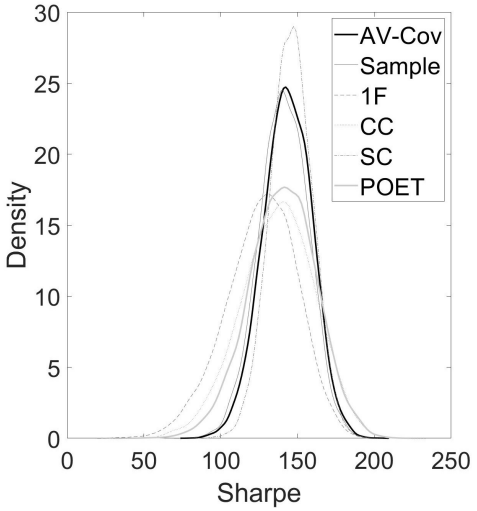


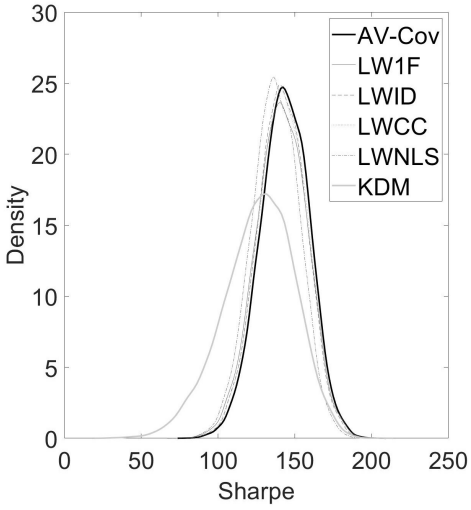
Panel D: Estimation window = 480





Panel E: Estimation window = 960





C Results on monthly data frequency

Following Ledoit and Wolf (2017), we change the excess return frequency from daily to monthly. While keeping the out-of-sample period the same from 06/1976-06/2019, we use the longer available history to estimate the covariance matrix over $T = 120$ most recent months, i.e. 10 years of data, as common in the literature. The data sets as described in Table C.16 are the same for the Fama-French factor-mimicking and industry portfolios. We reduce the CRSP data set to the 30 largest stocks, representing the Dow Jones Industrial Average index. We thus ensure that the number of assets N remains smaller than the available monthly observations τ. In July of each year we select the 30 largest stocks in terms of market capitalization from all NYSE, AMEX, and NASDAQ stocks in the Center for Research in Security Prices (CRSP) database that fulfill the following specific criteria. We filter out stocks with a price of less than \$5 as for daily observations, as well as stocks with missing excess returns in the 120 preceding months or the 12 months subsequent to the selection date. The stocks constitute the investment universe for one year. For our rolling-sample procedure, we set the estimation window to $\tau = 120$ observations, corresponding to ten years of monthly data. Starting in June 1966, we estimate the portfolio weights of each strategy using only the information in the estimation window, which comprises the most recent τ excess returns. We hold the estimated portfolio weights constant for a one-month holding period, and save the corresponding out-of-sample excess returns. We then move the estimation window forward by a month and repeat the aforementioned procedure over the entire sample period. The specification is thus similar to the simulation study with an estimation window of $\tau = 120$. The remaining details are as described in the main text.

We find that the results on a monthly data frequency are qualitatively similar to the ones presented in Section 4 for daily observations. Our averaging rule achieves on all averaging levels comparable variances and Sharpe ratios. Comparing AV-Wgt to the benchmarks in Panel C of Table C.17, we find that AV-Wgt achieves on all five data sets lower variances than LWNLS, being significantly lower for four out of five. The

variance is also lower or similar to the other shrinkage estimators LW1F, LWID, LWCC, or KDM. In terms of out-of-sample risk-adjusted performance, AV-Wgt achieves consistently higher Sharpe ratios compared to the ID strategy in Panel B of Table C.18. In most cases, the performance is also superior, or at least similar, to the benchmarks in Panel C. The monthly average turnover for AV-Wgt as shown in Table C.19 is larger than for the benchmarks for the smaller 6FF or 10Ind portfolios but tend to be lower for the 25FF or 30Ind portfolios and 30 stocks (STOCK30). As with daily returns, the short interest of AV-Wgt in Table C.20 is the smallest of all strategies (excluding ID and SC) for the 30- industry portfolio and the stocks, and on a similar level as LW1F and LWNLS for the remaining portfolios. Summing up, our averaging rule on monthly data constitutes a further improvement over the existing strategies and shrinkage approaches similar as for daily data.

Table C.16: List of the empirical data sets for monthly data

This table lists the data sets for the empirical evaluation of our proposed averaging rule in comparison to existing portfolio strategies, their abbreviations, the number of assets in each data set, and the data sources. All data sets span the period 06/1966 - 06/2019, constitute monthly data, and apply in the case of portfolio data the value weighting scheme to the respective constituents. Data from Kenneth French is taken from his website (`http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`) and represents different cuts of the U.S. stock market. The STOCK30 data set contains the 30 largest single stocks in terms of market capitalization in July of every year after filtering out stocks that have a price of less than \$5, or missing returns in the preceding 120 and subsequent 12 trading months to the selection date. All stock prices are taken from the Center of Research in Security Prices (CRSP).

#	Data set	Abbreviation	N	Source
1	6 Fama and French portfolios of firms sorted by size and book-to-market	6FF	6	K. French
2	25 Fama and French portfolios of firms sorted by size and book-to-market	25FF	25	K. French
3	10 industry portfolios representing the U.S. stock market	10Ind	10	K. French
4	30 industry portfolios representing the U.S. stock market	30Ind	30	K. French
5	30 Stocks with the largest market capitalization	STOCK30	30	CRSP

Table C.17: Monthly out-of-sample variances in empirical data sets

This table reports the monthly out-of-sample variances of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panel B and C, as well as for the data sets, are reported in Table 3 and 9, respectively. We report statistical significance for the null hypothesis wherein the (log) expected out-of-sample portfolio variance of the averaging rule on the portfolio weight level, AV-Wgt, is greater than that of the non-linear shrinkage approach of Ledoit and Wolf (2017), LWNLS: i: $H_0 : log(\hat{\sigma}^2_{AV-Wgt}) - log(\hat{\sigma}^2_{LWNLS}) \leq 0$. We follow Ledoit and Wolf (2011) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt	13.9847	12.2703	11.0577	10.6377	12.8335
AV-Inv	13.9706	12.2774	11.0736	10.6116	12.8364
AV-Cov	14.2756	12.2592	11.0897	10.6961	12.8539
Panel B: Single strategies					
Sample	14.3117	13.7620	11.4345	12.6086	15.2445
1F	20.5461	29.6997	12.5939	13.0024	15.4703
CC	19.3826	32.6633	13.5744	16.4668	15.9903
ID	22.1005	23.7947	16.5779	20.7166	17.7152
SC	17.5683	16.6497	11.3624	11.5421	12.7908
POET	14.9073	13.2813	11.6811	11.7436	14.0681
Panel C: Benchmarks					
LW1F	14.4877	12.7579	11.2287	10.9466	13.3622
LWID	14.3537	12.0929	10.9141	10.9056	13.3110
LWCC	17.6658	19.2178	11.3035	11.3676	12.7354
LWNLS	14.3093**	12.7787**	11.1915	11.1344**	13.3429**
KDM	20.5461	29.6988	12.5939	13.0024	15.4703

Table C.18: Monthly out-of-sample Sharpe ratios in empirical data sets

This table reports the monthly out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panel B and C, as well as for the data sets, are reported in Table 3 and 9, respectively. We use the 30-day T-bill rate as the risk-free rate. We report statistical significance for the null hypothesis wherein the Sharpe ratio of the averaging rule on the portfolio weight level, AV-Wgt, is greater than that of the 1/N portfolio, corresponding to the ID strategy: $H_0 : SR_{AV-Wgt} - SR_{ID} \leq 0$. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt	0.2668	0.2632	0.2035	0.1872	0.1667
AV-Inv	0.2681	0.2636	0.2030	0.1870	0.1672
AV-Cov	0.2664	0.2673	0.2005	0.1850	0.1634
Panel B: Single strategies					
Sample	0.2799	0.2925	0.2023	0.1609	0.1103
1F	0.1571	0.1608	0.2124	0.1863	0.1748
CC	0.1608	0.1456	0.1936	0.1811	0.1634
ID	0.1631***	0.1656***	0.1681	0.1535	0.1578
SC	0.1775	0.1816	0.1964	0.1824	0.1739
POET	0.3135	0.2711	0.1818	0.1733	0.1829
Panel C: Benchmarks					
LW1F	0.2678	0.2825	0.2051	0.1830	0.1402
LWID	0.2308	0.2754	0.2090	0.1983	0.1434
LWCC	0.1887	0.2152	0.2028	0.1912	0.1555
LWNLS	0.2600	0.2522	0.2013	0.1912	0.1584
KDM	0.1571	0.1608	0.2124	0.1863	0.1748

Table C.19: Average monthly turnover in empirical data sets based on monthly observations

This table reports the average monthly turnover of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the portfolio strategies in Panels B and C, as well as for the data sets, are reported in Table 3 and 9, respectively. Turnover is measured as the average percentage of total wealth traded in each period. The reported numbers are in percentages.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt	18.23	40.99	11.02	20.52	40.27
AV-Inv	18.47	41.57	10.99	20.28	40.75
AV-Cov	23.865	39.87	11.56	21.35	40.75
Panel B: Single strategies					
Sample	21.88	79.46	15.47	45.92	29.74
1F	12.36	24.54	9.25	14.79	12.69
CC	10.32	23.23	7.54	13.96	10.64
ID	1.47	1.67	2.26	2.85	4.36
SC	4.09	6.96	4.87	6.74	8.98
POET	27.22	43.96	16.37	23.96	17.85
Panel C: Benchmarks					
LW1F	19.73	55.67	13.60	29.33	19.06
LWID	9.03	32.09	10.41	26.35	18.92
LWCC	10.89	28.60	9.63	22.20	14.26
LWNLS	19.56	54.99	13.78	28.95	19.14
KDM	12.36	24.54	9.25	14.79	12.69

Table C.20: Average monthly short interest in empirical data sets

This table reports the average monthly short interest of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK30 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panel B and C, as well as for the data sets, are reported in Table 3 and 9, respectively. Short interest is measured by the average amount of wealth that is held in short positions. The reported numbers are in percentages.

	6FF	25FF	10Ind	30Ind	STOCK30
Panel A: Averaging rule across three levels					
AV-Wgt	108.23	195.87	36.47	57.28	27.98
AV-Inv	111.16	198.87	35.54	56.94	27.93
AV-Cov	107.66	199.12	35.76	60.05	29.03
Panel B: Single strategies					
Sample	140.33	328.78	64.29	159.97	85.36
1F	88.05	158.19	53.68	82.89	44.79
CC	76.54	144.93	41.03	77.17	38.58
ID	0.00	0.00	0.00	0.00	0.00
SC	0.00	0.00	0.00	0.00	0.00
POET	193.73	269.34	70.25	96.15	53.75
Panel C: Benchmarks					
LW1F	123.08	260.15	59.75	113.88	55.48
LWID	63.08	171.05	41.86	97.42	51.26
LWCC	75.19	159.06	38.56	86.43	40.78
LWNLS	116.09	253.59	56.81	105.21	51.73
KDM	88.05	158.18	53.68	82.89	44.79

D Portfolio strategies and averaging with short-sale constraint

In order to investigate whether our averaging rule can be used for long-only investment manager or under regulatory regimes that prohibit short-sales, we impose a short-sale constraint on all strategies that are considered in our averaging rule and benchmarks.[23]. That means that in addition to the ID and SC strategy, also the Sample, 1F, and POET strategy have now non-negative portfolio weights, i.e. the short interest is zero. The results for the out-of-sample variance, Sharpe ratio, and turnover are shown in Tables D.21-D.23.

Imposing the short-sale restriction results in higher variance for all portfolio strategies, except for 1F. The variance of our rule is between 11 and 85% higher on the 10IND and 25FF data set, respectively. In comparison to the benchmarks, our rule achieves, on average, lower portfolio variance. Except for the 25FF data set, the variance of AV-Wgt is lower than for LWNLS across the remaining four data sets, being significantly lower on the 30IND and STOCK500 data set.

The constraint for non-negative portfolio weights leads to 3 to 34% lower Sharpe ratios for the STOCK500 and 25FF data set, respectively. Only for the 30IND portfolio, the Sharpe ratio is 1% higher than without the short-sale constraint. Our rule also compares favorably to all benchmarks across all data sets. Additionally, the out-of-sample performance of AV-Wgt is better than for the 1/N portfolio across all data sets, being significantly higher on four out five.

The turnover is significantly reduced by 50 to 97% for the STOCK500 and 6FF data sets. The reduction in turnover applies to all portfolio strategies and is now comparable to the strategies with natural short-sale constraints, i.e. the ID and SC strategy.

[23] We exclude strategies based on the inverse covariance matrix, i.e. AV-Inv and KDM, as a short-sale restriction is not possible.

Table D.21: Daily out-of-sample variances with short-sale constraints

This table reports the daily out-of-sample variances of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C, after imposing short-sale constraints on the single strategies. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The reported variances are scaled by 10e+03. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. We report statistical significance for the null hypothesis wherein the log out-of-sample variance of AV-Wgt is greater than that of the non-linear shrinkage approach of Ledoit and Wolf (2017), LWNLS: $H_0 : log(\hat{\sigma}^2_{AV-Wgt}) - log(\hat{\sigma}^2_{LWNLS}) \leq 0$. We follow Ledoit and Wolf (2011) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5, and 10% level is denoted by , , and , respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across to levels					
AV-Wgt	0.7705	0.7051	0.5807	0.5362	0.4176
AV-Cov	0.7696	0.7086	0.5806	0.5420	0.4145
Panel B: Single strategies					
Sample	0.7700	0.7072	0.5812	0.5415	0.4192
1F	0.7774	0.7170	0.5795	0.5442	0.5205
CC	0.7775	0.7128	0.5877	0.5676	0.5147
ID	1.0892	1.0908	0.9686	1.0537	1.1287
SC	0.7700	0.7072	0.5812	0.5415	0.4192
POET	0.7697	0.7080	0.5803	0.5419	0.4212
Panel C: Benchmarks					
LW1F	0.7702	0.7073	0.5808	0.5412	0.4173
LWID	0.7711	0.7086	0.5838	0.5425	0.4176
LWCC	0.7705	0.7064	0.5811	0.5427	0.4158
LWNLS	0.7695	0.7072	0.5809	0.5405**	0.5771***

Table D.22: Daily out-of-sample Sharpe ratios with short-sale constraints

This table reports the daily out-of-sample Sharpe ratios of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C, after imposing short-sale constraints on the single strategies. The results are shown for to the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. We use the 30-day T-bill rate as the risk-free rate. We report statistical significance for the null hypothesis wherein the Sharpe ratio of AV-Wgt is greater than that of the 1/N portfolio, corresponding to the ID strategy: $H_0 : SR_{AV-Wgt} - SR_{ID} \leq 0$. We follow Ledoit and Wolf (2008) and use their proposed bootstrap procedure, using a block length of 5 and 1,000 iterations. Statistical significance at the 1, 5 and 10% level is denoted by ***, **, and *, respectively.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across two levels					
AV-Wgt	0.0511	0.0316	0.0421	0.0404	0.0496
AV-Cov	0.0507	0.0304	0.0416	0.0398	0.0496
Panel B: Single strategies					
Sample	0.0507	0.0300	0.0418	0.0404	0.0503
1F	0.0504	0.0332	0.0413	0.0395	0.0429
CC	0.0505	0.0336	0.0411	0.0393	0.0449
ID	0.0343***	0.0189***	0.0332*	0.0320	0.0363**
SC	0.0507	0.0300	0.0418	0.0404	0.0503
POET	0.0507	0.0307	0.0419	0.0398	0.0476
Panel C: Benchmarks					
LW1F	0.0508	0.0302	0.0419	0.0404	0.0494
LWID	0.0510	0.0303	0.0416	0.0403	0.0499
LWCC	0.0511	0.0309	0.0419	0.0405	0.0497
LWNLS	0.0507	0.0301	0.0419	0.0405	0.0487

Table D.23: Average monthly turnover with short-sale constraints

This table reports the average monthly turnover of our averaging rule on the portfolio weight (AV-Wgt), the inverse (AV-Inv), and the covariance (AV-Cov) level in Panel A, of the single strategies in Panel B, and of the benchmarks in Panel C, after imposing short-sale constraints on the single strategies. The results are shown for the Fama-French factor-mimicking and industry portfolios, as well as the STOCK500 data set. The out-of-sample period is from 06/1976 - 06/2019. The abbreviations for the strategies in Panels B and C, as well as for the data sets, are reported in Tables 3 and 9, respectively. Turnover is measured as the average percentage of total wealth traded in each period. The reported numbers are in percentages.

	6FF	25FF	10Ind	30Ind	STOCK500
Panel A: Averaging rule across two levels					
FAWgt	0.46	0.90	1.08	1.84	5.38
FACov	0.47	0.95	1.11	1.85	5.39
Panel B: Single strategies					
Sample	0.46	0.94	1.10	1.86	3.65
1F	0.42	0.84	1.05	1.75	3.00
CC	0.44	0.76	0.97	1.43	2.49
ID	1.48	1.68	2.32	2.87	5.45
SC	0.46	0.94	1.10	1.86	3.66
POET	0.44	0.96	1.08	1.82	3.48
Panel C: Benchmarks					
LW1F	0.46	0.94	1.10	1.85	3.54
LWID	0.51	0.98	1.19	1.94	3.72
LWCC	0.46	0.91	1.08	1.76	3.26
LWNLS	0.46	0.95	1.11	1.90	3.92

References

Abarbanell, J. S., & Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance*, *47*(3), 1181–1207. doi: https://doi.org/10.2307/2328982

Aharoni, G., Grundy, B., & Zeng, Q. (2013). Stock returns and the Miller Modigliani valuation formula: Revisiting the Fama French analysis. *Journal of Financial Economics*, *110*(2), 347–357. doi: https://doi.org/10.1016/j.jfineco.2013.08.003

Ahern, K. R., & Sosyura, D. (2015). Rumor has it: Sensationalism in financial media. *Review of Financial Studies*, *28*(7), 2050–2093. doi: https://doi.org/10.1093/rfs/hhv006

Ang, A., Chen, J., & Xing, Y. (2006). Downside risk. *Review of Financial Studies*, *19*(4), 1191–1239. doi: https://doi.org/10.1093/rfs/hhj035

Antoniou, C., Doukas, J. A., & Subrahmanyam, A. (2016). Investor sentiment, beta, and the cost of equity capital. *Management Science*, *62*(2), 347–367. doi: https://doi.org/10.1287/mnsc.2014.2101

Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, *36*(2), 334–357. doi: https://doi.org/10.1016/j.ijforecast.2019.05.010

Audrino, F., & Tetereva, A. (2019). Sentiment spillover effects for US and European companies. *Journal of Banking and Finance*, *106*, 542–567. doi: https://doi.org/10.1016/j.jbankfin.2019.07.022

Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets*, *7*(3), 271–299. doi: https://doi.org/10.1016/j.finmar.2003.11.005

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680. doi: https://doi.org/10.1111/j.1540-6261.2006.00885.x

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151. doi: https://doi.org/10.1257/jep.21.2.129

Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, *104*(2), 272–287. doi: https://doi.org/10.1016/j.jfineco.2011.11.002

Balduzzi, P., & Lynch, A. W. (1999). Transaction costs and predictability: Some utility cost calculations. *Journal of Financial Economics*, *52*(1), 47–78. doi: https://doi.org/10.1016/s0304-405x(99)00004-5

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

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343. doi: https://doi.org/10.1016/S0304-405X(98)00027-0

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance*, *32*(3), 663–682. doi: https://doi.org/10.1111/j.1540-6261.1977.tb01979.x

Becker, G. S., & Mulligan, C. B. (1997). The endogenous determination of time preference. *Quarterly Journal of Economics*, *112*(3), 729–758. doi: https://doi.org/10.1162/003355397555334

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies*, *30*(9), 3009–3047. doi: https://doi.org/10.1093/rfs/hhx031

Borochin, P., & Zhao, Y. (2015). Variation of the implied volatility function and return predictability. *SSRN Electronic Journal*. Retrieved from `https://optionmetrics.com/research/borochin{_}zhao{_}2016/`

Boyer, B., Mitton, T., & Vorkink, K. (2010). Expected idiosyncratic skewness. *Review of Financial Studies*, *23*(1), 169–202. doi: https://doi.org/10.1093/rfs/hhp041

Bramante, R., Petrella, G., & Zappa, D. (2013a). On the interpretation and estimation of the market model R-square. *Electronic Journal of Applied Statistical Analysis*, *6*(1), 57–66. doi: https://doi.org/10.1285/i20705948v6n1p57

Bramante, R., Petrella, G., & Zappa, D. (2013b). On the use of the market model R-square as a measure of stock price efficiency. *Review of Quantitative Finance and Accounting*, *44*(2), 379–391. doi: https://doi.org/10.1007/s11156-013-0410-8

Brealey, R. A., Cooper, I. A., & Kaplanis, E. (2014). The Behaviour of Sentiment-Induced Share Returns: Measurement When Fundamentals are Observable. In J. Guerard (Ed.), *Portfolio construction, measurement, and efficiency* (pp. 291–313). Cham: Springer International Publishing. doi: https://doi.org/10.2139/ssrn.2463113

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *Journal of Finance*, *54*(2), 655–671. doi: https://doi.org/10.1111/0022-1082.00120

Brodie, J., Daubechies, I., De Mol, C., Giannone, D., & Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(30), 12267–12272. doi: https://doi.org/10.1073/pnas.0904287106

Brown, G. W., & Cliff, M. T. (2005). Investor sentiment and asset valuation. *Journal of Business*, *78*(2), 405–440. doi: https://doi.org/10.1086/427633

Bushee, B. J., & Friedman, H. L. (2016). Disclosure standards and the sensitivity of returns to mood. *Review of Financial Studies*, *29*(3), 787–822. doi: https://doi.org/10.1093/rfs/hhv054

Cakici, N. (2015). The five-factor Fama-French model: International evidence. *SSRN Electronic Journal*. doi: https://doi.org/10.2139/ssrn.2601662

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, *52*(1), 57–82. doi: https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Chan, L. K., Karceski, J., & Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, *12*(5), 937–974. doi: https://doi.org/10.1093/rfs/12.5.937

Changsheng, H., & Yongfeng, W. (2012). Investor sentiment and assets valuation. *Systems Engineering Procedia*, *3*, 166–171. doi: https://doi.org/10.1016/j.sepro.2011.11.023

Chen, H., De, P., Hu, Y., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403. doi: https://doi.org/10.1093/rfs/hhu001

Ciccone, S. (2003). Does Analyst Optimism About Future Earnings Distort Stock Prices? *Journal of Behavioral Finance*, *4*(2), 59–64. doi: https://doi.org/10.1207/S15427579JPFM0402_02

Cochrane, J. H. (2009). *Asset pricing (revised edition).* Princeton University Press.

Crone, S. F., & Koeppel, C. (2014). Predicting exchange rates with sentiment indicators: An empirical evaluation using text mining and multilayer perceptrons. In *2014 ieee conference on computational intelligence for financial engineering & economics (cifer)* (pp. 114–121). IEEE. doi: https://doi.org/10.1109/CIFEr.2014.6924062

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*(5), 1461–1499. doi: https://doi.org/10.1111/j.1540-6261.2011.01679.x

Da, Z., Engelberg, J., & Gao, P. (2012). *Internet search and momentum effect.* doi: https://doi.org/10.2139/ssrn.1785924

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, *28*(1), 1–32. doi: https://doi.org/10.1093/rfs/hhu072

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain* (1st ed.). Harper Perennial. Paperback. doi: https://doi.org/10.1001/jama.1995.03520420079045

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, *53*(6), 1839–1885. doi: https://doi.org/10.1111/0022-1082.00077

Daniel, K., & Titman, S. (1997). Evidence on the characteristics of cross sectional variation in stock returns. *Journal of Finance*, *52*(1), 1–33. doi: https://doi.org/10.1111/j.1540-6261.1997.tb03806.x

Daszynska-Zygadlo, K., Szpulak, A., & Szyszka, A. (2014). Investor sentiment, optimism and excess stock market returns. Evidence from emerging markets. *Business and Economic Horizons (BEH)*, *10*(4), 362–373. doi: https://doi.org/10.22004/ag.econ.246045

De Bondt, W. F. (1998). A portrait of the individual investor. *European Economic Review*, *42*(3-5), 831–844. doi: https://doi.org/10.1016/S0014-2921(98)00009-9

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738. doi: https://doi.org/10.1086/261703

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management Science*, *55*(5), 798–812. doi: https://doi.org/10.1287/mnsc.1080.0986

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies*, *22*(5), 1915–1953. doi: https://doi.org/10.1093/rfs/hhm075

DeMiguel, V., Martin-Utrera, A., & Nogales, F. J. (2013). Size matters: Optimal calibration of shrinkage estimators for portfolio selection. *Journal of Banking and Finance*, *37*(8), 3018–3034. doi: https://doi.org/10.1016/j.jbankfin.2013.04.033

DeVault, L., Sias, R., & Starks, L. (2019). Sentiment metrics and investor demand. *Journal of Finance*, *74*(2), 985–1024. doi: https://doi.org/10.1111/jofi.12754

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., & Zoladz, P. R. (2007). The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plasticity*, *2007*, 60803. doi: https://doi.org/10.1155/2007/60803

Dolan, R. J. (2002). Neuroscience and psychology: Emotion, cognition, and behavior. *Science*, *298*, 1191–1194. doi: https://doi.org/10.1126/science.1076358

Elton, E. J., & Gruber, M. J. (1973). Estimating the dependence structure of share prices - implications for portfolio selection. *Journal of Finance*, *28*(5), 1203–1232. doi: https://doi.org/10.1111/j.1540-6261.1973.tb01451.x

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *Journal of Finance*, *66*(1), 67–97. doi: https://doi.org/10.1111/j.1540-6261.2010.01626.x

Fama, E. F., & French, K. R. (1992). The cross section of expected stock returns. *Journal of Finance*, *47*(2), 427–465. doi: https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3–56. doi: https://doi.org/10.1016/0304-405X(93)90023-5

Fama, E. F., & French, K. R. (1995). Size and book-to-market factors in earnings and returns. *Journal of Finance*, *50*(1), 131–155. doi: https://doi.org/10.1111/j.1540-6261.1995.tb05169.x

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. doi: https://doi.org/10.1016/j.jfineco.2014.10.010

Fama, E. F., & French, K. R. (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics*, *123*(3), 441–463. doi: https://doi.org/10.1016/j.jfineco.2016.11.004

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607–636. Retrieved from `https://www.jstor.org/stable/1831028`

Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *75*(4), 603–680. doi: https://doi.org/10.1111/rssb.12016

Ferson, W. E., & Korajczyk, R. A. (2002). Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business*, *68*(3), 309. doi: https://doi.org/10.1086/296667

Finter, P., Niessen-Ruenzi, A., & Ruenzi, S. (2012). The impact of investor sentiment on the German stock market. *Zeitschrift für Betriebswirtschaft*, *82*(2), 133–163. doi: https://doi.org/10.2139/ssrn.1650164

Frazzini, A. (2006). The disposition effect and underreaction to news. *Journal of Finance*, *61*(4), 2017–2046. doi: https://doi.org/10.1111/j.1540-6261.2006.00896.x

Frost, P. A., & Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, *21*(3), 293. doi: https://doi.org/10.2307/2331043

García, D. (2013). Sentiment during recessions. *Journal of Finance*, *68*(3), 1267–1300. doi: https://doi.org/10.1111/jofi.12027

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, *57*(5), 1121. doi: https://doi.org/10.2307/1913625

Griffin, J. M., Hirschey, N. H., & Kelly, P. J. (2011). How important is the financial media in global markets? *Review of Financial Studies*, *24*(12), 3941–3992. doi: https://doi.org/10.1093/rfs/hhr099

Griffin, J. M., Kelly, P. J., & Nardari, F. (2010). Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. *Review of Financial Studies*, *23*(8), 3225–3277. doi: https://doi.org/10.1093/rfs/hhq044

Hansen, B. E. (2008). Least-squares forecast averaging. *Journal of Econometrics*, *146*(2), 342–350. doi: https://doi.org/10.1016/j.jeconom.2008.08.022

Hillert, A., Jacobs, H., & Müller, S. (2014). Media makes momentum. *Review of Financial Studies*, *27*(12), 3467–3501. doi: https://doi.org/10.1093/rfs/hhu061

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

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *Review of Financial Studies*, *28*(3), 791–837. doi: https://doi.org/10.1093/rfs/hhu080

Huynh, T. D., & Smith, D. R. (2017). Stock price reaction to news: The joint effect of tone and attention on momentum. *Journal of Behavioral Finance*, *18*(3), 304–328. doi: https://doi.org/10.1080/15427560.2017.1339190

Jaffe, J., Keim, D. B., & Westerfield, R. (1989). Earnings yields, market values, and stock returns. *Journal of Finance*, *44*(1), 135–148. doi: https://doi.org/10.1111/j.1540-6261.1989.tb02408.x

Jagannathan, R., & Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, *58*(4), 1651–1684. doi: https://doi.org/10.1111/1540-6261.00580

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91. doi: https://doi.org/10.1111/j.1540-6261.1993.tb04702.x

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, *132*(1), 126–149. doi: https://doi.org/10.1016/j.jfineco.2018.10.001

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior and Organization*, *176*, 63–90. doi: https://doi.org/10.1016/j.jebo.2020.03.002

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology*, *45*(1), 20–31. doi: https://doi.org/10.1037/0022-3514.45.1.20

Joseph, K., Babajide Wintoki, M., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, *27*(4), 1116–1127. doi: https://doi.org/10.1016/j.ijforecast.2010.11.001

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. doi: https://doi.org/10.2307/1914185

Kan, R., & Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, *42*(3), 621–656. doi: https://doi.org/10.1017/s0022109000004129

Keynes, J. M. (1936). *The general theory of employment interest and money.* Macmillan Cambridge University Press. doi: https://doi.org/10.1007/978-3-319-70344-2

Kliger, D., & Levy, O. (2003). Mood-induced variation in risk preferences. *Journal of Economic Behavior and Organization*, *52*(4), 573–584. doi: https://doi.org/10.1016/S0167-2681(03)00069-6

Kourtis, A., Dotsis, G., & Markellos, R. N. (2012). Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. *Journal of Banking and Finance*, *36*(9), 2522–2531. doi: https://doi.org/10.1016/j.jbankfin.2012.05.005

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. *Journal of Finance*, *73*(3), 1183–1223. doi: https://doi.org/10.1111/jofi.12612

Laborda, R., & Olmo, J. (2014). Investor sentiment and bond risk premia. *Journal of Financial Markets*, *18*(1), 206–233. doi: https://doi.org/10.1016/j.finmar.2013.05.008

Lakonishok, J., Shleifer, A., & Vishny, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, *49*(5), 1541–1578. doi: https://doi.org/10.1111/j.1540-6261.1994.tb04772.x

Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, *10*(5), 603–621. doi: https://doi.org/10.1016/S0927-5398(03)00007-0

Ledoit, O., & Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, *30*(4), 110–119. doi: https://doi.org/10.3905/jpm.2004.110

Ledoit, O., & Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411. doi: https://doi.org/10.1016/S0047-259X(03)00096-4

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859. doi: https://doi.org/10.1016/j.jempfin.2008.03.002

Ledoit, O., & Wolf, M. (2011). Robust performances hypothesis testing with the variance. *Wilmott*, *2011*(55), 86–89. doi: https://doi.org/10.1002/wilm.10036

Ledoit, O., & Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Review of Financial Studies*, *30*(12), 4349–4388. doi: https://doi.org/10.1093/rfs/hhx052

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies*, *19*(4), 1499–1529. doi: https://doi.org/10.1093/rfs/hhj038

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, *81*(1), 146–159. doi: https://doi.org/10.1037/0022-3514.81.1.146

Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, *15*(5), 337–341. doi: https://doi.org/10.1111/j.0956-7976.2004.00679.x

Lettau, M., Maggiori, M., & Weber, M. (2014). Conditional risk premia in currency markets and other asset classes. *Journal of Financial Economics*, *114*(2), 197–225. doi: https://doi.org/10.1016/j.jfineco.2014.07.001

Li, J. (2015). Sparse and stable portfolio selection with parameter uncertainty. *Journal of Business and Economic Statistics*, *33*(3), 381–392. doi: https://doi.org/10.1080/07350015.2014.954708

Lintner, J. (1965, feb). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, *47*(1), 13–37. doi: https://doi.org/10.2307/1924119

Livermore, J. L. (1940). *How to trade in stocks; the Livermore formula for combining time element and price*. Duel, Sloan & Pearce.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, *1*(1), 41–66. doi: https://doi.org/10.1093/rfs/1.1.41

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, *90*(2), 426–432. doi: https://doi.org/10.1257/aer.90.2.426

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, *41*(5), 867–887. doi: https://doi.org/10.2307/1913811

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Topics in Catalysis*, *8*(4), 323–361. doi: https://doi.org/10.1016/0304-405X(80)90007-0

Michaelides, A., Milidonis, A., Nishiotis, G. P., & Papakyriakou, P. (2015). The adverse effects of systematic leakage ahead of official sovereign debt rating announcements. *Journal of Financial Economics*, *116*(3), 526–547. doi: https://doi.org/10.1016/j.jfineco.2014.12.005

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, *48*(3), 261–297. Retrieved from `https://www.jstor.org/stable/1809766`

Morck, R., Yeung, B., & Yu, W. (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics*, *58*(1-2), 215–260. doi: https://doi.org/10.1016/S0304-405X(00)00071-4

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica*, *34*(4), 768–783. doi: https://doi.org/10.2307/1910098

Nofsinger, J. R. (2005, sep). Social Mood and Financial Economics. *Journal of Behavioral Finance*, *6*(3), 144–160. doi: https://doi.org/10.1207/s15427579jpfm0603_4

Nooijen, S. J., & Broda, S. A. (2016). Predicting Equity Markets with Digital Online Media Sentiment: Evidence from Markov-switching Models. *Journal of Behavioral Finance*, *17*(4), 321–335. doi: https://doi.org/10.1080/15427560.2016.1238370

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, *108*(1), 1–28. doi: https://doi.org/10.1016/j.jfineco.2013.01.003

Odean, T. (1998). Volume, volatility, price, and profit when all traders are above average. *Journal of Finance*, *53*(6), 1887–1934. doi: https://doi.org/10.1111/0022-1082.00078

Ottaviani, M., & Sørensen, P. N. (2015). Price reaction to information with heterogeneous beliefs and wealth effects: Underreaction, momentum, and reversal. *American Economic Review*, *105*(1), 1–34. doi: https://doi.org/10.1257/aer.20120881

Pástor, u., & Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, *111*(3), 642–685. doi: https://doi.org/10.1086/374184

Peterson, R. L. (2007). *Inside the investor's brain.* John Wiley & Sons.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. doi: https://doi.org/10.1037/h0077714

Schmeling, M. (2009). Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, *16*(3), 394–408. doi: https://doi.org/10.1016/j.jempfin.2009.01.002

Sen, P. K., Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1981). Introduction to bivariate and multivariate analysis. *Journal of the American Statistical Association*, *76*(375), 752. doi: https://doi.org/10.2307/2287559

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, *5*(1), 1–55. doi: https://doi.org/10.1093/rfs/5.1.1

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, *9*(2), 277–293. doi: https://doi.org/10.1287/mnsc.9.2.277

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, *19*(3), 425–442. doi: https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Sharpe, W. F. (1967). Portfolio analysis. *Journal of Financial and Quantitative Analysis*, *2*(2), 76–84. doi: https://doi.org/10.2307/2329895

Shefrin, H., & Belotti, M. L. (2008). *A Behavioral Approach to Asset Pricing.* Elsevier. doi: https://doi.org/10.1016/B978-0-12-374356-5.X5001-3

Shen, J., Najand, M., Dong, F., & He, W. (2017). News and social media emotions in the commodity market. *Review of Behavioral Finance*, *9*(2), 148–168. doi: https://doi.org/10.1108/RBF-09-2016-0060

Shi, F., Shu, L., Yang, A., & He, F. (2020). Improving minimum-variance portfolios by alleviating overdispersion of eigenvalues. *Journal of Financial and Quantitative Analysis*, *55*(8), 2700–2731. doi: https://doi.org/10.1017/S0022109019000899

Shi, Y., Ho, K. Y., & Liu, W. M. (2016). Public information arrival and stock return volatility: Evidence from news sentiment and Markov regime-switching approach. *International Review of Economics and Finance*, *42*, 291–312. doi: https://doi.org/10.1016/j.iref.2015.12.003

Shu, H. C. (2010). Investor mood and financial markets. *Journal of Economic Behavior and Organization*, *76*(2), 267–282. doi: https://doi.org/10.1016/j.jebo.2010.06.004

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior and Organization*, *107*, 730–743. doi: https://doi.org/10.1016/j.jebo.2014.06.004

Sinha, A. (2016). Learning and the yield curve. *Journal of Money, Credit and Banking*, *48*(2-3), 513–547. doi: https://doi.org/10.1111/jmcb.12308

Stein, C. (1975). Estimation of a covariance matrix, Rietz Lecture. In *39th annual meeting ims, atlanta, ga, 1975.*

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Soviet Mathematics*, *34*(1), 1373–1403. doi: https://doi.org/10.1007/BF01085007

Sture Holm. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. Retrieved from `https://www.jstor.org/stable/4615733`

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking and Finance*, *73*, 147–164. doi: https://doi.org/10.1016/j.jbankfin.2016.09.010

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, *62*(3), 1139–1168. doi: https://doi.org/10.1111/j.1540-6261.2007.01232.x

Timmermann, A. (2006). Chapter 4 Forecast combinations. In *Handbook of economic forecasting* (pp. 135–196). North Holland, Amsterdam. doi: https://doi.org/10.1016/S1574-0706(05)01004-9

Titman, S., Wei, K. C., & Xie, F. (2004). Capital investments and stock returns. *Journal of Financial and Quantitative Analysis*, *39*(4), 677–700. doi: https://doi.org/10.1017/s0022109000003173

Tu, J., & Zhou, G. (2011). Markowitz meets Talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, *99*(1), 204–215. doi: https://doi.org/10.1016/j.jfineco.2010.08.013

Veronesi, P. (1999). Stock market overreaction to bad news in good times: A rational expectations equilibrium model. *Review of Financial Studies*, *12*(5), 975–1007. doi: https://doi.org/10.1093/rfs/12.5.975

Welch, B. L. (1947). The generalisation of student's problems when several different population variances are involved. *Biometrika*, *34*(1-2), 28–35. doi: https://doi.org/10.1093/biomet/34.1-2.28

Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin*, *31*(1), 121–135. doi: https://doi.org/10.1177/0146167204271309

Winkielman, P., Knutson, B., Paulus, M., & Trujillo, J. L. (2007). Affective influence on judgments and decisions: Moving towards core mechanisms. *Review of General Psychology*, *11*(2), 179–192. doi: https://doi.org/10.1037/1089-2680.11.2.179

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, *52*(2), 276–291. doi: https://doi.org/10.1016/0749-5978(92)90039-A

Xing, X., Hu, J., & Yang, Y. (2014). Robust minimum variance portfolio with L-infinity constraints. *Journal of Banking and Finance*, *46*(1), 107–117. doi: https://doi.org/10.1016/j.jbankfin.2014.05.004

Yen, Y. M. (2016). Sparse weighted-norm minimum variance portfolios. *Review of Finance*, *20*(3), 1259–1287. doi: https://doi.org/10.1093/rof/rfv024

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. doi: https://doi.org/10.1002/cne.920180503

Yu, J., & Yuan, Y. (2011). Investor sentiment and the mean-variance relation. *Journal of Financial Economics*, *100*(2), 367–381. doi: https://doi.org/10.1016/j.jfineco.2010.10.011

Curriculum Vitae

Personal Information

Name:	Christian Koeppel
Date of Birth	12th August 1990
Place of Birth	Munich (Germany)
Nationality	German

Education

02/2016 - present	**University of St. Gallen**, St. Gallen, Switzerland
	PhD in Finance
10/2012 - 10/2013	**University of Lancaster**, Lancaster, United Kingdom
	MSc in Quantitative Finance
09/2009 - 08/2012	**International School of Management**, Munich, Germany
	BA in International Management

Work Experience

07/2019 - present	**Julius Baer**, Zurich, Switzerland
	Product Manager Investment Management
07/2018 - 06/2019	**EPAM Systems**, Zurich, Switzerland
	Senior Manager Business Consulting Wealth Management
01/2016 - 06/2018	**Deutsche Bank Suisse S.A.**, Zurich, Switzerland
	Product Manager Wealth Management Technology
10/2013 - 12/2015	**Deloitte Consulting GmbH**, Munich, Germany
	Senior Consulting Financial Services