Essays on Return Predictability and Asset Pricing

Essays on Return Predictability and Asset Pricing

by
Eric James Wilson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in McMaster University 2025

Doctor of Philosophy, 2025

DeGroote School of Business, McMaster University

Hamilton, Ontario

Title: Essays on Return Predictability and Asset Pricing

Number of Pages: 15, 154

Author: Eric James Wilson

Email: wilsoe40@mcmaster.ca

ORCID iD: 0009-0001-9354-5995

Doctoral Committee:

Professor Ron Balvers, Co-Supervisor

Professor Jiaping Qiu, Co-Supervisor

Assistant Professor Yoontae Jeon

Assistant Professor Ken Li

Abstract

This dissertation studies return predictability amongst hedge fund and stock returns. In Chapter 1, I introduce a new measure of hedge fund manager skill that predicts hedge fund performance out-of-sample (OOS). In Chapter 2, we introduce an economically motivated loss function for predicting market risk premia for asset allocation. Lastly, in Chapter 3 I document new evidence that the consensus set of test assets used to price the cross-section of stock returns does not span the Stochastic Discount Factor (SDF). In doing so, I introduce hedge fund portfolios as a new class of test assets and show that it gets us closer to spanning the SDF.

Chapter 1 introduces a new measure of hedge fund manager skill, Edge, that predicts hedge fund performance OOS. In contrast to the standard approach that estimates skill based on an estimate of alpha, the Edge measure tells us how hedge fund managers produce alpha. A hedge fund manager has Edge if they produce positive alpha without being negatively affected by market downside risks. I document a new finding in the hedge fund literature: Hedge funds can be separated, ex-ante, into two groups, with respect to Edge. Only 3% of hedge funds possess Edge. OOS, hedge funds with (without) Edge have higher (lower) Sharpe Ratios and positive (negative) skewness. Hedge fund managers with Edge exhibit highly persistent performance, charge higher fees and run larger funds. I show the OOS performance of Edge is due to it's robustness to an unidentified form of hedge fund model misspecification: weak latent factors.

Next, I transition from Chapter 1 whereby I study return predictability among hedge fund returns to Chapter 2 in which I investigate return predictability in the time series of aggregate U.S. stock market returns.

Chapter 2, joint with Yoontae Jeon and Laleh Samarbakhsh, proposes a new approach for using stock market return predictors to maximize investor's utility gains by adding an economically motivated penalization term to the conventional OLS loss function. Our approach is computationally easy to implement and delivers superior OOS utility gains measured by

the certainty equivalent (CE). Moreover, the new methodology demonstrates that the advantage relative to the OLS approach becomes larger when an investor considers higher moments of portfolio returns and has a larger degree of risk aversion coefficient. Our results point toward the importance of aligning the loss function with the OOS evaluation metric when using return predictor variables.

In the final chapter I pivot from studying return predictability in the time series of aggregate U.S. stock market returns in the previous chapter to examining return predictability in the cross-section of U.S. stock returns.

Chapter 3 uses a comprehensive set of anomaly portfolios as test assets to determine their ability to span the SDF. I make three contributions. (1) I estimate the SDF from the leading principal component (PC) factors extracted from a set of anomaly portfolios the implied upper bound and show that they do not span the SDF. (2) I introduce hedge fund portfolios as a new class of test assets and show that it gets us closer to spanning the SDF by generating a Sharpe Ratio that is 3.4x higher than the traditional anomaly portfolios. (3) I show that the lower ranked PCs which are commonly excluded in standard factor models emerge as an important feature in spanning stock returns during periods of market turbulence.

Acknowledgements

I am grateful to the many individuals that made this dissertation possible. First and foremost I want to thank my co-supervisors Ron Balvers and Jiaping Qiu, for their steadfast support, encouragement, and guidance throughout this journey. Ron's openness and enthusiasm to my ideas have been instrumental in giving me the confidence to pursue this research agenda. Moreover, Ron has been so generous with his time in always making himself available to provide feedback on my work. I am deeply indebted to Jiaping for his mentorship that started from my initial conversation with him in the summer before starting the PhD program. I am especially thankful to Jiaping for asking some of the most difficult and relevant questions that have pushed my research further. I have learned a tremendous amount from both Jiaping and Ron and owe them for my success in the program.

I thank my committee members Yoontae Jeon and Ken Li. It was a game changer for me when Yoontae joined McMaster. I will never forget how he took a deep interest in my work after I first presented my job market paper. I'm fortunate that we share many research interests and I have learned so much from Yoontae and grown as a researcher because of all of the time he has spent with me. I thank Ken for all of his support and providing detailed comments on my work along the way. I want to acknowledge Clarence Kwan for his unparalleled level of dedication to reviewing my work and enthusiasm for my research. I would also like to thank the finance faculty and my fellow PhD students at McMaster.

There are many people in the finance community outside of McMaster that I also want to pay tribute to. First among them, I want to thank Peter Cziraki who has been a constant in my academic life since my time at the University of Toronto. I am incredibly grateful for Peter's unwavering support over the years. I owe a tremendous debt to Craig Doidge who's support and feedback on my job market paper and job talk without question positioned me to succeed on the market. To this end, I want to acknowledge Ambrus Kecskés who has been so generous with his time, advice and feedback during my time on the job market. Finally, I want to thank Ilias Tsiakas who is the reason I am an academic in the first place.

Most importantly, this dissertation would not have been possible without the unconditional love and support of my wife Medina, my siblings and my parents. It is to them that I dedicate my PhD.

Table of Contents

List of Tables

List of Figures

Declaration of Authorship

I, Eric James Wilson, declare that this thesis, entitled, "ESSAYS ON RETURN PREDICTABILITY AND ASSET PRICING," and the work presented in it, are my own. I confirm that the thesis comprises the following chapters:

- Hedge Funds With(out) Edge: A New Measure of Hedge Fund Manager Skill;
- The Value of Economic Regularization for Stock Return Predictability (joint work with Yoontae Jeon and Laleh Samarbakhsh); and
- What is the Implied Upper Bound of the Stochastic Discount Factor?

This thesis is entirely my own original work unless otherwise indicated. Any use of the work of other authors is acknowledged at their point of use.

Chapter 1

Hedge Funds With(out) Edge: A New Measure of Hedge Fund Manager Skill

1.1 Introduction

This paper develops a new measure of hedge fund manager skill, Edge, that predicts hedge fund performance out-of-sample (OOS). Overcoming poor OOS performance is the central hurdle in identifying hedge fund managers with skill, and it is the focus of this paper. We observe hedge fund returns, but not the underlying strategies used to generate these returns. Hedge fund managers can produce positive alpha in-sample (IS) by either implementing a strategy that is not harmed from market turbulence, or is negatively impacted by taking on leveraged exposure to market downside risks. To avoid the mistake of attributing skill to the latter group of hedge fund managers, I rely on estimates of both the intercept and the the coefficient loading on the benchmark. The slope coefficient is informative since it also picks up the covariation with risk factors that we are unable to control for, which have a disproportionate impact during periods of market turbulence. By using estimates of the intercept and slope coefficient, Edge tells us how hedge fund managers produce alpha.

I base the analysis on a hedge fund benchmark called "Short VIX," an investment strategy that shorts near-dated constant maturity VIX futures contracts. Short VIX earns positive returns when markets are calmer than consensus expectations, and vice versa. I show Short VIX is a closer approximation to the average hedge fund's returns relative to traditional hedge fund benchmarks such as Jurek and Stafford (2015) that writes out-of-the-money (OTM) put options on the S&P 500. Although I rely on Short VIX, the Edge approach can be successfully adopted using alternative hedge fund benchmarks which include multi-factor hedge fund factor models.

The central question is whether hedge fund managers have skill. Are hedge funds simply producing alpha from hidden risks which materialize infrequently? How many hedge funds generate alpha that does not disappear OOS? What do hedge fund managers with positive OOS alpha look like, and where does their alpha come from?

The incentives of a hedge fund manager can explain why there are many funds that produce positive IS alpha that disappears OOS. Hedge fund managers are compensated with a management fee based on a percentage of Assets under Management (AUM) and a performance fee. Hence, hedge fund managers are incentivized to raise as much capital as possible to earn greater management fees. Given the information asymmetry between investors and hedge fund managers, investors often evaluate hedge funds based on Sharpe Ratio (SR) over a sufficiently long track record (e.g. a minimum of three years). It follows, a hedge fund with a higher SR over a longer track record will raise more money, all else equal. Therefore, unskilled hedge fund managers are incentivized to implement strategies with high SRs which appear as positive alpha, by taking on exposure to market downside risks.

The challenge is to find a method that can identify hedge fund managers with positive OOS alpha. By construction, any estimate of hedge fund skill is a consequence of the benchmark. The benchmark constructed from either observable or latent risk factors will inevitably omit return variation that is inconsequential during the IS period. However, some of this omitted return variation from strategies with leveraged exposure to downside risks, positively covaries with the benchmark, and has a disproportionately negative impact on OOS returns during times of market turbulence. The group of hedge funds that implement high SR strategies from leveraged exposure to market downside risks will suffer large losses during these periods. The basic insight of this paper is that identifying skill based on only the estimate of IS hedge fund alpha will lead to poor OOS performance so long as there is a majority of hedge fund managers who implement high SR strategies which are negatively affected by market downside risks.

I overcome the problem of identifying hedge fund managers that produce positive alpha OOS by further relying on the information within the coefficient estimate of the benchmark.

We know unskilled hedge fund managers with positive alpha have some positive exposure to the benchmark which becomes larger during periods of market stress. Therefore, we can identify the skilled managers by selecting hedge funds with a positive intercept whose exposure to the benchmark is either zero or strictly negative. I call this latter group of managers, hedge funds with Edge. The label Edge conveys that we are identifying hedge fund managers that implement a strategy with lasting alpha which persists OOS. In contrast, "skill" is synonymous with alpha which is highly time-varying. More generally, the term Edge is used to reflect a strategy that has a positive expected value based on the true unknown distribution (MacLean et al., 2011). With that, hedge funds with Edge implement strategies that have a positive expected OOS return based on the true underlying return distribution, in contrast to those without Edge whose expected OOS returns are negative.

In the cross-section, only 3% of hedge funds have Edge, based on AUM. Over the OOS period from May 2008 to December 2022, hedge funds with Edge generate a compound return, in excess of the riskless rate, of 5.7% whereas hedge funds without Edge produce a compound return of -0.2%. This finding suggest that 97% of hedge funds produce OOS returns that are indistinguishable from the riskless return with significantly more risk. Only a small minority of hedge funds are capable of generating positive OOS alpha.

Nearly half (49%) of hedge funds with Edge are represented by global macro strategies, with the next four largest strategies represented by Fund of Funds (11%), Options Strategy (9%), Multi-Strategy (8%) and Equity Market Neutral (7%). The disproportionate number of global macro hedge funds is intuitive given that this strategy is designed to benefit from periods of market turbulence. The remaining half of hedge funds with Edge points to a more diverse group of strategies with no obvious trait that is common among the strategies.

I find that compared to all hedge funds, those with Edge tend to be larger in AUM (US$ 1.1 bn vs. $282 mm), suggesting that the Edge measure is not picking up low capacity hedge funds that are of little to no consequence to hedge fund investors. In fact, the Edge approach captures some of the largest hedge funds with the greatest invesment capacity. I also find that hedge funds with Edge charge higher management (1.7% vs. 1.4%) and performance fees (18.3% vs. 14.2%). Interestingly, hedge fund investors are willing to pay up for fund managers with Edge, thereby suggesting that they can infer their quality at some level.

Lastly, hedge funds with Edge demonstrate persistence in the returns they generate with both higher SRs and positive skewness up to five years ahead. This evidence addresses the concern that the percentage of hedge funds with Edge of 3% is too small and might simply be due to sampling error. With a 95% confidence interval, we expect 5% of hedge funds have skill simply due to chance. However, if we assume independent and identically distributed hedge fund returns, the proportion of hedge funds with skill due to sampling error falls to 0.25% with two-year ahead returns based on a 5% level of significance. In contrast, the proportion of hedge funds with Edge remains persistent at between 3-4% up to five years ahead. Moreover, the superior OOS performance of hedge funds with Edge grows to its peak at three-year ahead returns with a SR of 0.77 and skewness of 1.12. All together, this piece of evidence suggest that the Edge approach is capturing lasting alpha that cannot be attributed to sampling error.

The Edge measure's efficacy in determining skill cannot be explained by either exposure to higher moment risk or macroeconomic uncertainty which have been shown to be key drivers of hedge fund returns. Kelly and Jiang (2012) document the large negative exposure hedge funds bear to downside tail risk. Avramov et al. (2011) connects hedge fund performance to macroeconomic factors in order to detect hedge fund manager skill. I verify that those with only a positive intercept have highly significant exposure to both higher moment risk and macroeconomic uncertainty. By solely relying on the intercept as a performance measure, we identify hedge funds that produce higher SRs at the expense of greater negative skewness. In contrast, hedge funds with Edge are able to produce returns independent of higher moment risk and macroeconomic uncertainty, which allows them to generate both higher SRs and positive skewness independent of these two well-known sources of hedge fund returns.

The efficacy of the Edge approach can be explained by its robustness to weak latent factors, a previously unidentified form of hedge fund model misspecification. Giglio et al. (2021) show that there is significant common variation in hedge fund returns even after controlling for the Fung and Hsieh (2004) seven-factor model. I build on this result by showing that the strong factor structure in hedge fund returns (residuals) weakens during periods of market turbulence such as the 2008 Financial Crisis and the 2020 COVID-19 Stock Market Crash. I show that hedge funds that load positively on the benchmark (Short VIX) generate returns that covary positively with the lower ranked PCs (weak latent factors) which results in negative skewness for this group of hedge funds. By avoiding these hedge funds that are negatively affected by weak latent factors, the Edge measure is able to identify a small group of hedge funds that either benefits or is not harmed from periods of market turbulence.

Literature

This paper's distinctive contribution is to propose a new measure of hedge fund manager skill, Edge, that is able to identify those hedge funds with superior OOS performance. I contrast my results with the most closely related papers in this literature.[1]

Agarwal et al. (2024) introduces a new measure of hedge fund manager skill, the unobserved performance (UP) of a hedge fund, based on the difference between a hedge fund's reported return and the portfolio return inferred from its reported long-equity holdings. They find that hedge funds with high UP earn a 6.36% annualized return that is higher than hedge funds with low UP. Although this new skill measure produces higher returns, it comes at the expense of more negative skewness during the 2008 Financial Crisis. Moreover, this new skill measure can only be applied to hedge funds who report long only equity positions. In contrast, the Edge approach employs a returns-based methodology that can be applied

[1] Several important studies within this literature that are also related to this paper include Agarwal et al. (2013); Agarwal and Naik (2000); Agarwal et al. (2023, 2024); Ardia et al. (2024); Avramov et al. (2011); Barth et al. (2023); Bollen et al. (2024); Brown et al. (1999); Chen et al. (2017); Chen and Liang (2007); Chen and Zhang (2023); Edwards and Caglayan (2001); Fung and Hsieh (2001, 2002, 2004); Giglio et al. (2021); Griffin and Xu (2009); Grinblatt et al. (2020); Jagannathan et al. (2010); Jiang et al. (2021); Titman and Tiu (2011).

broadly across the full universe of hedge funds.

Chen et al. (2017) infer hedge fund manager skill using a mixture model that assumes the ability of a hedge fund manager comes from one of several skill groups that can each be affected by some degree of luck. They find that 48% of hedge funds possess skill. In contrast, the Edge measure of skill identifies a much smaller group of skilled hedge funds of 3%. This is a direct result of conditioning on the slope coefficient, in addition to the intercept. Their approach recognizes an important issue that returns can also be affected by luck. However, the Edge approach takes into account another pervasive issue regarding hedge fund returns: the inverse relationship between the intercept and market downside risks among hedge fund managers without skill.

The recent studies highlighted above and the rest of the literature on hedge fund manager skill has tried to address the problem of positive IS alphas turning into negative OOS alphas by either: (i) luck or (ii) model misspecification. The common thread among these two approaches is the implicit assumption that a strong factor structure holds which has resulted in poor OOS performance. I contribute to the hedge fund literature by showing how the deterioration in OOS performance can instead be attributed to the presence of a previously unidentified form of hedge fund model misspecification, weak latent factors.[2]

Giglio et al. (2021) estimates hedge fund manager skill by estimating alpha from a factor model that adds latent factors to the well known Fung and Hsieh (2001, 2002, 2004) seven (FH7) factor model. These latent factors are estimated by the leading principal components (PCs) of the FH7 residuals covariance matrix. I build on the results of Giglio et al. (2021) by showing how the strong factor structure in hedge fund residuals weakens significantly during periods of market turbulence. This new evidence regarding a weaker factor structure indicates that the efficacy of the Edge approach is its robustness to weak latent factors. Importantly, this new evidence suggests that any measure of hedge fund manager skill based

[2]Several important studies within this literature include Anatolyev and Mikusheva (2022); Bryzgalova (2015); Bryzgalova et al. (2023b); Giglio et al. (2023a,b); Kan and Zhang (1999); Kleibergen (2009); Lettau and Pelger (2020).

entirely on the intercept will fail OOS due to the violation of the strong factor assumption.

The remainder of the paper is organized as follows. In Section 1.2 the Edge methodology, data, and Short VIX benchmark are detailed. Section 1.3 displays the main empirical results of the paper. Section 1.4 goes on to show why Edge has attractive OOS performance. Lastly, Section 3.5 concludes.

1.2 Edge Methodology, Data and Short VIX Benchmark

This section provides an overview of the Edge methodology, data sources used in the paper, CFTC hedge fund position level data in VIX futures and the construction of the benchmark Short VIX strategy. Summary statistics of the leading hedge fund replicators along with the new Short VIX benchmark are also provided.

1.2.1 Edge Methodology

This section details where the Edge methodology comes from and the specific measure used to determine whether or not a hedge fund manager possesses Edge. Recall that Edge is defined as a hedge fund manager who produces a positive and significant intercept without positive exposure to market downside risks. To operationalize this definition with an empirical measure I begin with the original market timing test of Treynor and Mazuy (1966) in order to reflect exposure to market downside risks:

$$r_t - r_{f,t} = a + b\left(r_{m,t} - r_{f,t}\right) + c\left(r_{m,t} - r_{f,t}\right)^2 + u_t, \quad \text{where } \mathrm{H}_0\text{: } \gamma \leq 0, \tag{1}$$

where r_t, $r_{f,t}$ and $r_{m,t}$ denote the hedge fund, risk-free rate and market return, respectively, all measured at date t. $c > 0$ indicates superior market timing ability. Given that the market has negative skewness, a negative gamma reflects positive exposure to market downside risks.

In light of the dynamic among hedge fund managers and investors described earlier in this paper, I make two modifications to Equation 1 to successfully apply it to hedge funds. First, I substitute $r_{m,t}$ with $r_{SVIX,t}$, the returns from shorting VIX futures.[3] This addresses the robust stylized fact that hedge fund returns can be successfully replicated by concave payoffs on the S&P 500.[4] Moreover, the Short VIX benchmark reflects an investment strategy that is attractive to a hedge fund manager without Edge given that it offers a high SR relative to the S&P 500 at the expense of positive exposure to market downside risks. With that substitution we get the following specification,

$$r_t - r_{f,t} = \alpha + \beta \left(r_{SVIX,t} \right) + v_t, \tag{2}$$

Equation 2 can be viewed as informative as Equation 1, at a minimum, since the Short VIX benchmark is able to subsume the linear and quadratic terms in the market portfolio (see Section 1.3), thereby capturing the most notable nonlinearities in hedge fund returns. Importantly, the beta of Equation 2 reflects significant positive exposure to market downside risks which is the crux of the Edge measure. Hence, a hedge fund manager that has a strictly positive Short VIX beta implicitly has large positive exposure to market downside risks.

The second modification I make is to add a quadratic term in the new Short VIX benchmark, $\left(r_{SVIX,t} \right)^2$, to account for any remaining nonlinearities that might affect hedge fund returns from higher moment risk beyond what Short VIX can capture linearly.[5] In addition, the quadratic term also controls for any timing with respect to the Short VIX benchmark that could affect the parameter estimates. All together, these two modifications yield the

[3]I construct the $SVIX$ benchmark by building a strategy that shorts second-dated constant maturity VIX futures at a monthly frequency. Monthly rollovers are captured by computing monthly roll-adjusted returns based on constant exposure to one contract of the second nearest-dated CBOE VIX futures contract. Daily log roll-adjusted returns are summed within each month to produce the SVIX monthly return series.

[4]I have chosen to use the Short VIX futures based strategy as a benchmark instead of the Jurek and Stafford (2015) option-based measure since Short VIX has a return profile (i.e. SR and Skewness) that more closely matches hedge fund data.

[5]The hedge fund literature (see for example Kelly and Jiang (2012)) has shown extensively how tail risk (i.e. higher moment risk) is a key driver of hedge fund returns.

main empirical test and specification of the paper:

$$r_t - r_{f,t} = a + b\left(r_{SVIX,t}\right) + c\left(r_{SVIX,t}\right)^2 + \varepsilon_t, \text{ where } \mathrm{H}_0\text{: } a \leq 0 \text{ or } b > 0, \tag{3}$$

By Equation 3, the test of a hedge fund manager's skill is the following rejection of the null hypothesis (H_0): (i) evidence of a positive and significant intercept ($a > 0$); and (ii) evidence of no positive exposure to market downside risks ($b \leq 0$). I propose this rejection of the null hypothesis as a new measure of investment skill which I name Edge.[6]

1.2.2 Sources

The data used in this paper is from the following sources: i) Bloomberg, ii) the Hedge Fund Research (HFR) Database, (iii) TASS, (iv) OptionMetrics (v) the Fung and Hsieh (2001) trend-following factors[7], (vi) Kenneth French's Online Data Library[8], (vii) the Federal Reserve Bank of St. Louis (FRED), (viii) CRSP and (ix) the CFTC Commitments of Traders in Financial Futures (TFF) report. From Bloomberg, I collect end-of-month spot Foreign Exchange (FX) rates (with USD as the base currency), daily closing prices of the S&P 500, Russell 2000 index, MSCI Emerging Market index, CBOE VIX Spot and Futures (second- and third-nearest-dated contract) from May 1, 2004[9] to December 31, 2022. I exclude the front month VIX futures contract to ensure that the strategy is realistic in its implementation and avoids the risk associated with liquidation prior to expiration.

From the HFR Database, I collect the HFR 500 Composite Index end of month Net Asset Values (NAVs) from January 2005 (the start date of the database) to December 2022.

[6]It is important to note that the quadratic term in Equation 3 introduces a bias in estimating the intercept (i.e. negative market timers will have larger intercepts, vice versa). I have chosen to base the main specification of this paper using Equation 3 since the Edge approach does not solely rely on an estimate of the intercept but instead relies on joint estimation of the intercept and the coefficient on a proxy for market downside risks. In light of the bias-variance tradeoff, the Edge methodology results in lower variance at the expense of more biased estimates (i.e. the efficacy of the Edge methodology is its predictive power).

[7]http://people.duke.edu/ dah7/DataLibrary/TF-Fac.xls

[8]https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[9]The first trading date of the CBOE VIX futures occurred on March 26, 2004.

The HFR 500 Composite Index consists of the largest funds that report to HFR Database that are open to new investments and offer quarterly liquidity or better. The HFRI 500 is a representative and broadly diversified benchmark for the hedge fund industry. The HFRI 500 Composite is equally weighted across its constituents and is comprised of the following hedge fund strategies: (i) Equity Hedge, (ii) Event-Driven, (iii) Macro, (iv) Relative Value and (v) Emerging Markets. From FRED, I collect the 10-year treasury constant maturity yield and the Moody's Baa yield to construct the Bond Market Factor and Credit Spread Factor, respectively as part of the FH8 model. In addition, I collect the 1- and 3-Month US Treasury Bill Yields, Moody's Seasoned Aaa Corporate Bond Yield, Default Spread (BAA-AAA) as part of the macroeconomic variables from Avramov et al. (2011); Fama and French (1989). I collect total cash dividends, closing prices and shares outstanding from CRSP for all publicly-listed companies to construct the dividend yield for the CRSP value-weighted index. From Kenneth French's website, I collect the Fama and French (2015) five factors (FF5), Carhart (1997) cross-sectional momentum, in addition to the risk-free rate, all at a monthly frequency.

It is widely documented in the hedge fund literature that hedge fund data suffers from various biases due to the fact that reporting to hedge fund databases is voluntary.[10] To remedy this, I have chosen the most widely studied hedge fund database, the TASS[11], accessed in February 2023. Furthermore, I have chosen to produce my empirical results with the most widely studied hedge fund database given that the paper's main aim is to shed new light on the main stylized fact in the hedge fund literature (i.e. high SRs at the expense of negative skewness). Section 3.5 details the original dataset and methodology used to clean the commercial hedge fund data of biases that have been previously documented in the hedge fund literature. The final cleaned TASS panel data used for the fund-level empirical work is reduced to 334,354 hedge fund month observations. This represents 4,078 unique hedge

[10]See Chapter 1 of Pedersen (2015) for a summary of the main biases present in hedge fund data.

[11]Joenväärä et al. (2021) review seven commercial hedge fund databases and find that TASS is the most widely used commercial database in academic research (79% of 92 papers reviewed used TASS) since it offers one of the highest quality datasets in terms of coverage and lack of survivorship bias after 1994.

funds of which 811 survived (3,267 died). Nearly 90% of the final set of unique hedge funds is spanned by three countries, with respect to the fund's base currency: U.S. Dollar (63%), Euro (20%) and Swiss Franc (4%). All monthly returns used in this study are in excess of the one-month risk-free rate, unless stated otherwise.

I use option price data from OptionMetrics to construct the OTM S&P 500 Put-Writing strategy in addition to the Gormsen and Jensen (2022) higher moment risk measures of the S&P 500. The sample period is from May 2004 to December 2021 at a daily frequency. The sample ends in December 2021 due to data limitations. I build the OTM put-writing measure using the same methodology as Jurek and Stafford (2015) with one modification. Instead of writing S&P 500 (SPX) put options at fixed Z-Scores, I instead choose the option with a fixed moneyness and expiration. Specifically, I choose the option whose delta is closest to the upper bound of the following range: $\Delta^{\text{put}} \in [-0.4, -0.2)$ as in Koijen et al. (2018) with a maturity that is closest to one month to expiration. The final strategy assumes leverage of two times unleveraged asset capital and produces a highly similar return profile with Jurek and Stafford (2015), in terms of SR and skewness.

1.2.3 Hedge Funds Short VIX Futures

Figure 1 displays the net long (total longs minus total shorts) CBOE VIX futures positions (grey line) of the leveraged funds group (i.e. hedge funds) from the CFTC Commitment of Traders in Financial Futures (TFF) report. The data is reported weekly spanning from August 29, 2006 to December 27, 2022. The red line shows the average over the entire sample period equal to a net long VIX position of -53,937 contracts, with a minimum and maximum net long position of -195,486 and 63,753, respectively. This data confirms that hedge funds are on average shorting VIX futures thereby providing the initial piece of evidence that motivates using the Short VIX benchmark to evaluate hedge funds. The shaded regions indicate periods when hedge funds hold a positive net long position. The initial two years of Figure 1 show that hedge funds held a positive net long VIX position consistently every week until September 23, 2008. From September 30, 2008 onwards, hedge funds were net short 93% of the time (81% over the full sample period), and have not held a positive net long position since April 17, 2018. The time-series dynamic of hedge fund positions in VIX futures highlights that hedge funds are shorting VIX futures the vast majority of the time.

1.2.4 Short VIX Measure

Here, I detail the construction of the benchmark used in this paper to identify hedge fund managers with Edge. The benchmark is an investment strategy that shorts VIX futures contracts with monthly rollovers.[12]

Figure 2 displays the cumulative logarithmic monthly returns generated from the benchmark strategy of shorting second-dated constant maturity VIX futures at a monthly frequency. The returns to this strategy are significantly higher than the market portfolio with a SR of 0.95 and cumulative returns of 920% over the full sample period from May 2004 to December 2022. The strong positive returns come with significant tail risks (negative skewness of -1.29) which is clearly indicated in the substantial drawdowns in two periods: (i) Around the 2008 Financial Crisis from June 2007 to March 2009 investors would have lost 158% applying this strategy; and (ii) At the beginning of the COVID-19 Global Pandemic from January 2020 to March 2020 investors would have lost 97%. Of course, the large magnitude of each drawdown reflects the Short VIX's unscaled annualized volatility of 52%.

It is important to understand what underpins the return profile of the Short VIX strategy. The vast majority of the time, the VIX term structure (or futures curve) is upward sloping[13] due to dealers' hedging demand of longer-term VIX futures to hedge equity market downside risk (Cheng, 2019).[14] The presence of these market participants results in a highly liquid

[12]Monthly rollovers are captured by computing monthly roll-adjusted returns based on constant exposure to one contract of the second nearest-dated CBOE VIX futures contract. Daily logarithmic roll-adjusted returns are summed within each month to produce the Short VIX monthly return series.

[13]At a daily frequency, 83% of the time, from May 2004 to December 2022, the nearest dated VIX futures contract is cheaper than the second nearest dated contract.

[14]Alternatively, it might also be a manifestation of differences in the Variance Risk Premium (VRP) across maturities. It remains an active area of research as to what drives the shape of the VIX term structure.

VIX futures market. Hence, an investor can collect the roll yield by shorting a longer-dated contract (i.e. 2 months from now) and holding it for one-month as the contract falls in value. The investor can liquidate the contract and repeat this process by buying the new second dated contract. The main risk is that a material and unforeseen market event occurs (i.e. COVID-19), in which case the VIX spikes upward leading the VIX futures curve to invert or shift upward. In this scenario, the investor who's shorted VIX futures suffers large losses. In the presence of significant hedging demand, these episodes are rare. It is this dynamic that results in the Short VIX trading strategy's return profile: high SR and negative skewness.

1.2.5 Summary Statistics

A Short VIX strategy with monthly rollovers is an attractive benchmark for hedge fund returns since it offers a hedge fund-like return profile (i.e. high SR with less negative skewness relative to competing hedge fund replicators). The most common benchmark that's used to replicate hedge fund returns is writing S&P 500 put options (Jurek and Stafford, 2015) or writing a lookback straddle on equities (Fung and Hsieh, 2001). Table 1 displays the summary statistics (annualized) for hedge fund returns and leading candidates to replicate hedge fund returns from January 2005 to December 2022. This includes the HFRI 500 Index, TASS equal-weighted portfolio, Short FH Stock Index Lookback Straddle (PTFSSTK), OTM S&P 500 Put-Writing Strategy and Short VIX strategy.

The HFRI 500 Composite provides a useful hedge fund benchmark given that it is widely used in research (see Jurek and Stafford (2015)) as a measure of aggregate hedge fund performance. Over the sample period, the HFRI 500 has a SR of 0.61 with negative skewness of -1.18, thus highlighting the concave payoff profile often associated with hedge fund returns. It is noteworthy that the equal-weighted hedge fund portfolio constructed from the cleaned TASS data is highly correlated (0.97) with the HFRI 500. The TASS data has a lower SR of 0.21 which is a direct result of applying the Couts et al. (2020) 3-step unsmoothing method that has increased volatilities significantly while leaving the mean unchanged. The HFRI

500 mean return which is double the mean return of the TASS data reflects the selection bias inherent in the HFRI 500. This selection bias is by construction, in that the HFRI 500 represents the largest funds that choose to report to the HFR database.

The short Straddle, short Put and Short VIX represent three potential substitutes that an investor could use to replicate hedge fund returns. The SRs range from 0.82 (short Straddle) to 0.91 (Short VIX) with skewness ranging between -2.44 (short Straddle) to -1.26 (Short VIX) and correlations with the HFRI 500 from 0.40 (short Straddle) to 0.74 (short Put). The Short VIX strategy has the second highest correlation (0.72) with the HFRI 500 and more importantly has a highly similar return profile with respect to SR and skewness.[15] In particular, the Short VIX displays a SR of 0.91 and skewness of -1.26 compared to the HFRI 500 which has a SR of 0.61 and skewness of -1.18. The clearest alternative to using the Short VIX would be to use the Short Put. While it has a similar SR of 0.85, it has significantly more negative skewness of -2.26. The Short VIX's more closely matched skewness to hedge fund returns is indicative of a benchmark that will have greater statistical power relative to the short Put and short Straddle hedge fund replicators. This is confirmed by Table A3 which shows how the Short VIX is able to eliminate all significant return variation (i.e. insignificant intercept) at the panel hedge fund month level.

1.3 Main Empirical Results

In this section I document the performance attributes of the Short VIX benchmark. Subsequent to this, I then present the main finding of the paper that hedge funds can be classified into two groups, ex-ante, based on the new Edge measure.

[15]Figure A2 plots the rolling 36-month correlations between the HFRI 500 and the Short VIX strategy returns from December 2007 to December 2022. The average correlation is 0.73 ranging between 0.51 and 0.89.

1.3.1 What is the Short VIX Benchmark Capturing?

Let us consider testing the Short VIX hedge fund benchmark using the Treynor and Mazuy (1966) market-timing model:

$$r_{\mathrm{SVIX},t} = a + b\left(r_{m,t} - r_{f,t}\right) + c\left(r_{m,t} - r_{f,t}\right)^2 + u_t, \tag{4}$$

where $c > 0$ indicates superior market timing.

Table 2 reports the empirical estimates from estimating Equation 4. The performance attributes of the Short VIX benchmark are the following: (i) intercept, a, of 4.0% per month; (ii) slope coefficient, b, of 2.47; and (iii) market timing coefficient, c, of -8.118. All parameter estimates are significant at the 1% level. In a nutshell, the Short VIX benchmark is a levered market position (as indicated by the b coefficient in excess of one) with negative market timing (i.e. negative c coefficient). Moreover, the significant positive intercept indicates that the Short VIX is able to subsume the Treynor and Mazuy (1966) specification thereby lending empirical support for the modification of substituting the market factor for Short VIX between Equations 1 and 2. Table A2 provides additional evidence that the Short VIX benchmark generates highly significant alphas with respect to the traditional linear asset pricing factor models (Carhart, 1997; Fama and French, 1992, 2015). The additional return variation and negative market timing attributes that characterize the Short VIX will prove useful in its ability to capture the nonlinearities in hedge fund returns across an array of strategy types.

1.3.2 Hedge Fund Exposure to the Short VIX Benchmark

Here I examine hedge fund return exposure to Short VIX at the fund-level by hedge fund strategy type. Table 3 displays the panel regression results at the hedge fund month level by primary category (i.e. hedge fund strategy), with hedge fund fixed effects and standard errors clustered at the fund level. Panel A is with respect to the Short VIX as the benchmark whereas the OTM short Put is used as the benchmark in Panel B for comparison purposes. The regression specification is as follows,

$$r_t^i = a + h^i + b\,(\text{Benchmark}_t) + c\,(\text{Benchmark}_t)^2 + \varepsilon_t^i, \tag{5}$$

where r_t^i is the hedge fund excess return, h^i is a hedge fund-specific fixed effect, a is unexplained return variation, Benchmark_t refers to either the Short VIX or OTM short Put monthly return, b is the coefficient loading on the benchmark and c is indicative of timing with respect to the benchmark. The sample period is May 2004 to December 2022 at a monthly frequency.

There are several key takeaways from Table 3. First, all strategies (except Managed Futures) load significantly, at conventional significance levels, on Short VIX as indicated by the b coefficient. Dedicated Short Bias is highlighted in red since it is the only strategy that has a significant negative b coefficient loading, which is intuitive since Dedicated Short Bias should perform well from asset prices falling. Given the performance attributes of the Short VIX, this evidence indicates that most hedge funds hold a leveraged position in the market and are negative market timers. Second, hedge funds have negative exposure to the quadratic term in the Short VIX (negative gamma) across all strategies, except for Dedicated Short Bias, Global Macro and Managed Futures. This evidence suggest that hedge funds have exposure to higher moment risk that was not fully captured by linear exposure to the Short VIX benchmark thereby lending support to the second modification made between Equations 2 and 3. Third, there is no remaining unexplained variation left for Emerging Markets, Managed Futures and Multi-Strategy after estimating Equation 5. Convertible Arbitrage is only marginally significant at the 10% level.

Lastly, across all strategies the Short VIX benchmark (Panel A) produces a clear pattern whereby hedge funds are on average loading positively on Short VIX (positive b coefficient) and have negative exposure to higher moment risks, consistent with the empirical record documented in the hedge fund literature (i.e. hedge funds produce higher SRs by exposing themselves to downside risks in the market). In contrast, in Panel B the OTM short Put produces mixed results across all strategies with both a positive b and c coefficients. The mixed results using the OTM short Put indicates that it has less statistical power relative to the Short VIX in identifying hedge funds with Edge. All together, Table 3 provides strong evidence that suggest that hedge funds have significant exposure to the Short VIX benchmark across a diverse group of strategies. Moreover, hedge fund loadings tend to increase on the Short VIX at the worst possible time.

1.3.3 Hedge Funds With and Without Edge

In this section, I document the main finding of the paper: hedge funds can be characterized by two groups, ex-ante, based on Edge. Hedge funds with (without) Edge have both higher (lower) SRs and positive (negative) skewness. Subsequent to presenting this new result, I then provide additional evidence that highlights the robustness of this result.

1.3.3.1 Annual OOS Returns

The Edge measure successfully partitions hedge funds into two very distinct groups, OOS, thereby providing us with a more nuanced understanding of hedge funds. In Figure 3, one group (Panel A) is entirely consistent with the stylized fact regarding how hedge fund returns can be characterized as concave payoffs in the S&P 500 with a high SR at the expense of large negative skewness (SR of 0.16 with Skewness of -0.72).[16] In sharp contrast, the other group (Panel B) features both higher SRs and positive skewness (SR of 0.53 with Skewness of 0.45). This latter group, which was unobservable until this paper, is comprised of hedge funds that employ a diverse range of strategies. By introducing the Edge measure, I document a new finding that sheds more light on hedge fund performance at the aggregate level: Hedge

[16] The SRs are significantly lower in magnitude from having applied the 3-step unsmoothing methodology recently proposed by Couts et al. (2020) to remove the autocorrelation in reported hedge fund returns at the fund- and strategy-level.

funds are comprised by instead *two* groups, ex-ante, based on Edge. With the introduction of Edge, this study complements the existing hedge fund literature that characterizes hedge fund performance as largely described by higher risk-adjusted returns at the expense of more negative skewness.

1.3.3.2 OOS Cumulative Portfolio Returns

Figure 4 presents the OOS cumulative returns based on three equal-weighted portfolios: (i) All hedge funds in the cleaned TASS Database (grey line); (ii) hedge funds with Edge (blue line); and (iii) hedge funds without Edge (red line). The Edge measure is based on parameter estimates from three-year rolling regressions of Equation 3, with an annual rebalancing of the portfolio each December. Recall, a hedge fund with Edge is identified as one that produces unexplained return variation (i.e. a positive intercept) without exposure to market downside risks.[17]. Figure 4 showcases the main finding of the paper: hedge funds with Edge have a higher SR (0.53) and positive skewness (0.45) versus those without Edge which have a lower SR (0.16) and negative skewness (-0.72). It is interesting to note that the grey and red lines are on top of one another. In other words, the main stylized fact documented in the literature (i.e. hedge fund returns can be represented by concave payoffs in the S&P 500 with relatively higher SRs at the expense of more negative skewness) is a reflection of those hedge funds that do not possess Edge. This new piece of OOS evidence provides a more nuanced understanding of hedge funds. It turns out that there are two groups of hedge funds with respect to Edge. Most interestingly, the one group of hedge funds with Edge is able to produce positive skewness without sacrificing SR.

1.3.3.3 Portfolio Turnover and the Composition of Hedge Funds with Edge

The next step in better understanding the new Edge portfolio is detailing its portfolio turnover and composition. Figure 5 displays the annual turnover (%) of the Edge port-

[17]More specifically, hedge funds with Edge are identified as having a positive intercept that is significant at the 10% level and a non-positive slope coefficient loading.

folio which represents the number of hedge funds that are invested and/or redeemed each year (grey bars) based on whether or not a hedge fund has Edge. That is, a hedge fund is added to the Edge portfolio if it has Edge based on the parameter estimates at year-end in December and is redeemed if the parameter estimates no longer satisfy the Edge criteria. The red line corresponds to the total unique number of hedge funds within the Edge portfolio for a given year. In spite of the average annual turnover being relatively high at 58%, the total number of hedge funds is low ranging from 2 to 12 hedge funds forming the portfolio within a given year. The small number of funds makes this strategy more feasible to implement in practice.

Table 4 displays the details regarding the portfolio composition of each hedge fund portfolio exhibited in Figure 4. Panel A lists the primary hedge fund strategy (percentage of hedge fund months) by investment portfolio. Panel B displays key portfolio characteristics: AUM (USD mm), Fund Age (Years), Average Monthly Number of Funds, and the top three currency denominations (percent representation) within each portfolio. Panel A shows that the majority (73.4%) of fund strategies for the entire hedge fund universe is comprised of Fund of Funds (32.5%), Long/Short Equity Hedge (29.1%) and Multi-Strategy (11.8%). This composition changes to the top three hedge fund strategies now represented by Global Macro (49.3%), Fund of Funds (11.4%) and Options Strategy (8.9%) for those hedge funds with Edge. Global Macro's representation increases dramatically by 45.1% which increases the skewness of the portfolio given that global macro strategies tend to outperform during periods of market stress. In spite of the dominant presence of Global Macro, the other half of the Edge portfolio is spread relatively equally amongst a diverse group of strategies: Fund of Funds, Options Strategy, Multi-Strategy, Equity Market Neutral, Long/Short Equity Hedge and Fixed Income Arbitrage. Lastly, hedge funds without Edge are largely represented by Fund of Funds (32.6%), Long/Short Equity Hedge (29.3%) and Multi-Strategy (11.8%). This portfolio is highly similar to the equal-weighted portfolio based on the entire hedge fund universe.

Shifting our attention to Panel B of Table 4, only 3.23% (0.83%) of hedge funds each month have Edge based on AUM (number of funds). This small minority of hedge funds with Edge is a consequence of the fact that the majority of hedge funds are negative market timers and increase their positions on Short VIX at the worst possible time. In addition, hedge funds with Edge tend to be larger ($US 1.1 bn) in comparison to all hedge funds ($US 282mm) and those without Edge ($US 278 mm). Hedge funds with Edge charge higher management and performance fees relative to those without Edge, respectively (1.7% and 18.3% vs. 1.4% and 14.2%). Hedge funds with Edge have a relatively shorter track record of 11 years compared to the two other portfolios with track records of 13 years. In addition a similar proportion of hedge funds are closed to new investment each month for both those with and without Edge, respectively (4.5% vs. 5.1%). Finally, a larger proportion of hedge funds with Edge have a US Dollar denomination (86% vs. 71%).

1.3.3.4 Why do so few Hedge Funds have Edge?

The evidence presented thus far makes clear that only a small minority of hedge funds possess Edge. To help reconcile why only a small group of hedge funds with Edge exist in the sample, I examine the individual Edge parameter estimates (a and b) both cross-sectionally and in the time-series over the OOS period (May 2008 to December 2008).

Figure 6 displays the histograms of the two components that comprise the Edge measure: (i) the t-statistic for a; and (ii) the estimate of the b coefficient.[18] Panels A and B correspond to the t-statistic for a and estimate of b, respectively. Each histogram is based on the average value by hedge fund over the OOS period. With respect to Panel A, the histogram for a is relatively symmetric compared to the b coefficient which looks like a truncated distribution with respect to zero (i.e. almost all of the mass of the b coefficient's distribution lies in the domain of positive b coefficient estimates). This exercise makes clear that few hedge

[18]To aid with visual inspection, I have displayed the histogram of the b coefficient conditional on a having a minimum level of statistical significance of 10%. To further enhance the visual representation of the distribution of the b coefficient, I have dropped one hedge fund observation (for presentation purposes only) from Panel B which has an average b coefficient estimate of 2.44.

funds have an average b that is non-positive conditional on having an a that is significant at the 10% level. To provide context, Figure A3 shows the histograms for the unstandardized parameter estimates for a and b in Panels A and B, respectively. Both histograms are significantly more skewed. a has the most dramatic skewness of 18.05 whereas b has positive skewness of 2.94. The high degree of asymmetry of the unstandardized parameter estimates (especially a) makes it a noisier OOS measure of Edge. For this reason, I have chosen to base the Edge measure off of a positive a that is significant at the 10% level. Since the Edge measure with respect to b is based on a non-positive estimate there is no additional requirement concerning statistical significance. Moreover, b has a less skewed distribution, which indicates that it is less noisy compared to the unstandardized estimate of a.

Figure 7 shows the proportion of hedge funds (by number of hedge funds within a given month) that satisfy the Edge measure (blue line) next to its component parts (the grey line corresponds to funds with a positive a and the red line refers to hedge funds with a non-positive b). The left hand vertical axis corresponds to the grey line whereas the right vertical axis refers to the red and blue line. On average, 10% of hedge funds have a positive intercept with significant variation over time. At the beginning of the OOS period in May 2008, hedge funds have a positive intercept of 50% which declines to 6% in January 2009 reflecting the aftermath of the 2008 Financial Crisis. In contrast, only 3% of hedge funds have a nonpositive coefficient loading on Short VIX, ranging between 1% and 5%. As a consequence, 1% of hedge funds (by number of hedge funds) have Edge, on average. That is, the small minority of hedge funds with Edge is primarily driven by the fact that most hedge funds with positive a are negative market timers.

1.3.3.5 Performance Persistence in Hedge Funds with Edge

One of the main concerns with the small estimate (3%) of hedge fund managers with Edge is that it could simply be due to sampling error. To address this issue, I perform the following exercise where I compare the proportion of hedge funds identified as having Edge compared to those having Edge due to luck (based on a 95% confidence interval). Figure 8 displays the results from this exercise at longer horizons from 1-year (baseline specification) to 5-years. I assume hedge fund returns are independent and identically distributed each year. The proportion of hedge funds with Edge due to luck (red bars) shows the 5% identified in Year 1 declines exponentially towards zero at subsequent horizons. In contrast, hedge funds with Edge based on the main specification (grey bars) is highly stable ranging from 3% to 4% over longer horizons. This evidence mitigates any concerns that the Edge approach has simply identified hedge funds that were lucky ex-post. The next set of results examines the persistence in outperformance of those hedge funds with Edge.

If hedge fund managers have Edge (or skill), we would expect persistence in their superior performance. In order to examine this, Table 5 shows the performance summary statistics (SR and skewness) for equal-weighted hedge fund portfolios that are formed from investment parameters ranging from 1-year prior (the baseline specification in Figure 4) to 5-years prior. To provide context, Panel A forms a hedge fund portfolio using the traditional positive intercept criterion based on Equation 3. Panel B forms the hedge fund with Edge portfolio based on both positive intercept and non-positive slope coefficient. Intuitively, in both panels SRs decline as the hedge fund portfolio is based on parameter estimates with a longer lag from one-year to five-years. However, the SRs in Panel B are larger in magnitude compared to Panel A for all horizons (except two-years). Skewness remains significantly negative for all horizons in Panel A. In contrast, Panel B's positive skewness declines montonically beginning with the two-year lagged parameters and turns slightly negative (-0.11) with five-year lagged investment parameters. All together, Table 5 highlights how hedge funds with Edge are able to generate significant persistence in their returns (up to five years) which is consistent with having identified hedge fund managers with skill.

1.3.4 Hedge Fund Exposure to Higher Moment Risk and Macro Uncertainty

Higher moment risk (see for example Kelly and Jiang (2012)) and macroeconomic uncertainty (see for example Avramov et al. (2011)) have been shown to be fundamental drivers of the variation in the higher risk-adjusted returns of hedge funds. In light of the evidence presented thus far, one potential concern would be whether either (or both) of these fundamental drivers of hedge fund abnormal returns is able to explain the Edge portfolio's superior OOS performance. This section addresses this issue directly by testing whether hedge funds with Edge have significant exposure to either higher moment risks and/or macroeconomic uncertainty.

I adopt the empirical measures of higher moment risk from Gormsen and Jensen (2022). Specifically, I measure the ex ante (implied) and ex post (realized) moments of the S&P 500 using option prices. Table 6 reports the results from panel regressions of hedge fund excess returns on changes in the realized higher moments (Panel A) and changes in the implied higher moments (Panel B). The three columns of results from left to right are based on hedge funds belonging to the following groups respectively: (i) hedge funds without Edge; (ii) hedge funds with Edge; and (iii) hedge funds with a positive intercept. I estimate the following panel regression to test for higher moment risk exposure among hedge funds,

$$r_t^i = \alpha + a^i + b_t + \beta_1 \left(\Delta_t \text{Variance}\right) + \beta_2 \left(\Delta_t \text{Skewness}\right) + \beta_3 \left(\Delta_t \text{Kurtosis}\right) + \varepsilon_t^i, \tag{6}$$

with hedge fund and year fixed effects and standard errors clustered at the hedge fund level. r_t^i are hedge fund excess returns, a^i is a hedge fund-specific fixed effect and b_t is a year fixed effect. The monthly changes of the higher order moments are computed by taking first differences, $\Delta_t = \text{Higher Moment}_t - \text{Higher Moment}_{t-1}$.

There are three important takeaways from Table 6. First, hedge funds with Edge have no exposure to both implied and realized higher moment risk at any conventional level of statistical significance. Second, hedge funds without Edge have significant exposure to all higher order moments of risk at the 1% level and are negatively impacted, on net, by higher moment risk. Hedge funds without Edge are negatively impacted by realized variance, kurtosis and implied variance and skewness while benefiting from realized skewness and implied kurtosis. The moments that have a negative effect have larger relative magnitudes. Third, hedge funds with a positive intercept have exposure to higher moment risk that is highly similar to those funds without Edge. The important difference to note is that hedge funds with a positive intercept have a significantly larger proportion of their returns explained by higher moment risk with an R^2 of 21% (24%) for realized (implied) moments. That is, a significant portion of hedge funds generate a positive intercept by loading up on exposure to higher moment risk which is ultimately borne out in higher SRs at the expense of negative skewness. This empirical evidence is consistent with the empirical evidence that documents how hedge fund returns are driven by tail risk (see for example Kelly and Jiang (2012)). In contrast, hedge funds with Edge are able to generate returns from some other source that is independent of higher moment risk. Hence, this additional piece of evidence is consistent with the idea that Edge is able to better reflect hedge fund manager skill (i.e. the ability to generate uncorrelated returns).

Table 7 reports the results from the same panel regression estimation as in Table 6 (over the full OOS period) except the regressors are replaced by macroeconomic variables from Avramov et al. (2011); Fama and French (1989) to reflect macroeconomic uncertainty. The macroeconomic regressors include the following set: (i) Δ_tDividend Yield; (ii) Δ_tDefault Spread; (iii) Δ_tTerm Spread; (iv) Δ_t3-Month Treasury Bill; and (v) Δ_tVIX Spread. Most importantly, the high SRs and positive skewness of hedge funds with Edge cannot be explained by exposure to macroeconomic uncertainty (R^2 of 2% and an intercept of 1.8% per month that is significant at the 5% level). This result is especially surprising given that nearly half of hedge funds with Edge are represented by Global Macro hedge funds which are positioned to benefit from periods of elevated macroeconomic uncertainty. The other key takeaway from this table is that hedge funds without Edge and those with only a positive intercept have an R^2 of 11% and 23%, respectively. This evidence is consistent with Avramov et al. (2011) who show the significant explanatory power of hedge fund performance using macroeconomic variables. In particular, hedge funds with only a positive intercept are producing higher returns by gaining exposure to macroeconomic uncertainty.

1.4 Why Does Edge Perform Well as an OOS Measure?

The previous section showcases the efficacy of using the newly proposed Edge measure in identifying hedge fund managers with both higher SRs and positive skewness, OOS. This section investigates why this new measure, Edge, has such attractive OOS properties.

1.4.1 The Factor Structure of Hedge Fund Returns

Giglio et al. (2021) show empirically the extent of omitted variable bias (OVB) present in hedge fund returns whereby the benchmark model is the Fung and Hsieh (2001) seven factor model.[19] In particular, they document how there remains a strong factor structure in hedge fund returns after controlling for the FH8 factors which indicates that there is a significant OVB problem in hedge fund returns. To address this omitted factor issue they perform an eigenvalue decomposition via PCA on the FH8 residuals variance-covariance matrix. With that, they then add the leading three PCs to the FH8 model to account for the omitted factors.

Adding the leading PCs from the residuals covariance matrix neccessarily assumes that the leading PCs will account for the OVB both IS and OOS. Said differently, the robustness of this approach relies on a strong factor structure amidst the hedge fund return residuals

[19]The Emerging Market Risk Factor has been recently proposed to be added to this factor model, which I refer to as FH8 throughout this paper.

that can be adequately summarized by the leading PCs. Figure 9 displays the percentage of variance explained by the leading 15 PCs. As in Giglio et al. (2021), I first apply matrix completion to the respective factor model (i.e. FH8) residuals, via the nuclear-norm penalization algorithm, and then conduct PCA on the recovered matrix. The PCs are estimated over the OOS period between May 2007 and December 2022.[20] The blue line is based on the full period from 2007 to 2022, whereas the red line is estimated using only months in which the Short VIX benchmark return is negative over this period.[21] A negative Short VIX return is intended to proxy for a market dislocation (or increase in systematic risks) since the Short VIX return is only negative when either the VIX futures curve rises or inverts into backwardation.[22] Both of these circumstances represent a rise in implied volatility thereby indicating a rise in uncertainty in financial markets (i.e. increases the likelihood of deviations between market prices and fundamental values thereby representing a market dislocation).

The main takeawaway from Figure 9 is that the strong factor structure in the hedge fund return residuals *weakens* during market dislocations. Panel A shows the blue line (full period) is on top of the red line (market dislocations) for the first three PCs and then stays below the red line from PCs 4 to 15. This is clear evidence that the variance in hedge fund returns explained by the first few PCs shifts to the lower ranked PCs in periods of market stress. More precisely, Panel B shows that the variance explained by PCs 1-3 for the full period is nearly cut in half during market dislocations (63% to 35%). Hence, an approach that uses the leading PCs to account for omitted factors will inevitably fail to account for a significant portion of variation during market dislocations. Given the asymmetry in financial market returns this failure during market dislocations is the likely source of negative skewness

[20]This sample period covers the full estimation window of the OOS period in this paper. It is important to note that the OOS period mentioned previously begins in May 2008 to capture 1-year ahead returns that was based on estimation up until the previous year.

[21]The Short VIX has negative monthly returns one third of the time (63 out of 188 months) from May 2007 to December 2022.

[22]Pasquariello (2014) creates a model-free measure of financial market dislocations that is comprised of an equal-weighted average of violations of three no-arbitrage conditions (Covered Interest Rate Parity, Triangular Arbitrage Parity and American Depositary Receipt Parity). I have instead chosen to define dislocations with the negative returns from the Short VIX since it is by definition directly related to the benchmark.

of hedge fund returns that have been identified as having "skill" based on a positive and statistically significant intercept. In the absence of an efficient benchmark (Dybvig and Ross, 1985) any measure of hedge fund skill solely based on a positive intercept is likely to suffer from poor OOS performance characterized by negative skewness given that a positive intercept fails to account for the weaker factor structure during dislocations.

1.4.2 Negative Skewness and the Lower Ranked PCs

The natural next step is to verify whether or not the lower ranked PCs ultimately have any influence on hedge fund returns (i.e. resulting in negative skewness). Figure 10 shows the rolling 12-month pairwise correlations between the sum of PCs 10-15 from the FH8 residuals covariance matrix and three different return series from April 2009 to December 2022. Panels A, B and C correspond to the Short VIX returns, hedge funds without Edge portfolio returns and hedge funds with Edge portfolio returns, respectively. The hedge fund portfolio returns is the same set of returns in the main specification as displayed in Figure 4. Each panel provides a window into how these lower ranked PCs show up as negative skewness in return series. In Panel A, the correlation between Short VIX and PCs 10-15 is 0.46 and 0.43 in August and September 2009, respectively which coincides with the large negative returns experienced one-year prior in September and October 2008 due to the Financial Crisis. This high correlation between the two time series of returns subsides and is not exceeded until April 2020 with a correlation of 0.47 which rises to as high as 0.73 in December 2020. In short, the correlation between the lower ranked PCs and the Short VIX increases during market dislocations. This evidence provides further evidence that the Short VIX is able to capture higher moment risk that spills into the lower ranked PCs during crises.

Having established the significant comovement between the Short VIX benchmark and lower ranked PCs, the next step is to establish the relationship between the lower ranked PCs and the OOS hedge fund portfolio returns. Figure 10 Panel B and C displays the pairwise correlation between the lower ranked PCs and the portfolio returns of hedge funds without Edge and with Edge, respectively. If we focus our attention on the two major market dislocations during the OOS period (2008 Financial Crisis and 2020 Covid Crash) we see one clear contrasting effect between the two panels. In Panel B, the correlation between hedge funds without Edge and the lower ranked PCs ranges from 0.45 in April 2009 to 0.18 in September 2009. In sharp contrast, the correlation between hedge funds with Edge and the lower ranked PCs ranges from 0.17 in April 2009 to -0.24 in September 2009. The correlations during this period span the depths of the 2008 Financial Crisis in September and October 2008. The contrasting correlation between the lower ranked PCs and hedge funds with and without Edge can explain why hedge funds without Edge suffer significant losses relative to those with Edge during the 2008 Financial Crisis (see Figure 4). That is, the correlations increase at the outset of a dislocation and remain significantly positive over the dislocation between the lower ranked PCs and hedge funds without Edge. In contrast, the correlations are low at the outset of a dislocation and turn negative over the course of the dislocation between the lower ranked PCs and hedge funds with Edge.

The second major dislocation represented by the 2020 COVID-19 Crash illustrates the same contrasting relationship between the two groups of hedge funds. In Panel B of Figure 10, the correlation between hedge funds without Edge and the lower ranked PCs ranges from -0.42 in December 2019 to 0.63 in December 2020. Again, in contrast, the correlation between hedge funds with Edge and the lower ranked PCs ranges from 0.02 in December 2019 to -0.08 in December 2020. In sum, the correlation between the lower ranked PCs and the hedge funds without Edge increases dramatically to a large positive correlationn during market dislocations whereas the correlation is near zero at the outset of the crisis and decreases over the subsequent year with respect to hedge funds that possess Edge. All together, Figure 10 provides evidence that the negative skewness observed in hedge funds without Edge enters through the lower ranked PCs as negative residuals. This evidence highlights how the lower ranked PCs have a disproportionate negative effect on hedge funds without Edge during market dislocations.

1.4.3 How Does the Edge Measure Mitigate the Issue of the Lower Ranked PCs?

The evidence presented thus far in this section has documented two empirical details: (i) the factor structure weakens in hedge fund returns during market dislocations; and (ii) the lower ranked PCs have a high degree of positive comovement with the Short VIX benchmark and hedge funds without Edge during major market dislocations. However, it is still not clear how the Edge measure is able to overcome the issue of the factor structure weakening during dislocations. The final set of empirical results of this section helps shed light on bridging this gap in our understanding.

Figure 11 displays SRs (blue bars) and Skewness (red bars) from various equal-weighted hedge fund portfolios. In this empirical exercise all hedge fund portfolios have been estimated with the leading three PCs of their respective residuals covariance matrix following Giglio et al. (2021). That is, the left hand side (Panels A and C) correspond to the FH8 Model with the addition of the three leading PCs. The right hand side (Panels B and D) correspond to the main specification of this paper (Equation 3) with the addition of the leading three PCs from this model's residuals' covariance matrix. The top row forms a hedge fund portfolio by selecting one with a stictly positive intercept whereas the bottom row forms a hedge fund portfolio by selecting a hedge fund with both a strictly positive intercept and non-positive coefficient on the benchmark. In the case of the FH8 model I have used the coefficient on the market factor as a proxy for the coefficient on the Short VIX benchmark. Each panel varies in the level of significance of the positive intercept ranging from no significance (i.e. only the sign of the coefficient matters) to statistical significance at the 1% level. The top row shows a distinct pattern that is consistent with Back et al. (2018): SRs (skewness) increase (decrease) as we increase the level of significance of the positive intercept. In spite of having controlled for the OVB proxied by the leading PCs, there is still the important tradeoff between higher SRs and negative skewness. It is important to also note that the Short VIX benchmark Model (Panel B) results in higher SRs relative to FH8 (Panel A) for all hedge fund portfolios except those that do not necessarily have a significant intercept.

The second important finding from this empirical exercise is that the bottom row displays hedge fund portfolios with significantly less negative skewness or significant positive skewness. To be clear, the bottom row builds hedge fund portfolios based on the Edge measure (i.e. dual requirement of a significantly positive intercept and non-positive coefficient on the benchmark). By conditioning on those hedge funds with a non-positive coefficient on either the market factor or Short VIX, the return series do not suffer from the same level of negative skewness. The only difference between the hedge funds in the top and bottom row is the additional requirement of a non-positive coefficient on the benchmark. This additional requirement is ultimately why the Edge measure has such attractive OOS properties with positive skewness. By definition, the requirement of a non-positive coefficient removes all hedge funds with a strictly positive coefficient on the benchmark. Assuming the benchmark is a proxy for systematic risks, these hedge funds that are removed all load positively on systematic risks that positively covary with one another. Given that systematic risks increase during a market dislocation, this systematic risks spills over into the lower ranked PCs. Hence, hedge funds that load positively on systematic risks are those hedge funds that load positively on the lower ranked PCs. This group of hedge funds (without Edge) end up with negative skewness from their positive exposure to both the benchmark and the lower ranked PCs. By removing this problematic group of hedge funds, the Edge measure is no longer vulnerable to the weakening of the factor structure during market dislocations. By avoiding increases in market downside risks, the Edge measure identifies hedge funds with a positive intercept that do not suffer from negative skewness.

The bottom row of Figure 11 also makes clear that the Edge measure (based on Short VIX), Panel D, performs significantly better than its counterpart using the FH8 model in Panel C. In spite of this improvement of skewness, Figure 11 also shows how the portfolio returns are less attractive compared to the main specification in this paper that does not rely on estimating latent factors. This should not be too surpising given the small number of hedge funds that have been determined to possess Edge in this paper. By adding the leading PCs from the residual covariance matrix, the coefficient estimates have higher variance leading to more unstable OOS performance. This set of empirical results highlights the benefit of a more parsimonious model that is easier to estimate. Moreover, it illustrates the effectiveness of how this new measure, Edge, is able to overcome the issue of OVB that has made it particularly difficult to predict hedge fund performance in previous studies.

1.5 Conclusion

I develop a new measure of hedge fund manager skill that predicts hedge fund performance OOS by addressing a new form of hedge fund model misspecification (weak latent factors). I document a novel empirical finding regarding hedge fund returns: The consensus view that hedge fund returns can largely be explained by offering investors concentrated exposure to market crashes is incomplete and too coarse a representation of hedge funds. This study presents evidence that shows hedge funds can instead be divided into two groups, ex-ante: one group, with Edge, that produces high SRs and positive skewness and another, without Edge, with relatively low SRs and negative skewness.

The new Edge methodology that allows us to predict OOS hedge funds with both higher SRs and positive skewness suggest we revisit how to characterize hedge fund performance and the role of weak latent factors in asset pricing. How is it possible that hedge funds can offer both higher SRs and positive skewness without exposure to either higher moment risks or macroeconomic uncertainty? This paper highlights an understudied area in the hedge fund literature as it relates to hedge funds that offer high SRs and positive skewness. In particular, future work needs to address two important questions: First, why do so few hedge funds have Edge? Second, why are hedge fund investors willing to allocate to hedge funds with negative skewness when they can instead invest in hedge funds that offer both higher SRs and convex payoffs? Tackling this line of inquiry will likely lead to a deeper understanding of both hedge funds and investor behavior.

1.6 Tables and Figures

Table 1: Summary Statistics of Hedge Fund Returns and Hedge Fund Substitutes
The table presents the summary statistics for the HFRI 500 Index, TASS equal-weighted portfolio, Short Fung and Hsieh (2001) Stock Index Lookback Straddle (PTFSSTK), Jurek and Stafford (2015) OTM S&P 500 Put-Writing Strategy and Short VIX strategy. All summary statistics (expressed in decimal form) are annualized from a monthly frequency and calculated over the sample period from January 2005 to December 2022.

	HFRI 500	TASS	Short Straddle	Short Put	Short VIX
Average	0.04	0.02	0.47	0.13	0.48
Volatility	0.06	0.11	0.57	0.16	0.52
SR	0.61	0.21	0.82	0.85	0.91
Skewness	−1.18	−0.81	−2.44	−2.26	−1.26
Correlation	1.00	0.97	0.40	0.74	0.72

Table 2: Short VIX and the Treynor and Mazuy (1966) Market-Timing Model
The table reports the results from the market timing test of Treynor and Mazuy (1966):
$r_{\mathrm{SVIX},t} = a + b\left(r_{m,t} - r_{f,t}\right) + c\left(r_{m,t} - r_{f,t}\right)^2 + u_t$
The sample period is April 2004 to December 2022. Coefficient estimates (in decimal form) of a, b and c are reported along with their respective Newey and West (1987) t-statistics and R^2.

	$\hat{a}$	$\hat{b}$	$\hat{c}$	R^2
Estimate	0.0397	2.470	-8.118	0.62
t-stat	5.52	17.97	-4.65	

Table 3: Panel Regressions (Full Sample)
The table reports the results from the panel regressions of the following Equation: $r_t^i = a + h^i + b\,(\text{Benchmark}_t) + c\,(\text{Benchmark}_t)^2 + \varepsilon_t^i$, for each hedge fund strategy with hedge fund fixed effects, where h^i is a hedge fund-specific fixed effect, a is unexplained return variation, Benchmark refers to the monthly returns from the Short VIX$_t$ in Panel A and OTM Short Put$_t$ in Panel B, b is the coefficient loading on the benchmark and c indicates hedge fund exposure to timing the benchmark. The sample period is May 2004 to December 2022 at a monthly frequency. Coefficient estimates (expressed in decimal form) of a, b and c along with their respective t-statistics and sample size are also reported. Standard errors are clustered at the hedge fund level. Asterisks denote the levels of statistical significance of the b coefficient: 10% level (*), 5% level (**) and 1% level (***).

Panel A: Short VIX as the Hedge Fund Benchmark

Hedge Fund Strategy	$\hat{a}$	t-stat	$\hat{b}$	t-stat	$\hat{c}$	t-stat	R^2
Convertible Arbitrage***	0.002	1.66	0.161	6.23	−0.236	−8.91	0.19
Dedicated Short Bias***	0.015	2.66	−0.367	−3.91	−0.051	−0.33	0.34
Emerging Markets***	0.000	-0.85	0.267	39.06	−0.255	−22.70	0.23
Equity Market Neutral***	0.001	3.40	0.077	11.08	−0.025	−2.55	0.10
Event Driven***	0.002	8.75	0.173	30.11	−0.145	−17.93	0.34
Fixed Income Arb***	0.003	10.55	0.075	18.71	−0.129	−8.02	0.18
Fund of Funds***	−0.002	−22.23	0.106	72.48	−0.072	−28.43	0.06
Global Macro***	0.003	7.14	0.038	4.41	−0.002	−0.11	0.01
L/S Equity Hedge***	−0.003	−16.53	0.218	60.78	−0.037	−6.46	0.29
Managed Futures	0.003	1.35	−0.008	−0.59	−0.037	−0.39	0.00
Multi Strat.***	0.000	−0.99	0.106	30.30	−0.073	−8.46	0.16
Options Strategy*	0.006	4.67	0.032	1.71	−0.171	−4.03	0.02
All Strategies***	−0.001	−11.43	0.144	77.32	−0.078	−28.57	0.13

Panel B: OTM Short Put as the Hedge Fund Benchmark

Hedge Fund Strategy	$\hat{a}$	t-stat	$\hat{b}$	t-stat	$\hat{c}$	t-stat	R^2
Convertible Arbitrage***	−0.003	−2.52	0.680	6.57	−0.343	−1.62	0.24
Dedicated Short Bias***	0.022	5.91	−1.501	−6.17	−2.980	−3.69	0.46
Emerging Markets***	−0.011	−21.31	1.143	38.40	0.913	10.24	0.27
Equity Market Neutral***	0.000	−0.85	0.254	8.41	0.444	5.79	0.07
Event Driven***	−0.002	−6.35	0.660	27.98	0.411	6.20	0.35
Fixed Income Arb***	0.002	7.08	0.265	16.13	−0.550	−5.05	0.20
Fund of Funds***	−0.004	−30.26	0.389	52.90	0.273	9.35	0.06
Global Macro***	0.001	1.64	0.222	7.73	0.706	8.46	0.03
L/S Equity Hedge***	−0.007	−32.61	0.870	57.80	1.491	37.09	0.33
Managed Futures**	−0.005	−4.15	0.101	2.03	2.039	3.15	0.02
Multi Strat.***	−0.003	−11.68	0.445	29.81	0.539	5.32	0.16
Options Strategy**	0.005	5.11	0.170	2.33	−0.090	−0.50	0.02
All Strategies***	−0.005	−36.86	0.574	71.45	0.676	28.49	0.14

Table 4: Hedge Fund Portfolio Composition
The table presents the portfolio characteristics for the three equal-weighted portfolios from Figure 4: All Hedge Funds, Hedge Funds with Edge and Hedge Funds without Edge, whereby Edge is defined by a positive intercept and nonpositive coefficient loading with respect to the Short VIX benchmark (main specification). Panel A displays the primary hedge fund strategy (% of hedge fund months) by investment portfolio. Panel B displays the following investment portfolio characteristics by investment portfolio: AUM ($ US mm), Fund Age (Years), Number of Funds (Mean), Management Fee (%), Performance Fee (%), Closed to New Investment (%), and the top three currency denominations (percentage of representation) in each portfolio. The sample period is May 2008 to December 2022.

Panel A: Primary Hedge Fund Strategy by Investment Portfolio

	All	HFs with Edge	HFs without Edge
Convertible Arbitrage	1.4	0.0	1.4
Dedicated Short Bias	0.1	0.0	0.1
Emerging Markets	7.3	1.1	7.4
Equity Market Neutral	3.3	7.1	3.2
Event Driven	6.7	0.0	6.8
Fixed Income Arbitrage	2.5	5.2	2.5
Fund of Funds	32.5	11.4	32.6
Global Macro	4.2	49.3	3.8
Long/Short Equity Hedge	29.1	6.0	29.3
Managed Futures	0.5	3.4	0.5
Multi-Strategy	11.8	7.5	11.8
Options Strategy	0.8	8.9	0.7

Panel B:Investment Portfolio Characteristics

	All	HFs with Edge	HFs without Edge
AUM (USD mm)	282	1,092	278
Fund Age (Years)	13	11	13
Number of Funds (Average)	720	6	714
Management Fee (%)	1.4	1.7	1.4
Performance Fee (%)	14.2	18.3	14.2
Closed to New Investment (%)	5.1	4.5	5.1
Top Three Curr. Denom. (%)			
US Dollar	71	86	71
Euro	17	6	17
Swiss Franc	4	1	5

Table 5: Multi-Year Ahead Performance
The table reports the results from estimating rolling 3-year regressions (with Newey and West (1987) standard errors) of Equation 3,
$r_t - r_{f,t} = a + b\left(r_{SVIX,t}\right) + c\left(r_{SVIX,t}\right)^2 + \varepsilon_t$,
Both panels display the SR and skewness of an equal-weighted hedge fund portfolio that is formed from parameter estimates that range from 1-year prior (as in Figure 4) to 5-years prior. Panel A selects hedge funds based on a positive intercept whereas in Panel B hedge funds are chosen based on the Edge measure (positive intercept and non-positive slope coefficient). The OOS period begins in May 2008 (1-year ahead returns) with all periods ending in December 2022, at a monthly frequency.

Panel A: Positive Intercept as the Performance Measure

	SR	Skewness
One-Year Ahead Returns	0.34	-1.39
Two-Year Ahead Returns	0.65	-0.51
Three-Year Ahead Returns	0.47	-1.01
Four-Year Ahead Returns	0.16	-0.40
Five-Year Ahead Returns	0.11	-1.91

Panel B: Hedge Funds with Edge

	SR	Skewness
One-Year Ahead Returns	0.53	0.45
Two-Year Ahead Returns	0.56	2.47
Three-Year Ahead Returns	0.77	1.12
Four-Year Ahead Returns	0.56	0.89
Five-Year Ahead Returns	0.35	-0.11

Table 6: Hedge Fund Exposure to Higher Moment Risk (OOS Period)
The table reports the results from the panel regressions of the following Equation: $r_t^i = \alpha + a^i + b_t + \beta_1 (\Delta_t \text{Variance}) + \beta_2 (\Delta_t \text{Skewness}) + \beta_3 (\Delta_t \text{Kurtosis}) + \varepsilon_t^i$, for each hedge fund strategy with hedge fund and year fixed effects, where r_t^i are hedge fund excess returns, a^i is a hedge fund-specific fixed effect, b_t is a year fixed effect and α is unexplained return variation. In Columns 2 to 4, the panel regressions are with respect to hedge funds without Edge, with Edge and hedge funds with a positive intercept, respectively. Edge is based on the main specification (positive alpha and nonpositive beta based on the Short VIX benchmark). In Panel A, the panel regression is based on the changes in realized higher order moments (variance, skewness and kurtosis). In Panel B the panel regression is with respect to the changes in implied higher order moments. The monthly changes of the higher order moments are computed by taking first differences, $\Delta_t = \text{Higher Moment}_t - \text{Higher Moment}_{t-1}$. The intercept for each panel regression and the respective higher order moment coefficient loading $\beta_1, \beta_2, \beta_3$ are reported (in decimal form) along with their respective t-statistics. Overall R^2, sample and cluster size are also reported. Standard errors are clustered at the hedge fund level. The sample period is from May 2008 to December 2021.

Panel A: Realized Moments

	HFs without Edge	HFs with Edge	HFs with $\hat{\alpha} > 0$
Intercept	−0.031	0.016	−0.021
t-stat	−16.76	2.23	−7.15
Δ Realized Variance	−2.320	0.030	−1.994
t-stat	−45.97	0.14	−17.68
Δ Realized Skewness	−0.010	0.000	−0.009
t-stat	−22.76	−0.52	−8.08
Δ Realized Kurtosis	−0.002	0.000	−0.002
t-stat	−15.90	0.05	−5.81
R^2	0.08	0.01	0.21

Panel B: Implied Moments

	HFs without Edge	HFs with Edge	HFs with $\hat{\alpha} > 0$
Intercept	−0.014	0.014	−0.012
t-stat	−7.33	1.98	−4.14
Δ Implied Variance	−0.003	0.000	−0.003
t-stat	−48.75	−0.73	−22.43
Δ Implied Skewness	0.029	−0.007	0.015
t-stat	25.13	−0.97	7.29
Δ Implied Kurtosis	0.001	0.000	0.001
t-stat	23.73	−0.94	6.95
R^2	0.12	0.01	0.24
No. of Clusters	2,743	52	761
N	119,870	1,009	11,083

Table 7: Hedge Fund Exposure to Macroeconomic Uncertainty (OOS Period)
The table reports the results from the panel regressions of the following Equation: $r_t^i = \alpha + a^i + b_t + \beta_1\,(\Delta_t\text{Dividend Yield}) + \beta_2\,(\Delta_t\text{Default Spread}) + \beta_3\,(\Delta_t\text{Term Spread}) + \beta_4\,(\Delta_t\text{3-Month T-Bill}) + \beta_5\,(\Delta_t\text{VIX Spread}) + \varepsilon_t^i$, for each hedge fund strategy with hedge fund and year fixed effects, where r_t^i are hedge fund excess returns, a^i is a hedge fund-specific fixed effect, b_t is a year fixed effect and α is unexplained return variation. The macroeconomic variables correspond to the same set from Fama and French (1989) with the VIX Spread added from Avramov et al. (2011). In Columns 2 to 4, the panel regressions are with respect to hedge funds without Edge, with Edge and hedge funds with a positive intercept, respectively. Edge is based on the main specification (positive alpha and nonpositive beta based on the Short VIX benchmark). The monthly changes of the macroeconomic variables are computed by taking first differences, $\Delta_t = \text{Macro Variable}_t - \text{Macro Variable}_{t-1}$. The intercept for each panel regression and the respective macroeconomic variable coefficient loadings β_1 to β_5 are reported (in decimal form) along with their respective t-statistics. Overall R^2, sample and cluster size are also reported. Standard errors are clustered at the hedge fund level. The sample period is from May 2008 to December 2022.

	HFs without Edge	HFs with Edge	HFs with $\hat{\alpha} > 0$
Intercept	−0.013	0.018	−0.007
t-stat	−6.87	2.16	−2.49
Δ Dividend Yield	−5.923	−0.588	−3.045
t-stat	−43.12	−0.63	−13.03
Δ Default Spread	−0.238	−2.372	−0.645
t-stat	1.04	−1.68	−1.50
Δ Term Spread	0.086	0.112	−0.135
t-stat	2.14	0.52	−2.16
Δ 3-Mo. T-Bill	7.011	−4.281	5.010
t-stat	28.42	−1.43	12.14
Δ VIX Spread	−0.002	0.000	−0.001
t-stat	−40.37	−1.53	−18.63
R^2	0.11	0.02	0.23
No. of Clusters	2,793	55	792
N	125,634	1,067	11,806

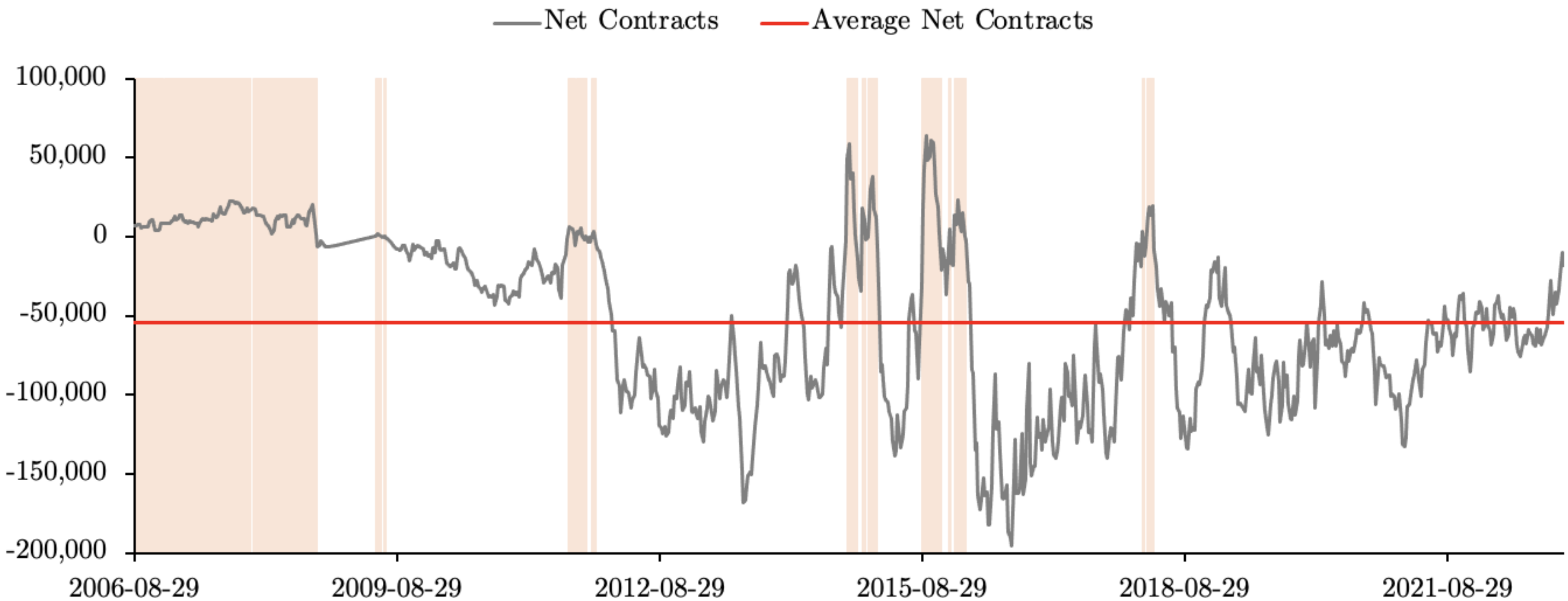


Figure 1: Net Long CBOE VIX Futures Positions of Leveraged Hedge Funds
This figure plots the weekly aggregate net position (long minus short) of the leveraged funds group in the CFTC Commitments of Traders in Financial Futures (TFF) report from August 29, 2006 to December 27, 2022. There is a gap in reporting from January 2009 to May 2009. The average net long position over the sample period is depicted by the red line and the shaded regions indicate periods when the net long position is positive. Units are in thousands of contracts (net notional position in millions).

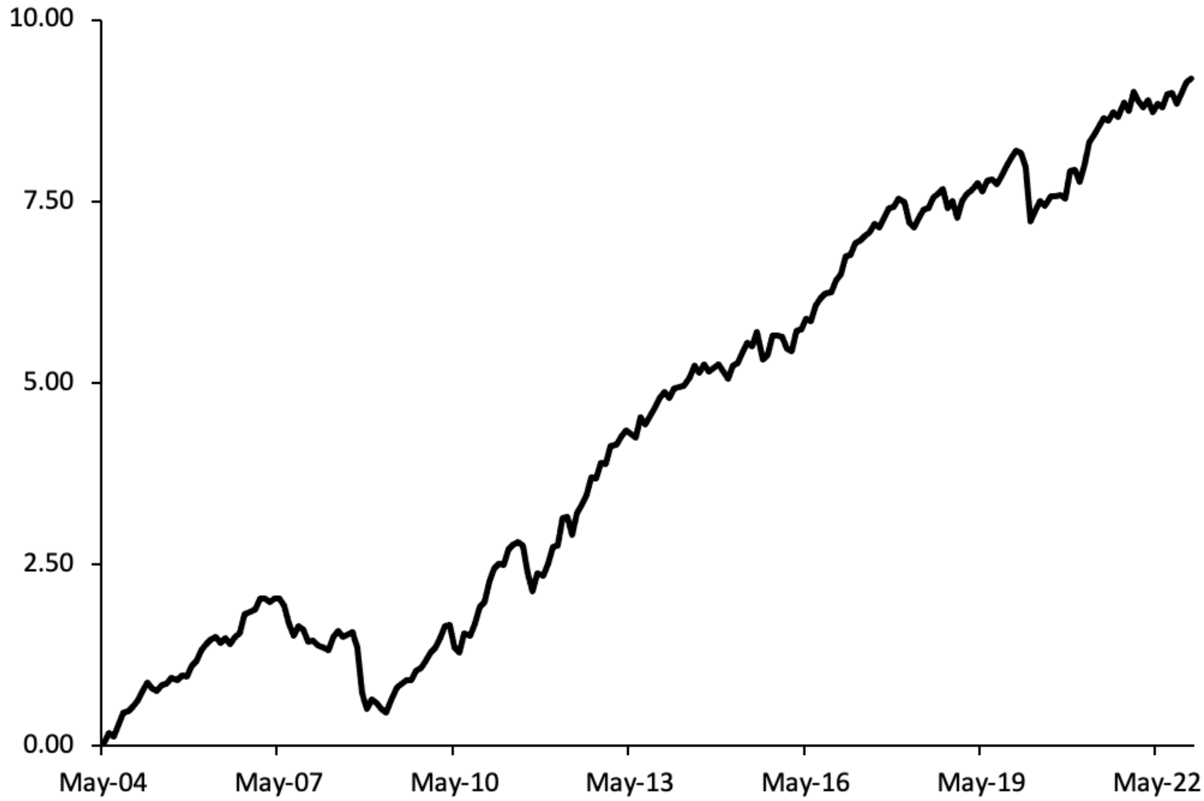
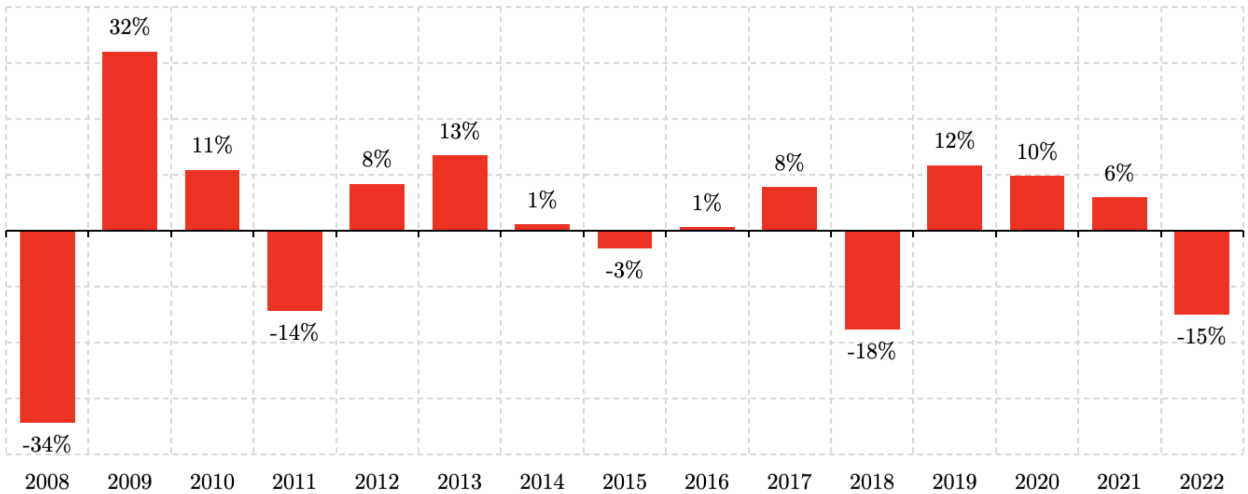
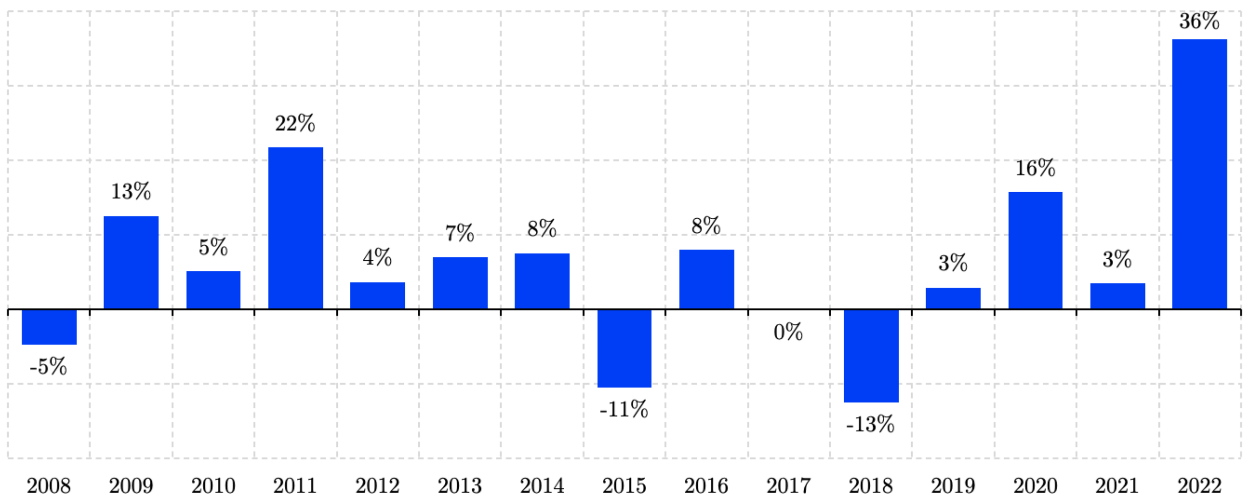


Figure 2: Short VIX Benchmark Cumulative Returns
The figure displays the cumulative monthly returns of the Short VIX futures strategy. The sample period is from May 2004 until December 2022.



(a)



(b)

Figure 3: Annual OOS Returns for Hedge Funds with(out) Edge
The figure shows the annual returns for the OOS period: May 2008 to December 2022 based on three-year rolling regressions of Equation 3. Panel (a) displays returns (red bars) from an equal-weighted portfolio of hedge funds without Edge. Panel (b) displays returns (blue bars) from an equal-weighted portfolio of hedge funds with Edge. Edge is based on the main specification of hedge funds with a positive intercept and nonpositive coefficient loading with respect to the Short VIX benchmark.

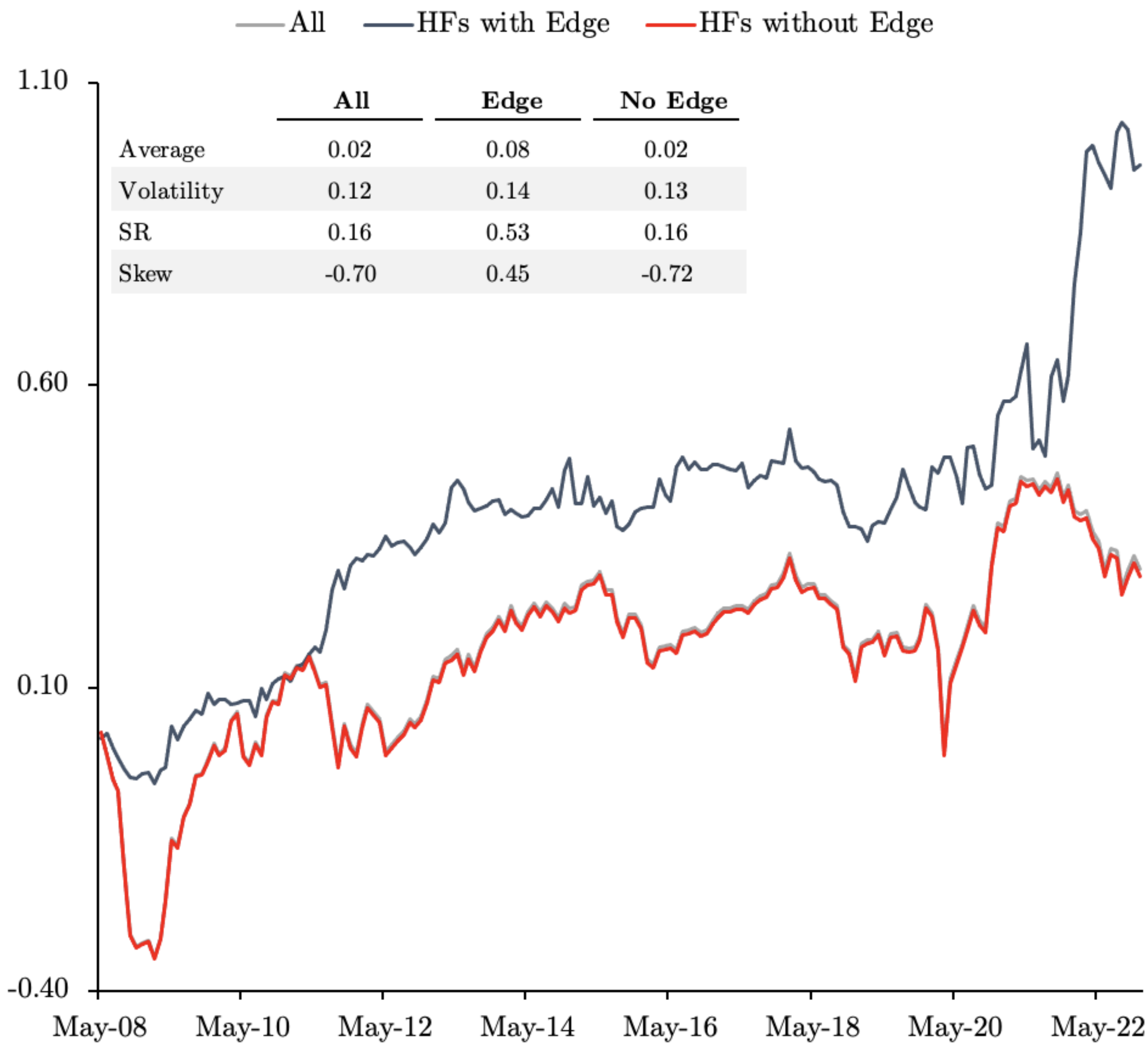


Figure 4: Cumulative Portfolio Returns (OOS)
The figure displays the cumulative returns to three equal-weighted portfolios of hedge funds. The grey line is comprised of all hedge funds in the TASS Database. The blue line is a portfolio of hedge funds with Edge ($\hat{a} > 0$ and $\hat{b} \leq 0$). The red line is a portfolio of hedge funds without Edge. The parameters that are the basis for the Edge methodology are estimated by 3-year rolling regressions of Equation 3 with Newey and West (1987) standard errors. The Edge portfolios have an annual rebalancing that takes place each December. Hedge funds are selected based on these December parameter estimates and then held over the subsequent year. The sample is from May 2008 until December 2022 to account for the initial estimation window.

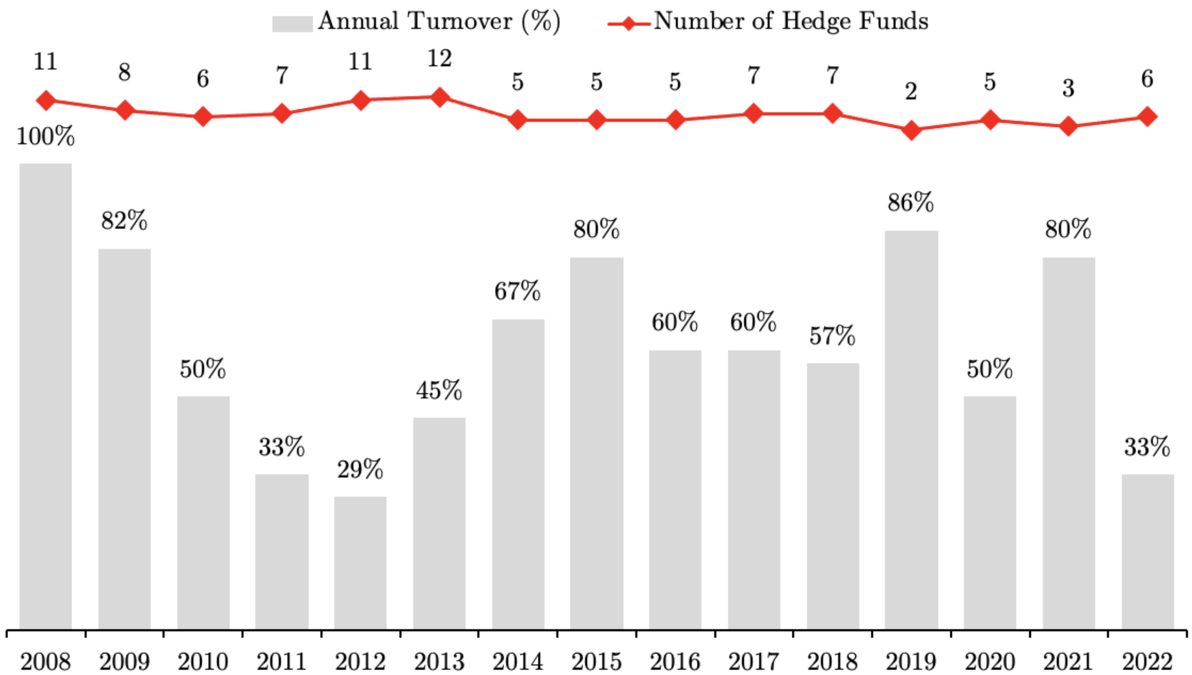
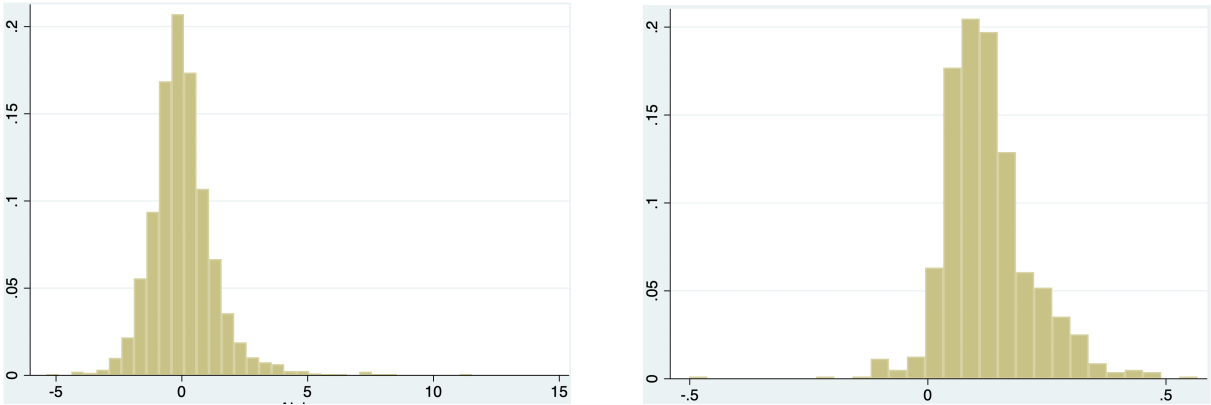


Figure 5: Annual Portfolio Turnover (%)
The figure displays the percentage of hedge funds (grey bars) that are invested and/or redeemed each year based on whether or not a hedge fund has Edge. In addition, the total unique number of hedge funds in the Edge portfolio is displayed (red line). The sample is from May 2008 until December 2022 to account for the initial estimation window.



(a) $\hat{a}_{t\text{-stat}}$ (b) $\hat{b} \mid \hat{a}_{t\text{-stat}} \geq 1.65$

Figure 6: Edge Parameter Histograms
The figure shows the histograms of the respective parameter that comprises the Edge measure. The vertical axis represents the fraction of observations (in decimal form) that fall within the respective bins. Panel A and B display the histogram of the t-statistics for the a estimate and the estimate of the b coefficient (conditional on a having a minimum level of statistical significance of 10%), respectively. Each histogram is based on the average value (t-statistic or estimate of the parameter), by hedge fund, over the OOS period (May 2008 to December 2022). For presentation purposes only, I have dropped one hedge fund observation from Panel B which has an average b coefficient estimate of 2.44. In doing so, the shape of the distribution for the b coefficient is easier to evaluate.

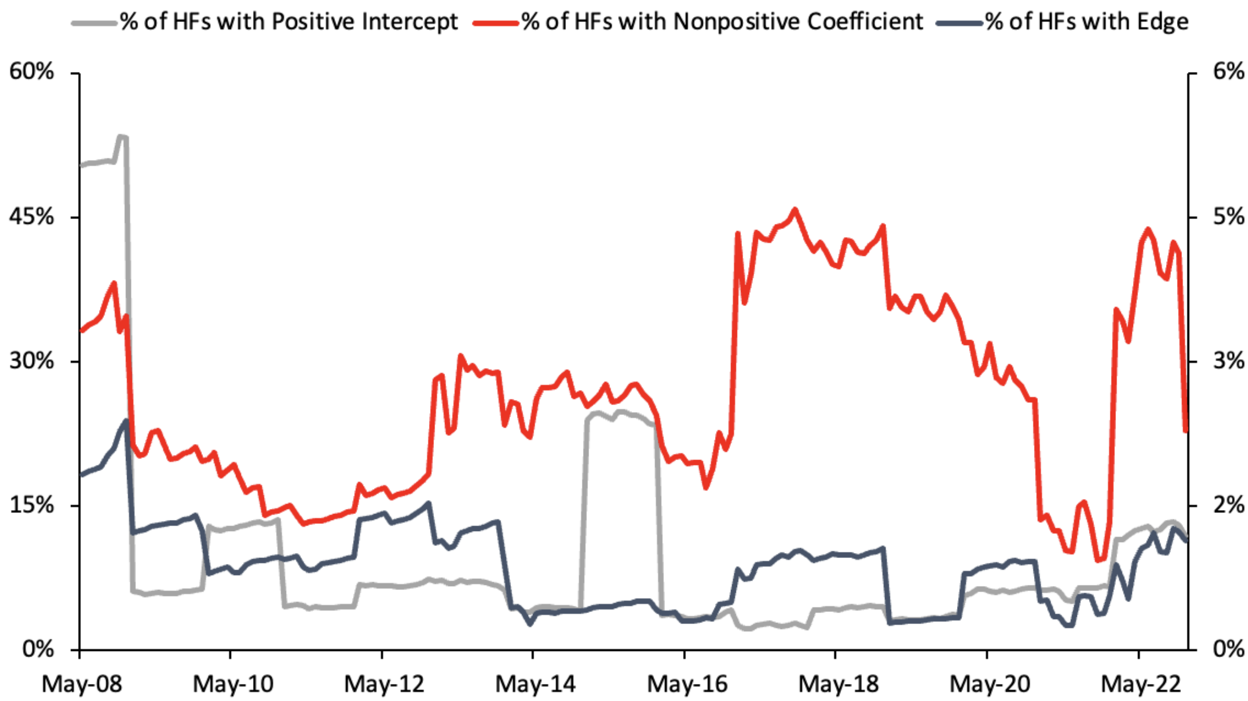


Figure 7: Edge Parameters Over Time
The figure displays the proportion of hedge funds that satisfy the Edge methodology and its component parts (a and nonpositive b). The Edge parameters are estimated via rolling 3-year regressions. The grey line corresponds to the LHS vertical axis whereas the red and blue lines correspond to the RHS vertical axis. The sample is from May 2008 until December 2022 to account for the initial estimation window.

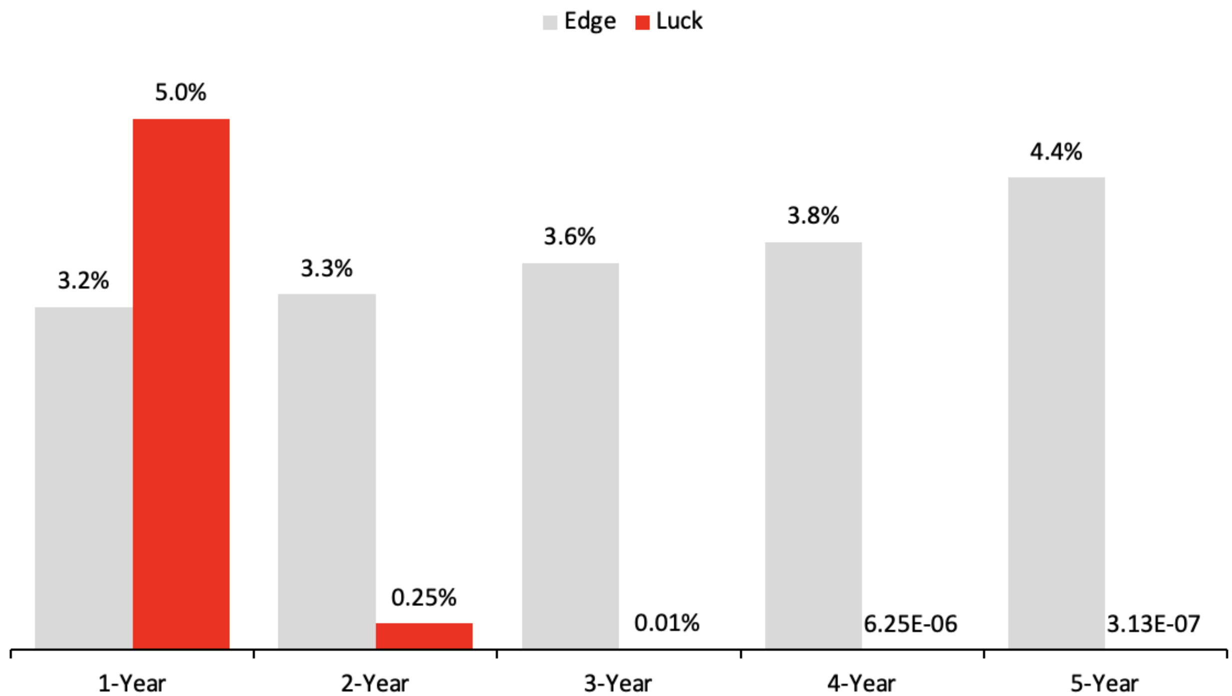
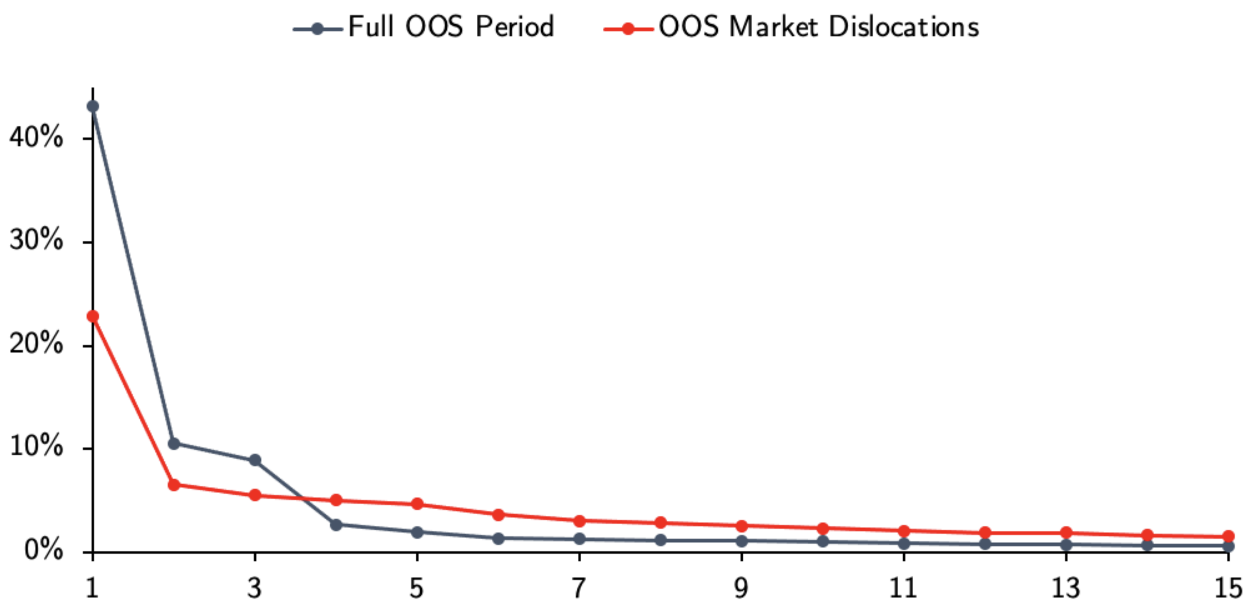
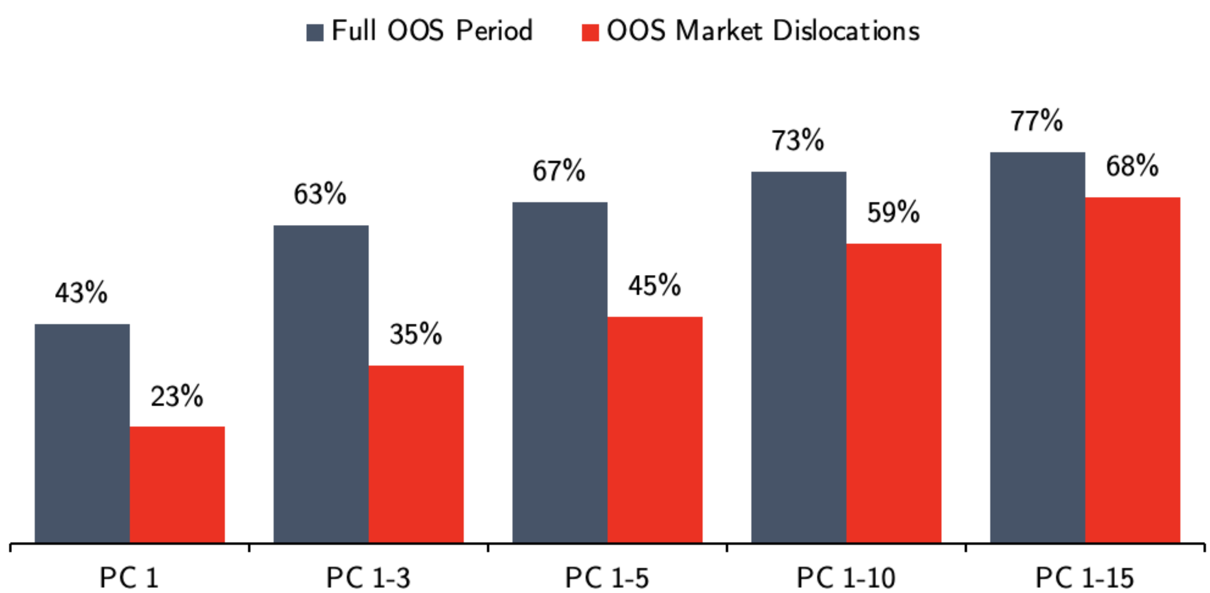


Figure 8: Edge Versus Luck
The figure displays the proportion of hedge funds, over longer horizons, that satisfy either the Edge methodology (grey bars) or identified as having Edge due to sampling error based on a 95% confidence interval (red bars). The red bars are based on the assumption that hedge fund returns are independent and identically distributed each year. The sample is from May 2008 until December 2022 to account for the initial estimation window.

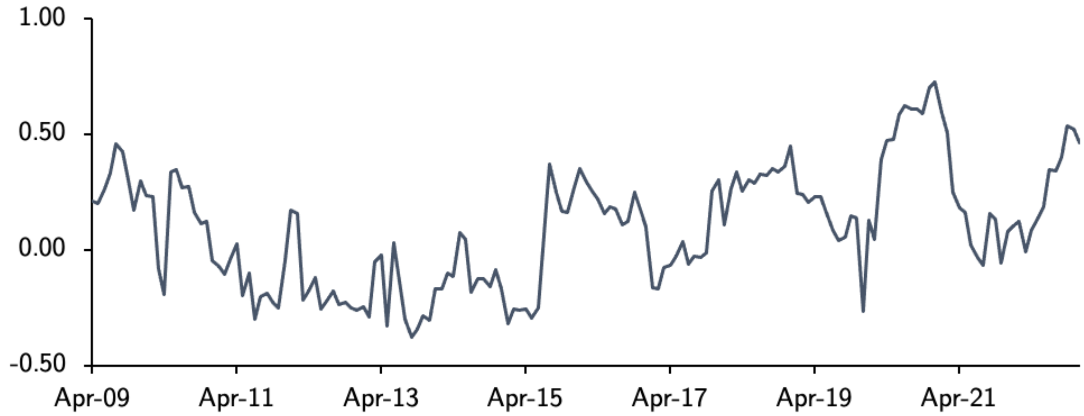


(a) Scree Plot

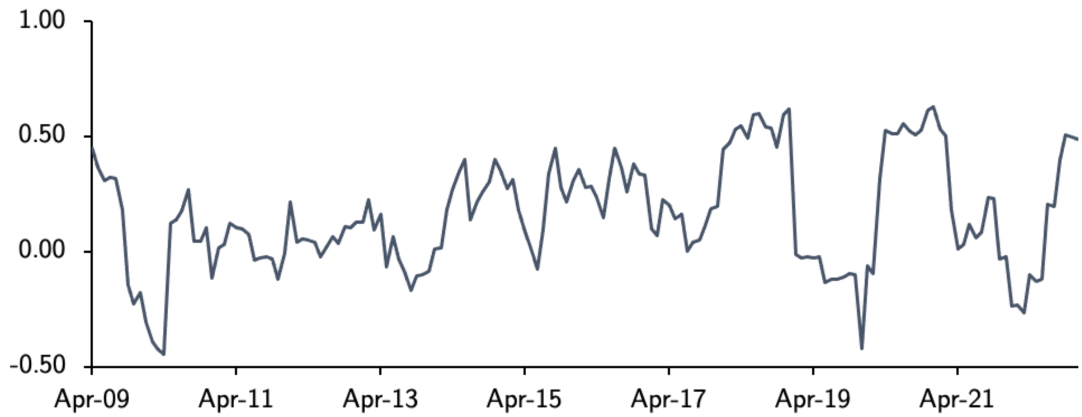


(b) Percentage of Variance Explained by PCs

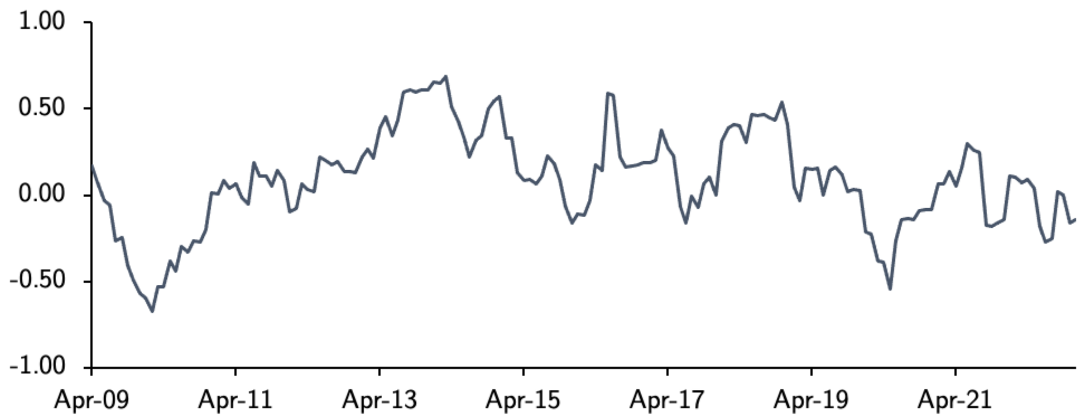
Figure 9: Variance Explained (%) by Leading PCs
The figure displays the proportion of variance explained by the leading fifteen PCs. Panel A displays the percentage of variance explained for each of the leading 15 PCs whereas Panel B displays the percentage of variance explained for different groupings of the leading 15 PCs. The eigenvalues are estimated from a variance-covariance matrix of the FH8 residuals over the OOS period between May 2007 and December 2022. The blue line corresponds to the full OOS period. The red line is based on an eigenvalue decomposition of the FH8 residuals that uses only the months in which the Short VIX benchmark return is negative during the OOS period.



(a) Short VIX

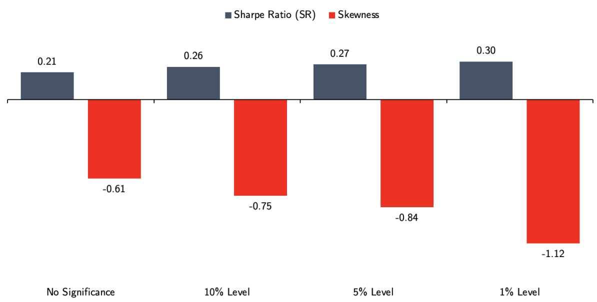


(b) Hedge Funds Without Edge

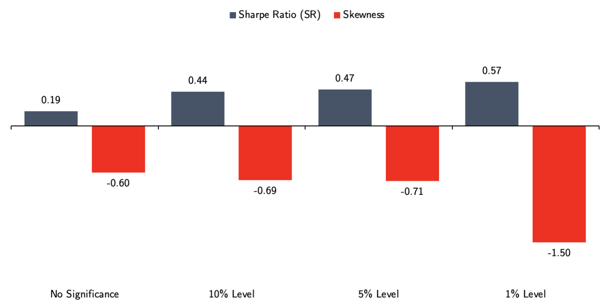


(c) Hedge Funds With Edge

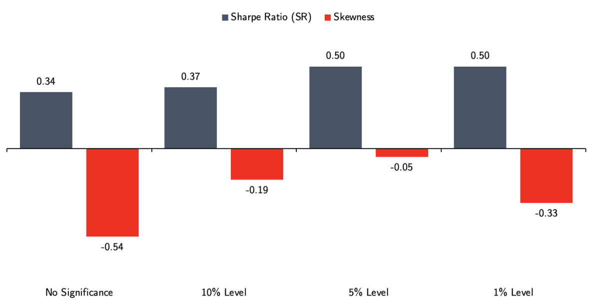
Figure 10: Rolling 12-Month Correlations With PC 10-15
Panel A, B and C display the rolling 12-month correlations between the sum of PCs 10-15 (from the FH8 residuals covariance matrix) and the following three return series, respectively: (i) Short VIX, (ii) Hedge Funds Without Edge Portfolio Returns and (iii) Hedge Funds With Edge Portfolio Returns. The hedge fund portfolio returns correspond to the main specification as shown in Figure 4.



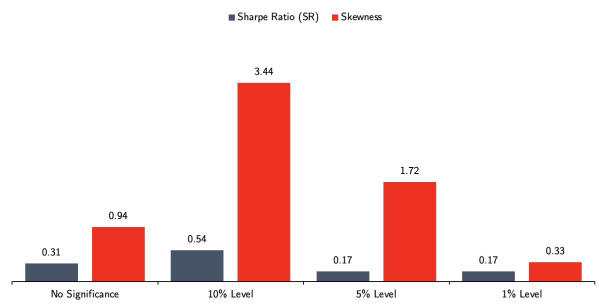
(a) FH8 $\hat{a} > 0$



(b) Short VIX $\hat{a} > 0$



(c) FH8 $\hat{a} > 0$ and $\hat{b}^{\text{mkt}} \leq 0$



(d) Short VIX $\hat{a} > 0$ and $\hat{b}^{\text{SVIX}} \leq 0$

Figure 11: SR and Skewness of OOS Hedge Fund Portfolios
The figure displays the SRs (blue bars) and Skewness (red bars) from the following equal-weighted hedge fund portfolio returns. Panels A and B form hedge fund portfolios by selecting hedge funds that have a strictly positive intercept. Panels C and D form hedge fund portfolios by selecting hedge funds that have both a strictly positive intercept and non-positive coefficient on the benchmark (or proxy for systematic risk). Panels A and C are based on the estimates of the FH8 Model with the addition of the three leading PCs from the residuals covariance matrix. The FH8 market coefficient loading, $\hat{b}^{\text{mkt}}$, is used as a proxy for the respective Edge parameter, $\hat{b}^{\text{SVIX}}$ since they both reflect systematic risk. Panels B and D are based on estimates of Equation 3 with the addition of the three leading PCs from the residuals covariance matrix. Each panel varies the level of significance of the intercept from No Signifiance (i.e. only the sign is required) to statistical significance at the 1% level. The Portfolio Returns correspond to the period from April 2011 to December 2022 to account for an additional 3-year rolling window to estimate the PCs from the initial OOS period.

Chapter 2

The Value of Economic Regularization for Stock Return Predictability

2.1 Introduction

One of the longest traditions in finance has been to propose some economic variable in order to predict future stock market returns. This effort has spawned a vast literature with many economically motivated predictors that have been shown to predict the equity premium in-sample (IS). In a path-breaking paper, Goyal and Welch (2008) document striking evidence that many of the most well-known predictors fail to predict the equity premium OOS, relative to an OOS prediction that only uses the historical average of the equity premium.[23] Many studies have responded to the evidence of Goyal and Welch (2008) (most notably Campbell and Thompson (2008)) by proposing economically motivated constraints to improve the OOS forecasting power of stock market return predictors. The common thread running along all of these studies is that the main OOS evaluation metric is based on OOS R^2, a measure of statistical fit. However, as Nagel (2021) notes, an improvement in OOS R^2 does not necessarily coincide with an improvement in the performance of an investor's portfolio.

To bridge the gap in evaluating the OOS performance of stock return predictors using either statistical fit (R^2) or an investor's portfolio performance (SR), we propose a new approach that places them both on an equal footing with one another. More specifically, we modify the standard least squares based loss function by adding a penalization term that maximizes the CE.[24] By modifying the loss function in this way, we are able to evaluate whether the stock market return predictors in question from Goyal and Welch (2008)

[23]Goyal et al. (2024) conduct a comprehensive review of more prominent return predictors that have been proposed since Goyal and Welch (2008) and find that their original finding is largely unchanged. That is, they find little to no evidence of equity premium predictability with this updated set of predictors.

[24]The CE is proportional to the SR for a mean-variance investor. Moreover, the CE can be viewed as a more general version of the SR used to measure portfolio performance for investors who care about higher-order moments.

translate into OOS economic gains for an investor.

The economic value offered by return predictors are often measured by the utility gain, or CE, of an investor using the predicted equity premium to allocate her assets between risky and risk-free assets. To demonstrate this point, Campbell and Thompson (2008) show that even a small amount of OOS statistical fit measured by R^2 can lead to a substantial improvement in the OOS CE gain. Moreover, Nagel (2021) emphasizes how an improvement in OOS R^2 can translate into *no* improvement in SR.[25] Therefore, it is natural to ask whether an investor should place less importance on statistical efficiency and instead seek to maximize utility gains. This is the main intuition underpinning the proposed methodology in this paper.

The traditional OLS approach predicts the next-period equity premium to be a linear function of the predictor variable (Equity Premium$_{t+1} = \alpha + \beta X_t + \epsilon_{t+1}$). We take this functional relationship as granted and propose the estimation of the associated coefficients to be carried out by adding an economically motivated penalization term to the standard least squares problem. The penalization term we use is the maximization of an investor's CE gain, instead of solely minimizing the sum of squared errors as in an OLS case.

Therefore, our approach is to be understood as a modification to the loss function in the linear predictive relationship. Note that there exists multiple different ways to improve OLS in the statistics literature when OLS is not feasible. These methods, which are often called regularization or broadly referred to as a machine learning algorithm, applies a tweak to the original loss function to make the optimization problem tractable. Our approach is in parallel to such statistical regularization approaches. In particular, we choose to augment the loss function by adding an economically motivated penalization term that reflects the investor's economic gains of a predictor. Therefore, we refer to our method as *economic regularization* in this paper.

[25] Most recently, Kelly et al. (2024) show in a more general setting, with respect to the case of model misspecification, how a positive OOS R^2 is not a necessary condition for an strategy that offers attractive risk-adjusted returns to an investor.

We empirically test our approach against OLS using the 14 return predictors studied in Goyal and Welch (2008). The results indicate substantial benefits for investors aligning the loss function with the evaluation metric. When tested for the OOS period up until 2023, a mean-variance investor with a risk aversion coefficient of 3 obtains an average of 2.38% per month of additional CE gain by using our proposed method over OLS. The economic gain becomes larger when an investor uses the CRRA utility function, delivering a 2.87% incremental CE gain. Overall, our results suggest that the benefit of using economic regularization is greater when an investor has preferences over higher moments and has a higher degree of risk aversion coefficient. These results directly stem from the setup of our approach that uniquely incorporates the specific choice of utility function and risk aversion coefficient into the estimation procedure.

The contribution of our paper is the introduction of a new method to evaluate return predictors from the investor's perspective, rather than finding a new predictor variable. Thus, our paper can be seen as an alternative way of re-visiting the time-series return predictability literature whereby the predictors are often judged to be good or bad based on their ability to improve statistical IS (or OOS) fit. As a consequence, many of the return predictors were not found to be useful based on the standard statistical criteria. Our proposed method suggests that it is worthwhile to revisit the proposed set of predictors and instead assess whether they can deliver superior economic gains when combined with the economically motivated IS estimation of coefficients.

Our economic regularization approach offers several advantages over the existing methodologies. First, as we are maximizing the IS utility gains directly, our approach delivers different estimates of the coefficients depending on the specific utility functions used, as well as a different coefficient of risk aversion. The previous portfolio allocation approach is a two-step procedure where an investor estimates the coefficients by running an OLS model, then uses the plug-in rule to estimate the optimal portfolio weights. The specific choice of the utility function and risk aversion coefficient only enters in the second step. In contrast, our approach combines these two-steps into a one-step procedure that estimates the coefficients while embedding the choice of utility function and risk aversion coefficient.

Second, our approach is also capable of incorporating the time-varying volatility of stock market returns. Again, the time-varying volatility is not considered until the second step whereby the weights are determined in the traditional approach, but our method unifies the time-varying nature of volatility into the estimation procedure. This can be seen as an alternative to the WLS-EV method proposed in Johnson (2019) that embeds the time-varying volatility into the OLS regression.

Our approach also differs from the stream of literature using more advanced statistical models or different utility functions to increase the power of using stock market return predictors. For instance, in an influential paper of DeMiguel, Garlappi, and Uppal (2009), they find that using minimum variance, 1/N, or Bayesian methods deliver superior performance over the conventional mean-variance approach in constructing optimal portfolio allocation. We differ from this branch of the literature in that we do not attempt to use different utility functions or statistical models, but rather propose to simply use an alternative parameter estimation method by using an economically motivated loss function. As a result, our approach is more flexible given that it can also be applied with different utility functions or using an alternative statistical model, which we demonstrate in a robustness check.

The closest paper to ours is perhaps Brandt, Santa-Clara, and Valkanov (2009). They propose a parametric portfolio choice rule to directly estimate the loading on the individual stock characteristics by maximizing a utility of the investor's portfolio return. Our approach builds upon their intuition to study a similar framework in the stock market return predictability context. Also, while their study focuses on the cross-section of stock returns we study the time-series property of using an economically motivated objective function.

The rest of the paper is organized as follows: we provide a detailed description of our proposed methodology in Section 2.2. Then, we present the empirical results in Section 2.3. Section 2.4 discusses various extensions and robustness checks of the result. We conclude in

Section 2.5.

2.2 Methodology

The traditional ordinary least squares (OLS) approach to forecasting stock market returns using an economic predictor is based on the following predictive regression specification

$$\text{Equity Premium}_{t+1} = \alpha + \beta X_t + \epsilon_{t+1}. \tag{7}$$

In the above specification, the coefficients α and β are estimated by regressing the next period's equity premium $\{R_{t+1} - R_t^f\}_{t=1}^{T-1}$, on the lagged predictor $\{X_t\}_{t=1}^{T-1}$, where R_{t+1} is the stock market return between time t and $t+1$ and R_t^f is the risk-free rate prevailing at time t for the same period. In estimating any econometric model, one has to specify the loss function to state the optimization problem to be solved. In the case of OLS, the corresponding loss function is the sum of squared residuals from the fitted relationship. In the form of an equation, estimating the coefficients via OLS can be written as the solution to the following optimization problem

$$\min_{\alpha,\beta} \sum_{t=1}^{T-1} (R_{t+1} - R_t^f - \alpha - \beta X_t)^2. \tag{8}$$

Although the OLS approach is the most statistically appealing way of studying the linear predictive relationships, the literature has documented strikingly weak empirical results when performance is measured using out-of-sample (OOS) utility gains for most of the predictors proposed. Goyal and Welch (2008) provide a comprehensive study to test if any of the economic predictors can beat the simple historical average of past equity premiums in predicting the equity premium out-of-sample. Their conclusion is largely pessimistic that they do not find statistically nor economically significant outperformance over the simple historical average for those predictors shown to deliver in-sample (IS) predictability. In fact, the majority of economically motivated equity premiums are shown to underperform the simple historical average OOS.

In order to overcome this issue, Campbell and Thompson (2008) propose an approach with an economically-motivated constraint to modify the standard OLS predictive regression. Campbell and Thompson (2008) suggest that by using the following two simple adjustments: 1) whenever the estimated sign of β is not consistent with the economic theory, then use the historical average of past equity premium as a next-period prediction instead, and 2) whenever the predicted next-period equity premium is negative, then set it equal to 0 instead. These two simple "economically intuitive" modifications to the OLS approach are then shown to greatly improve the OOS performance of economic predictors. Building on Campbell and Thompson (2008), Pettenuzzo, Timmermann, and Valkanov (2014) propose an alternative method by requiring that the conditional annualized Sharpe ratios for the market return to be bounded between 0 and 1. The proposed method of Pettenuzzo, Timmermann, and Valkanov (2014) is then shown to further improve the OOS predictive power over and above Campbell and Thompson (2008).[26] Hence, these studies showcase the efficacy of adding economic constraints in order to improve the performance of equity premium predictors out-of-sample.

Note that, in the aforementioned studies, the metric used to evaluate OOS performance is based on the regression's R^2, which measures the proportion of variation in the equity premium that is explained by the economic predictor(s). While R^2 may be the most appealing metric in a statistical sense, it does not necessarily translate into economic value placed by a stock market investor who is using the model. To address this issue, several studies shifted their focus to the economic value of using predictors for forecasting stock market returns.[27]

[26] Li and Tsiakas (2017) further improve the predictive power of economic predictors by adopting shrinkage estimation (via Elastic Net, LASSO and Ridge Regression) in combination with the economically motivated constraints suggested by Campbell and Thompson (2008).

[27] Cenesizoglu and Timmermann (2012) compare the economic value of multiple return prediction models. See Rapach and Zhou (2013) for comprehensive summary of literature. Neely et al. (2014) use technical indicators in forecasting equity risk premium while Rapach et al. (2016) document short interest as a strong predictor with OOS economic value. Huang and Zhou (2017) and Mei and Nogales (2018) use CRRA utility functions in their predictability models, among others.

Specifically, instead of using R^2 to evaluate each predictor, the certainty equivalent (CE) measure derived from utility functions was introduced to understand the economic gain of an investor using a proposed return predictor. One result of particular importance along this line was studied by Campbell and Thompson (2008) that showed even very modest OOS R^2 value can lead to substantial economic gains measured by CE.

Motivated by the observations above regarding the addition of economic constraints and economic evaluation criteria, we propose a new approach for using equity premium predictors, which we refer to as the *economic regularization* method. Our proposed approach estimates the coefficients α and β, similar to that in Equation 7, by modifying the least-squares based loss function with a penalization term that maximizes the in-sample certainty equivalent. In other words, we rely on an economically motivated loss function instead of a purely statistical loss function in order to maximize OOS economic gains for an investor.

Prior studies suggest the importance of aligning the loss function used for estimation and evaluation. For example, Engle (1993) points out the importance of choosing a loss function when defining a new model and Christoffersen and Jacobs (2004) show the importance of aligning the loss function with the evaluation metric for option pricing models. Motivation of our approach is based on the same intuition: if an investor is interested in potential economic gains OOS, then the IS estimation should use a loss function that also consider maximizing economic gains instead of statistical fit only.

In a recent article, Cederburg, Johnson, and O'Doherty (2023) demonstrate that statistically strong predictors can sometimes be economically unimportant if they tend to take extreme values in high-volatility periods, have low persistence, or follow distributions with fat tails. These findings lend further support as to the need of aligning loss functions for both the estimation and evaluation. Moreover, it suggests that statistically strong predictors IS does not necessarily translate to superior OOS economic performance.

As an example, consider an investor who has a mean-variance (MV) utility function with risk aversion coefficient γ. She faces a portfolio allocation problem where she has to find the optimal weights between investing in the stock market and the risk-free asset. At each time t, it can be easily shown that the optimal weight of her portfolio in the risky asset, the stock market, is given by

$$w_t = \frac{1}{\gamma} \frac{E_t[R_{t+1} - R_t^f]}{\sigma_{t+1|t}^2}, \tag{9}$$

where $E_t[R_{t+1} - R_t^f]$ is the conditional expectation of next period's equity premium and $\sigma_{t+1|t}^2$ is the conditional forecast of next period's stock return variance. Investors using the OLS approach (and economic predictor, X_t) will obtain estimates of the model coefficients $\hat{\alpha}$ and $\hat{\beta}$, then use plug-in rule to form expectations regarding the equity premium next period as $E_t[R_{t+1} - R_t^f] = \hat{\alpha} + \hat{\beta} X_t$. Then, the OOS CE metric corresponds to the utility gain of an investor who is using this expression to make a portfolio allocation OOS.

In contrast to the OLS approach, our proposed economic regularization method aims to estimate $\hat{\alpha}$ and $\hat{\beta}$ by modifying the least squares based loss function by adding a penalization term that maximizes the IS CE. Basic intuition comes from the machine learning literature that has proposed regularization techniques such as Ridge Regression (Hoerl and Kennard, 1970) or the class of least absolute shrinkage and selection operator (LASSO) introduced by Tibshirani (1996). In Ridge regression or LASSO estimation, one imposes an additional penalty term on the loss function to penalize when estimated coefficients β are too large, hence eliminating possibilities of unrealistic optimal coefficients with large magnitudes. In our setting, the LASSO estimator will solve the following minimization problem:

$$\min_{\alpha,\beta} \sum_{t=1}^{T-1} (R_{t+1} - R_t^f - \alpha - \beta X_t)^2 + \lambda |\beta|. \tag{10}$$

Motivated by the above, we propose to instead add an economically-motivated penalization term based on the CE to the standard least squares based loss function. Specifically, we solve the following optimization problem to estimate α and β:

$$\min_{\alpha,\beta} \sum_{t=1}^{T-1} (R_{t+1} - R_t^f - \alpha - \beta X_t)^2 \\ - \lambda \left[\frac{1}{T-1} \sum_{t=1}^{T-1} (w_t R_{t+1} + (1-w_t) R_t^f) - \frac{\gamma}{2} \frac{1}{T-2} \sum_{t=1}^{T-1} (w_t R_{t+1} + (1-w_t) R_t^f - \bar{R})^2 \right], \tag{11}$$

where the term in the square bracket corresponds to the in-sample certainty equivalent of an investor who allocates the asset using $w_t = \frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}$. In this case, the hyperparameter λ controls the extent to which how much penalization should be placed based on the CE criteria. Hence, higher values of λ force the estimators to favor coefficients that maximize the CE instead of minimizing the sum of least squares. It is straightforward to solve the above minimization problem and we outline the derivation in the Appendix A. Lastly, following Kozak et al. (2020a), we determine λ using K-fold cross validation methodology with K=3 for each equity premium predictor considered.

Our approach is an important departure from exclusively relying on OLS in order to lead to superior economic performance OOS. By remaining agnostic regarding the weights chosen with respect to either the sum of squared errors or the maximization of CE in the minimization problem, our method is able to both benefit from the penalization term while also mitigating the bias that is introduced. In doing so, our methodology leads to significant OOS CE gains.

The intuition of our approach is similar to Brandt et al. (2009), who propose parametric portfolio choice rule to construct optimal portfolios in the cross-section of stocks. In their approach, different characteristics of individual stocks are used to construct a parametric portfolio by solving a similar utility maximization problem IS. Our approach can be seen as a time-series extension (using equity premium predictors), of their cross-sectional application (using individual stock characteristics).

Lastly, our approach also carries the same spirit as the WLS-EV approach of Johnson

(2019). Johnson (2019) demonstrate that using the weighted least squares approach, taking the time-varying variance into consideration, significantly improves the OOS R^2 of the stock return predictive regression. We would like to emphasize that our approach also embeds the time-varying stock market variance implicitly through the functional form of the optimal weights. When constructing the IS objective function to be maximized, portfolio weights on the risky asset are determined by the ratio of the equity premium forecast and stock market variance. Hence, our objective function automatically takes the time-varying nature of the stock market variance into consideration by varying the weight placed on the risky asset at each period.

Having established the motivation for our new methodology it is now an empirical question of whether the proposed approach will perform better than OLS. In the next section, we take our approach to the data and study whether the proposed methodology can deliver superior CE gains OOS.

2.3 Empirical Results

2.3.1 Data

We use the same set of economic predictors studied in Goyal and Welch (2008).[28] We download the latest available data from Amit Goyal's website, ending in December 2023. From their original paper, we use 14 variables for the empirical study: divided price ratio (dp), dividend yield (dy), earnings price ratio (ep), dividend payout ratio (de), stock variance (svar), treasury bill rate (tbl), long term yield (lty), long term rate of return (ltr), term spread (tms), default yield spread (dfy), inflation (infl), log of book to market ratio (bm),

[28]The literature has more recently identified several economic variables that exhibit stronger return predictability then Goyal and Welch (2008)'s variables. These include the variance risk premium of Bollerslev et al. (2009), aggregate short interest of Rapach et al. (2016), and aggregate implied volatility spread of Han and Li (2020), among others. Since the focus of our paper is placed on the methodology rather than specific variables, we do not consider them in the main results. However, the proposed methodology is directly applicable to any set of predictors found in the literature. Accordingly, as a robustness check, we perform the main test of our paper using a more recently proposed large set of predictors from Cao et al. (2023).

cross-sectional premium (csp), and net equity expansion (ntis). Table 8 provides descriptive statistics of the variables during our sample period.

The equity premium is defined in the standard way as the monthly excess return of the CRSP value-weighted index (i.e. the difference between the realized return, including dividends, of the value-weighted CRSP companies and the prevailing risk-free rate). All returns are transformed to log-return before taking the difference. Since the focus of our paper lies on the usefulness of return predictors for the equity premium, we simply assume the next-period forecast of market variance, $\sigma^2_{t+1|t}$, is given by the previous month's stock variance, σ^2_t. This assumption is made throughout the paper when the conditional variance is needed to compute the optimal weights of the portfolio.

2.3.2 Main Results

To measure the additional economic benefits our method can bring, we proceed as follows. Consistent with the literature, we use the first 240 months (20 years) of observations as the initial estimation window. At each time t, if there are more than 240 observations available, we estimate the coefficients α and β for each of the 14 predictors using Equation (11). The estimation is done on a rolling basis with an expanding window. We assume a conservative estimate for the coefficient of risk aversion γ equal to 3.[29]

With the parameter estimates, we compute the OOS CE gain of a MV investor using the same expression as in Equation (11) except using the OOS observations. We denote this OOS CE using our method as CE(OLSCE). In parallel, we also compute two other certainty equivalent gains. First, we denote CE(OLS) for the CE gain of an investor using the traditional OLS method to allocate her assets. Second, we denote CE(Hist) for the CE gain of an investor who exclusively relies on the simple historical average of past equity premiums to allocate her assets, without using economic predictors.

Naturally two questions arise that need to be examined empirically. First, do CE(OLSCE)

[29]In Section 2.4, we demonstrate that increasing the coefficient of risk aversion actually delivers more favourable results with respect to our approach.

and CE(OLS) outperform CE(Hist)? That is, does the use of economic predictors deliver superior OOS economic gains for a mean-variance investor? Second, does our proposed methodology outperform the traditional OLS method? To see this clearly, we introduce the following notation:

$$\Delta CE(OLS) = CE(OLS) - CE(\text{Hist}) \tag{12}$$
$$\Delta CE(OLSCE) = CE(OLSCE) - CE(\text{Hist}).$$

Table 9 reports the above metrics for each of the 14 economic predictors, as well as their differences defined by Diff = $\Delta CE(OLSCE) - \Delta CE(OLS)$. First, the average OOS extra CE gain of the OLS method over the historical average is negative, being -0.05% per month. This means that investors using the historical average would have been better off relative to using OLS. This is particularly interesting since we reach the same conclusion as Goyal and Welch (2008) who evaluate these predictors based on statistical fit measured by OOS R^2 in contrast to our method of evaluation in terms of economic gain. Out of 14 predictors, five of them exhibit negative extra CE gains with the remaining nine displaying slightly positive values. The range of extra CE gain is modest, from the lowest of -0.85% using csp (cross-sectional premium) to the highest of 0.64% using ntis (net equity expansion). In contrast, the OOS incremental CE gain of our method over the historical average is significantly positive (average of 2.34% per month) thereby delivering superior economic performance relative to OLS and the historical average. The spread in CE gains between the best and worst performing predictor is higher using our method versus OLS. The worst performing predictor is now csp (cross-sectional premium) where it has 0.95% of $\Delta CE(CE)$ value. The best performing predictor, on the other hand, is ltr (long-term rate of return) that delivers 4.08% of extra CE gain. Out of 14 predictors, none of them show inferior performance relative to the historical average using our proposed method.

When comparing our method to OLS, the average difference is 2.38% per month where none of the 14 predictors show inferior performance to OLS. The relative gains in CE gains using our method relative to OLS is positive and similar in magnitude across all 14 predictors. Overall, the robust results of Table 9 suggests the potential advantage for a mean-variance investor by using our method over OLS. It is worthwhile to note that our method tends to deliver more stable CE gains compared to OLS, which indicates that a mean-variance investor would be better off using the full set of predictors. These results suggests that our proposed approach is able to more reliably identify predictors that offer a MV investor OOS CE gains.

To better understand the divergence between the two approaches, we report the time-series average of the estimated coefficients α and β in Table 10. Our main interest is in the magnitude of the estimated loading on the predictor variable, $\bar{\beta}$. Clearly, we see a larger average, in absolute value, for most of the estimated β coefficients using our method, labelled $\bar{\beta}$ OLSCE, compared to OLS, labeled $\bar{\beta}$ OLS. This should not be too surprising given that we introduce a bias with our method. The differences are sometimes quite extreme. For example, when using dfy (default yield spread) as the predictor, the average OLS loading is 0.086 while the proposed method changes sign and is meaningfully negative with an average loading of -0.665. This evidence highlights how the introduction of the CE-based penalty term has a significant impact on the estimated parameters relative to OLS.

Figures 12 and 13 plot the rolling estimates of α and β, respectively, for each of the 14 predictor variables using OLS (solid blue line) compared to our OLSCE approach (dashed orange line). In addition, NBER-defined recessions are represented by the grey shaded regions within each subfigure. Noticeable differences between the two approaches arise at the beginning of the sample. The OLSCE α estimates are significantly higher compared to OLS and remain higher throughout the entire sample period. However, the OLSCE β estimates are either higher or lower (in comparison to OLS) depending on the predictor variable. This is consistent with estimator's bias in Equation 26 derived in Section 3.5. The sign of the bias in Equation 26 is a consequence of the covariance between the predictor and the equity premium.

Both α and β coefficients estimated using the OLSCE approach show a greater degree of variation across time while the OLS approach produces a more stable pattern. This suggest that the OLSCE approach adjusts to the fluctuations in the business cycle faster than OLS. In contrast to OLS which places an equal weight across all observations, the OLSCE approach is able to directly apply a weighting for each period through the embedded stock market variance component. In doing so, the OLSCE method is able to capture the time-varying nature of the predictive relationship more effectively. It is important to note that given that the OLSCE approach is capable of producing a negative coefficient β estimate for an economic predictor (contrary to economic theory) such as the early part of the sample period for the dividend price (dp) ratio, it suggests that there is further room for improvement by applying the economic constraint of Campbell and Thompson (2008), which we later verify in Section 2.4.

Overall, our first set of empirical results point towards the benefit of aligning the IS loss function and OOS evaluation metric. In the next Section, we provide further extensions and robustness tests to validate our methodology.

2.4 Robustness

In this section, we conduct various robustness tests by using a different utility function, applying the economic constraint of Campbell and Thompson (2008), using various coefficients of risk aversion, sub-sample analysis, longer horizon forecasting and testing our methodology with another set of stock return predictors as in Cao et al. (2023). All robustness tests support our main results, with even stronger empirical findings. All together, this evidence indicates the efficacy of economic regularization in maximizing OOS CE gains for an investor.

2.4.1 CRRA Utility

Although mean-variance utility is one of the most popular choices for an optimal asset allocation problem, due to its analytical tractability, it suffers a drawback of not being able to capture risk appetite with respect to higher moments. In general, investors are not only concerned with the mean and variance of portfolio returns, but rather are concerned with its entire distribution. One parsimonious way to address this issue is to use the CRRA utility function defined by the following period-by-period utility

$$U(r_t) = \frac{(1+r_t)^{1-\gamma}}{1-\gamma}. \tag{13}$$

Under the CRRA utility framework, our optimization problem is modified with a penalization term defined in Equation (11) now stated as

$$\max_{\alpha,\beta} \sum_{t=1}^{T-1} U(w_t R_{t+1} + (1-w_t)R_t^f), \tag{14}$$

where w_t has the same functional form as in Equation (9).[30] Following Faias and Santa-Clara (2017), we define the OOS certainty equivalent by the following equation

$$CE = [(1-\gamma)\bar{U}]^{1/(1-\gamma)} - 1, \tag{15}$$

where $\bar{U}$ is the average CRRA utility of the OOS portfolio returns. We use the same risk aversion coefficient $\gamma = 3$ as in the mean-variance case.

Table 11 reports the results in a similar format to Table 9. Surprisingly, the difference between our approach and the traditional OLS estimator is even more pronounced in the CRRA case. The OLS model performs miserably by having an average extra CE gain of −2.30% (or loss of 2.30%) over the simple historical average model. However, it is impor-

[30] Campbell and Viceira (2002) derive the same functional form under the log-normally distributed returns and linear approximations, which we follow.

tant to note that the large extra CE loss is accentuated by by extreme observations. For instance, the ntis (net equity expansion) predictor has a -24.53% of $\Delta CE(CE)$ relative to the historical average. Five of the 14 predictors show an improvement over the historical average. Nevertheless, the performance of our method is quite promising in comparison. Not only is the average extra CE gain positive being 0.57%, but all of the 14 predictors (except ntis) show positive and significant improvements over the historical average. Moreover, the improvement is quite uniform across all predictors, rather than disperse like the mean-variance case. As a result, a CRRA investor using our method benefits by an average of 2.87% across 14 predictors over an investor using OLS model. Our results highlight the possibility that the true value of using the economic regularization method may lie in the power of optimizing over the higher moments of portfolio returns. While the mean-variance utility case still delivers significant differences between the two methods, the CRRA case thus further strengthens the importance of aligning the loss function with the evaluation metric. Moreover, it highlights the pitfalls of relying on a loss function that is based exclusively on statistical efficiency.

Figure 14 plots the time-series of period-by-period CRRA utility of OOS optimal portfolio returns. It is clear from the figure why investors would prefer the OLSCE approach over OLS. A CRRA investor who relies exclusively on OLS experiences an extremely volatile time-series of utility gains while the OLSCE approach delivers more stable utilities throughout the OOS period. Although OLS sometimes delivers higher utilities, its losses relative to OLSCE during bad times are significantly larger. As a result, when CRRA utility is used, investors show a greater willingness to use our economic regularization in comparison to traditional OLS. Lastly, OLSCE delivers higher CE gains over the simple historical average approach.

2.4.2 Campbell and Thompson Adjustment

Since our approach is based on maximizing economic gain, it is natural to also consider if our approach is merely capturing the extant methods that add an economic constraint.

To verify this, we apply the economic constraint of Campbell and Thompson (2008) (CT) to see whether it provides further improvements. Table 12 reports the results for a mean-variance investor with the risk aversion coefficient of $\gamma = 3$. Comparing the result with Table 9, without the CT adjustment, we see largely similar results. The average extra CE gain for the OLS method is unchanged at -0.05%, with the cross-sectional dispersion across predictors slightly reduced. For the case of our method, the average extra CE gain is largely similar compared to before (2.29% vs. 2.34%). As before, all 14 predictors show positive extra CE gain over the historical average while six out of 14 predictors suffer from negative extra CE gain in the OLS model.

The similarity between the findings in Table 9 and this subsection using the CT adjustment indicate that our methodology is able to accomplish the stability offered by previous approaches which impose additional economic constraints. The results indicate that our blended approach that modifies the least squares loss function with an economically motivated penalization term does not require additional economic constraints to improve OOS economic performance.

2.4.3 Different Risk Aversion Coefficient

To ensure our results are not driven by the specific choice of risk aversion parameter $\gamma = 3$, we consider higher orders of risk aversion ($\gamma = 5$ and $\gamma = 10$). In principle, our method should work even better with higher orders of risk aversion because our method directly incorporates the parameter γ in the estimation process. In contrast, OLS is based on a purely statistical framework, leaving no room for any specific risk aversion coefficient to play a role. Thus, increasing γ, or making the impact of investor's risk appetite more important, should deliver larger differences between our method and OLS.

Table 13 shows the results which confirms that the expected improvement is in fact borne out. Our method outperforms OLS by a large magnitude. We report the extra CE gain our method delivers over the OLS model for two cases of $\gamma = 5$ and $\gamma = 10$. For $\gamma = 5$, the average extra CE gain of our method over OLS model is 4.95% where all 14 predictors show positive gain. When γ is set equal to 10, the extra CE gain is magnified to average of 11.37% with all 14 predictors outperforming OLS. Hence, we conclude that our proposed method is going to be even more useful for more risk-averse investors.

2.4.4 Economic Expansions and Contractions

Next, we test whether our proposed method's advantage varies based on different parts of the business cycle. We split the sample into sub-periods of NBER-defined economic expansions versus contractions. Table 14 compares the mean values of the 14 economic predictors from 1946 to 2023 during expansion periods. According to the National Bureau of Economic Research, an expansion period is deemed a "normal state of the economy"[31]. Our model (OLSCE) exhibits a positive difference of 2.13% compared to OLS which has a negative average difference of -0.07%. The average difference in CE gains relative to the historical average between our method and OLS is 2.21% during expansions. Half of the predictors are significant at conventional levels of significance. The largest spread in mean differences is shown in ltr (long-term rate of return) of 3.12% which is statistically significant at the 10% level. During expansion periods, the CE measure allows investors to understand economic gain through return predictors. In sum, we find that economic predictors deliver superior gains for a mean-variance investor during expansion periods.

Table 15 compares the mean values of the 14 economic predictors from 1946 to 2023 during contraction periods, whereby a contraction period occurs between the peak and trough of a business cycle. The mean difference between models during the contraction period is 3.85% (3.95% vs. 0.11%). However, only one of 14 predictors (ltr) has a difference that is statistically significant. Consistent with expansions, the largest spread in mean differences is shown in ltr (long-term rate of return). However, this spread widens significantly to 9.84% and its statistical significance strengthens to the 5% level. We see all of the predictors

[31] https://www.nber.org/research/business-cycle-dating

(except for cross-sectional premium, csp) show positive extra CE using our new methodology during contractions. The evidence in this table is consistent with the earlier evidence that highlights how our approach offers larger benefits with utility functions that incorporate higher moments and larger degrees of risk aversion. That is, the marginal utility of risk-averse investors is higher during an economic contraction relative to an expansion. Hence, an approach that incorporates utility gains in its estimation should deliver larger OOS utility gains to an investor during periods when they value it the most (i.e. economic contractions). All together, the findings of this sub-period analysis further supports the utility of our model relative to OLS regardless of economic regime.

2.4.5 Longer-Horizon Forecasting

The baseline specification uses the one-month ahead equity premium to generate all of the empirical results discussed thus far. To demonstrate the predictive power of our apporach is not primarily a short-run phenomenon, we re-estimate the main results (in Section 2.3.2 with mean-variance preferences and γ equal to 3) at longer forecast horizons: 3, 6 and 12 months ahead.

Table 16 is comprised of three panels (A, B and C) that correspond to the three metrics: Δ CE(OLS), Δ CE(OLSCE) and Diff, respectively. Each panel displays the respective metric for each of the 14 predictors across the different forecast horizons next to the baseline specification (1 month ahead equity premium). Most importantly, the empirical results are largely unchanged. More specifically, OLS continues to underperform the historical average at longer horizons (an OOS CE gain of -0.10%, on average) and our proposed OLSCE approach is able to reliably produce OOS CE gains at longer horizons relative to the historical average (1.10% at 12 months, on average). It is important to note that the magnitude of the average OOS CE gain for OLSCE does fall significantly beyond the 1 month horizon CE gain of 2.34%. OLSCE outperforms OLS across 1, 3, 6 and 12 month horizons.

2.4.6 An Alternative Set of Equity Premium Predictors

The last robustness test of our methodology is applying our new approach to predictors that have been proposed subsequent to Goyal and Welch (2008). To that end we test our methodology using 17 alternative predictors that have been proposed in Cao et al. (2023) over the sample period 2005 to 2020: aggregate call order imbalance (acib), aggregate equity put option order imbalance (apib), aggregate implied volatility spread (ivs), cboe vix (vix), variance risk premium (vrp), index call option order imbalance (icib), index put option order imbalance (ipib), bw sentiment (bw sentiment), pls sentiment (pls sentiment), hlw disagreement (hlw disp), first principle component (PC) of the 22 predictors based on Goyal and Welch (2008) (gw pc 1), second PC of the 22 predictors based on Goyal and Welch (2008) (gw pc 2), third PC of the 22 predictors based on Goyal and Welch (2008) (gw pc 3), consumer sentiment index from the University of Michigan (michigan sentiment), manager sentiment index (manager sentiment), and aggregate purchase of DOTM SPX index put options (pnbo).

Tables 17, 18 and 19 presents the summary statistics, OOS CE gains, average estimated coefficients, respectively, for this alternative set of predictors. As previously, the CE gains are based on the main specification for a mean-variance investor with risk aversion coefficient equal to 3. The main takeaway from Table 18 is that our results are magnified using this alternative set of recently proposed predictors. The OOS CE gains are larger using our methodology compared to the Goyal and Welch (2008) predictor set (3.37% compared to 2.34%). The average difference between our new approach and OLS widens with this new set of predictors compared to the previous 14 predictor set (4.80% vs. 2.38%). All together, this evidence highlights the flexibility and robustness of our new approach in its ability to be tested with new economic variables that are proposed to predict the equity premium.

2.5 Conclusion

This paper proposes a new method to use stock market return predictors, that can enhance OOS utility gains of an investor. Instead of relying exclusively on the standard estimation via OLS, our proposed method modifies the least squares based loss function, via economic regularization, by imposing an economically-motivated penalty that maximizes IS CE of an investor. The empirical findings support the new methodology and shows significant OOS improvements in CE gains using the proposed method over OLS or the simple historical average. The empirical evidence is even stronger when using CRRA utility, higher orders of risk aversion or more recently proposed predictors of the equity premium.

Our findings contribute to the literature on return predictability by proposing a new way of leveraging the predictive content within economic variables regarding the equity premium. The approach is sufficiently general that it can be extended to different classes of utility functions. Our results point towards the benefit of future research exploring the development of other economically motivated loss functions in a broader context.

Table 8: Descriptive Statistics of 14 Predictor Variables

	Period	Mean	Median	Std. Dev.	Skewness
dp	1946–2023	−3.252	−3.165	0.459	−0.608
dy	1946–2023	−3.248	−3.162	0.456	−0.632
ep	1946–2023	−2.693	−2.704	0.384	−0.572
de	1946–2023	−0.558	−0.561	0.317	0.819
svar	1946–2023	0.003	0.001	0.005	7.107
tbl	1946–2023	0.033	0.030	0.030	1.113
lty	1946–2023	0.049	0.042	0.027	1.183
ltr	1946–2023	0.005	0.003	0.025	0.582
tms	1946–2023	0.016	0.016	0.013	−0.120
dfy	1946–2023	0.012	0.010	0.007	2.188
infl	1946–2023	0.003	0.002	0.006	0.802
bm	1946–2023	−0.725	−0.643	0.514	−0.388
csp	1957–2002	0.000	0.000	0.002	0.543
ntis	1947–2023	0.016	0.016	0.026	1.610

This table provides descriptive statistics of the variables during our sample period 1946–2023. The 14 predictor variables as per the empirical study by Goyal and Welch (2008), are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**).

Table 9: Comparison of OOS Certainty Equivalent Gains

	Period	Δ CE(OLS)	Δ CE(OLSCE)	Diff
dp	1946–2023	−0.45%	1.72%	2.17%
dy	1946–2023	−0.62%	1.82%	2.45%
ep	1946–2023	−0.17%	1.79%	1.96%
de	1946–2023	0.46%	1.80%	1.33%
svar	1946–2023	0.06%	2.12%	2.07%
tbl	1946–2023	0.15%	3.00%	2.86%
lty	1946–2023	0.01%	2.99%	2.98%
ltr	1946–2023	0.11%	4.08%	3.97%
tms	1946–2023	0.30%	3.20%	2.90%
dfy	1946–2023	0.00%	1.95%	1.95%
infl	1946–2023	0.11%	3.07%	2.95%
bm	1946–2023	−0.42%	1.69%	2.11%
csp	1957–2002	−0.85%	0.95%	1.80%
ntis	1947–2023	0.64%	2.54%	1.89%
Average:		−0.05%	2.34%	2.38%

This table presents statistics on forecast errors by comparing the OOS CE (OLSCE) gains for an investor who has mean-variance (MV) utility function with risk aversion coefficient $\gamma = 3$. The OOS certainty equivalent gain of an MV investor is computed using the same expression as in Equation (11) below, but using the OOS observations instead:

$$\min_{\alpha,\beta} \sum_{t=1}^{T-1} (R_{t+1} - R_t^f - \alpha - \beta X_t)^2 - \lambda \left[\frac{1}{T-1} \sum_{t=1}^{T-1} (w_t R_{t+1} + (1-w_t) R_t^f) - \frac{\gamma}{2} \frac{1}{T-2} \sum_{t=1}^{T-1} (w_t R_{t+1} + (1-w_t) R_t^f - \bar{R})^2 \right]. \tag{16}$$

CE(OLSCE) is the OOS certainty equivalent using our method, compute two other certainty equivalent gains. CE(OLS) denotes the CE gain of an investor using OLS method, and CE(Hist) for the CE gain of an investor using the simple historical average of past equity premium to allocate her assets, without using economic predictors. The deltas are defined as $\Delta CE(OLS) = CE(OLS) - CE(\text{Hist})$ and $\Delta CE(OLSCE) = CE(OLSCE) - CE(\text{Hist})$.

The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**). The **Diff** column shows noticeable differences between the two models across these variables' CE gains.

Table 10: Average Estimated Coefficients for OLS and OLSCE Approach

	$\bar{\alpha}$ OLS	$\bar{\alpha}$ OLSCE	$\bar{\beta}$ OLS	$\bar{\beta}$ OLSCE
dp	0.039	0.061	0.011	0.014
dy	0.052	0.080	0.015	0.020
ep	0.052	0.079	0.018	0.024
de	−0.001	0.005	−0.014	−0.023
svar	0.006	0.023	−0.099	−0.724
tbl	0.008	0.030	−0.106	−0.168
lty	0.015	0.030	−0.290	−0.172
ltr	0.005	0.022	−0.029	0.483
tms	0.003	0.021	0.131	0.420
dfy	0.004	0.027	0.086	−0.665
infl	0.006	0.028	−0.346	−2.107
bm	0.012	0.021	0.014	0.011
csp	0.003	0.012	2.354	3.327
ntis	0.010	0.024	−0.212	−0.228

This table reports the time-series average of the coefficients α and β is estimated for an investor who has mean-variance (MV) utility function with risk aversion coefficient $\gamma = 3$. It is notable that the magnitude of the average estimated loadings on the predictor variable, $\bar{\beta}$ is much larger when using our CE method, compared to the OLS method. The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**).

Table 11: Comparison of OOS Certainty Equivalent Gains: CRRA Investor

	Period	Δ CE(OLS)	Δ CE(OLSCE)	Diff
dp	1946–2023	0.54%	0.59%	0.05%
dy	1946–2023	0.28%	0.57%	0.29%
ep	1946–2023	−2.80%	0.57%	3.37%
de	1946–2023	−4.44%	0.48%	4.92%
svar	1946–2023	−0.16%	0.53%	0.69%
tbl	1946–2023	0.49%	0.64%	0.15%
lty	1946–2023	0.35%	0.65%	0.30%
ltr	1946–2023	−1.96%	0.86%	2.82%
tms	1946–2023	0.14%	0.63%	0.49%
dfy	1946–2023	−0.05%	0.47%	0.52%
infl	1946–2023	−0.10%	0.65%	0.55%
bm	1946–2023	−0.06%	0.51%	0.58%
csp	1957–2002	−0.12%	1.14%	1.26%
ntis	1947–2023	−24.53%	−0.26%	24.26%
Average:		−2.30%	0.57%	2.87%

This table presents statistics on forecast errors by comparing the OOS CE (OLSCE) gains for an investor who has CRRA utility function with risk aversion coefficient $\gamma = 3$, with the OOS OLS performance. The CRRA utility function is defined as $U(r_t) = \frac{(1+r_t)^{1-\gamma}}{1-\gamma}$, and the optimization problem is to maximize $\sum_{t=1}^{T-1} U(w_t R_{t+1} + (1-w_t)R_t^f)$. The OOS CE is $CE = [(1-\gamma)\bar{U}]^{1/(1-\gamma)} - 1$ where $\bar{U}$ is the average CRRA utility of the OOS portfolio returns. The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**).

Table 12: Comparison of OOS Certainty Equivalent Gains with CT Adjustment

	Period	Δ CE(OLS)	Δ CE(OLSCE)	Diff
dp	1946–2023	−0.43%	1.72%	2.16%
dy	1946–2023	−0.59%	1.82%	2.41%
ep	1946–2023	−0.16%	1.80%	1.96%
de	1946–2023	0.47%	1.80%	1.33%
svar	1946–2023	0.05%	2.12%	2.07%
tbl	1946–2023	0.06%	2.82%	2.76%
lty	1946–2023	−0.19%	2.86%	3.05%
ltr	1946–2023	0.13%	3.90%	3.77%
tms	1946–2023	0.29%	3.18%	2.90%
dfy	1946–2023	0.00%	1.95%	1.95%
infl	1946–2023	0.11%	2.92%	2.81%
bm	1946–2023	−0.31%	1.69%	2.00%
csp	1957–2002	−0.74%	0.88%	1.61%
ntis	1947–2023	0.65%	2.54%	1.89%
Average:		−0.05%	2.29%	2.33%

This table reports the results of comparing the OOS CE (OLSCE) gains after applying the Campbell and Thompson (2008) (CT) economic constraint. The presented forecast errors are for a mean-variance (MV) investor with the risk aversion coefficient of $\gamma = 3$. The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**).

Table 13: Comparison of OOS Certainty Equivalent Gains with Different γ Values

	Period	CE(OLSCE)-CE(OLS)	CE(OLSCE)-CE(OLS)
		$\gamma = 5$	$\gamma = 10$
dp	1946–2023	4.95%	0.91%
dy	1946–2023	1.72%	10.57%
ep	1946–2023	1.66%	12.19%
de	1946–2023	4.47%	10.67%
svar	1946–2023	4.75%	10.71%
tbl	1946–2023	6.25%	13.80%
lty	1946–2023	6.21%	13.54%
ltr	1946–2023	8.52%	18.39%
tms	1946–2023	6.39%	14.40%
dfy	1946–2023	4.47%	9.93%
infl	1946–2023	6.32%	13.93%
bm	1946–2023	5.20%	11.25%
csp	1957–2002	3.45%	7.31%
ntis	1947–2023	4.93%	11.62%
Average:		4.95%	11.37%

This table tests our model's robustness by comparing the OOS CE (OLSCE) gains for different choices of risk aversion values. Two cases with $\gamma = 5$ and $\gamma = 10$ are presented, showing additional CE advantage over OLS, for investors with higher risk-aversion. The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**).

Table 14: Mean Comparison: Economic Expansion Period

	Period	OLSCE	OLS	Diff (OLSCE-OLS)	
dp	1946–2023	1.55%	−0.55%	2.09%	***
dy	1946–2023	1.59%	−0.78%	2.38%	***
ep	1946–2023	1.72%	−0.23%	1.95%	
de	1946–2023	1.66%	0.47%	1.19%	
svar	1946–2023	1.94%	0.06%	1.87%	
tbl	1946–2023	2.63%	0.05%	2.58%	*
lty	1946–2023	2.78%	−0.03%	2.80%	**
ltr	1946–2023	3.28%	0.16%	3.12%	*
tms	1946–2023	2.93%	0.25%	2.67%	
dfy	1946–2023	1.70%	−0.02%	1.72%	
infl	1946–2023	2.74%	0.09%	2.65%	
bm	1946–2023	1.55%	−0.56%	2.11%	**
csp	1957–2002	1.27%	−0.69%	1.96%	**
ntis	1946–2023	2.55%	0.75%	1.80%	
Average:		2.13%	−0.07%	2.21%	

This table tests the mean differences by comparing the two empirical strategies during the expansion period from 1946 - 2023. The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**). Statistical significance of mean difference is indicated at 1%, 5%, and 10% levels.

Table 15: Mean Comparison: Economic Contraction Period

	Period	OLSCE	OLS	Diff (OLSCE-OLS)	
dp	1946–2023	3.12%	0.15%	2.96%	
dy	1946–2023	3.56%	0.31%	3.25%	
ep	1946–2023	2.60%	0.20%	2.40%	
de	1946–2023	3.06%	0.40%	2.66%	
svar	1946–2023	3.82%	0.00%	3.82%	
tbl	1946–2023	5.46%	0.70%	4.76%	
lty	1946–2023	4.35%	0.22%	4.13%	
ltr	1946–2023	9.63%	−0.20%	9.84%	**
tms	1946–2023	5.49%	0.56%	4.92%	
dfy	1946–2023	3.98%	0.07%	3.91%	
infl	1946–2023	5.56%	0.26%	5.31%	
bm	1946–2023	2.91%	0.41%	2.50%	
csp	1957–2002	−1.15%	−1.59%	0.44%	
ntis	1946–2023	2.98%	0.03%	2.95%	
Average:		3.95%	0.11%	3.85%	

This table tests the mean differences by comparing the two empirical strategies during the contraction period from 1946 - 2023. The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**). Statistical significance of mean difference is indicated at 1%, 5%, and 10% levels.

Table 16: Comparison of OOS Certainty Equivalent Gains (%) over Longer Horizons

Panel A: Δ CE(OLS)

	dp	dy	ep	de	svar	tbl	lty	ltr	tms	dfy	infl	bm	csp	ntis	Avg.
1 mo	−0.45	−0.62	−0.17	0.46	0.06	0.15	0.01	0.11	0.30	0.00	0.11	−0.42	−0.85	0.64	−0.05
3 mo	−0.16	−0.08	0.26	0.42	0.10	0.30	0.29	0.12	0.39	0.03	0.07	−0.29	−0.17	0.48	0.13
6 mo	−0.07	−0.03	0.59	0.36	0.16	0.23	0.38	0.26	0.20	0.18	0.05	0.06	−0.16	0.82	0.21
12 mo	−0.79	−0.78	0.12	0.14	0.19	−0.12	0.12	0.07	0.14	−0.22	0.01	−0.58	0.09	0.24	−0.10

Panel B: Δ CE(OLSCE)

	dp	dy	ep	de	svar	tbl	lty	ltr	tms	dfy	infl	bm	csp	ntis	Avg.
1 mo	1.72	1.82	1.79	1.80	2.12	3.00	2.99	4.08	3.20	1.95	3.07	1.69	0.95	2.54	2.34
3 mo	−0.66	1.28	1.60	1.28	0.92	1.58	1.55	1.44	2.19	0.95	1.84	0.33	−0.39	1.05	1.07
6 mo	2.10	−1.78	2.66	1.54	1.27	2.65	2.57	2.05	2.31	1.32	1.44	0.95	−0.30	1.73	1.46
12 mo	0.47	0.33	1.80	1.41	0.95	1.60	1.29	1.32	2.44	1.28	1.00	0.44	0.00	1.04	1.10

Panel C: Diff

	dp	dy	ep	de	svar	tbl	lty	ltr	tms	dfy	infl	bm	csp	ntis	Avg.
1 mo	2.17	2.45	1.96	1.33	2.07	2.86	2.98	3.97	2.90	1.95	2.95	2.11	1.80	1.89	2.38
3 mo	−0.50	1.36	1.34	0.85	0.83	1.28	1.27	1.32	1.81	0.92	1.76	0.62	−0.22	0.58	0.94
6 mo	2.17	−1.75	2.07	1.19	1.11	2.42	2.19	1.79	2.11	1.14	1.39	0.89	−0.14	0.91	1.25
12 mo	1.26	1.11	1.68	1.27	0.76	1.72	1.17	1.24	2.30	1.50	1.00	1.02	−0.09	0.80	1.19

This table tests our model's robustness by comparing the OOS CE gains over horizons that are successively longer (3, 6 and 12 months) than the baseline specification (i.e. one month horizon). Panels A, B and C correspond to Δ CE(OLS), Δ CE(OLSCE) and Diff, respectively. The sample period is the same as in the previous tables (1946 to 2023). The 14 variables from Goyal and Welch (2008)'s empirical study are: divided price ratio (**dp**), dividend yield (**dy**), earnings price ratio (**ep**), dividend payout ratio (**de**), stock variance (**svar**), treasury bill rate (**tbl**), long term yield (**lty**), long term rate of return (**ltr**), term spread (**tms**), default yield spread (**dfy**), inflation (**infl**), log of book to market ratio (**bm**), cross-sectional premium (**csp**), and net equity expansion (**ntis**).

Table 17: Descriptive Statistics of 17 Alternative Predictor Variables

	Period	Mean	Median	Std. Dev.	Skewness
acib	2005–2020	−0.121	−0.122	0.107	-0.102
apib	2005–2020	−0.106	−0.117	0.100	0.448
ivs	2005–2020	−0.005	−0.007	0.014	−1.054
vix	2005–2020	0.192	0.163	0.088	2.016
vrp	2005–2020	0.090	0.089	0.379	−7.737
icib	2005–2020	0.092	0.078	0.186	0.390
ipib	2005–2020	0.248	0.255	0.170	−0.069
bw sentiment	2005–2020	−0.085	−0.098	0.314	−0.122
pls sentiment	2005–2020	−0.303	−0.330	0.417	0.493
gm sentiment	2005–2020	0.029	0.025	0.016	0.703
hlw disp	2005–2020	−0.163	−0.293	0.703	0.140
gw pc 1	2005–2020	−0.542	−0.478	0.916	−1.116
gw pc 2	2005–2020	0.000	−0.177	1.200	2.885
gw pc 3	2005–2020	0.423	0.568	1.039	−1.481
michigan sentiment	2005–2020	83.009	84.100	12.056	−0.393
manager sentiment	2005–2020	0.156	0.199	0.711	−0.413
pnbo	2005–2020	−0.131	−0.053	0.517	−1.395

This table provides descriptive statistics of the variables during the sample period 2005–2020. The 17 alternative predictor variables are from Cao et al. (2023): aggregate call order imbalance (**acib**), aggregate equity put option order imbalance (**apib**), aggregate implied volatility spread (**ivs**), cboe vix (**vix**), variance risk premium (**vrp**), index call option order imbalance (**icib**), index put option order imbalance (**ipib**), bw sentiment (**bw sentiment**), pls sentiment (**pls sentiment**), hlw disagreement (**hlw disp**), first principle component of the 22 predictors based on Goyal and Welch (2008) (**gw pc 1**), second principle component of the 22 predictors based on Goyal and Welch (2008) (**gw pc 2**), third principle component of the 22 predictors based on Goyal and Welch (2008) (**gw pc 3**), consumer sentiment index from the University of Michigan (**michigan sentiment**), manager sentiment index (**manager sentiment**), and aggregate purchase of DOTM SPX index put options (**pnbo**).

Table 18: Comparison of OOS Certainty Equivalent Gains

	Period	Δ CE(OLS)	Δ CE(OLSCE)	Diff
acib	2005–2020	1.42%	3.30%	1.89%
apib	2005–2020	−10.27%	2.76%	13.03%
ivs	2005–2020	−1.71%	3.82%	5.53%
vix	2005–2020	0.29%	5.68%	5.40%
vrp	2005–2020	−4.57%	3.25%	7.82%
icib	2005–2020	−3.19%	3.29%	6.48%
ipib	2005–2020	−2.80%	3.25%	6.05%
bw sentiment	2005–2020	−0.92%	4.17%	5.09%
pls sentiment	2005–2020	0.40%	3.09%	2.69%
gm sentiment	2005–2020	−1.73%	3.29%	5.02%
hlw disp	2005–2020	2.82%	2.85%	0.03%
gw pc 1	2005–2020	−1.69%	3.07%	4.76%
gw pc 2	2005–2020	−3.45%	3.47%	6.92%
gw pc 3	2005–2020	−2.03%	3.34%	5.37%
michigan sentiment	2005–2020	−0.14%	3.61%	3.76%
manager sentiment	2005–2020	2.80%	2.57%	−0.23%
pnbo	2005–2020	0.50%	2.53%	2.03%
Average:		−1.43%	3.37%	4.80%

This table presents statistics on forecast errors by comparing the OOS CE (OLSCE) gains for an investor who has mean-variance (MV) utility function with risk aversion coefficient $\gamma = 3$ as in Table 9. The OOS CE gain of an MV investor is computed using the same expression as in Equation (11) below, but using the OOS observations instead:

$$\min_{\alpha,\beta} \sum_{t=1}^{T-1} (R_{t+1} - R_t^f - \alpha - \beta X_t)^2 - \lambda \left[\frac{1}{T-1} \sum_{t=1}^{T-1} (w_t R_{t+1} + (1-w_t) R_t^f) - \frac{\gamma}{2} \frac{1}{T-2} \sum_{t=1}^{T-1} (w_t R_{t+1} + (1-w_t) R_t^f - \bar{R})^2 \right]. \tag{17}$$

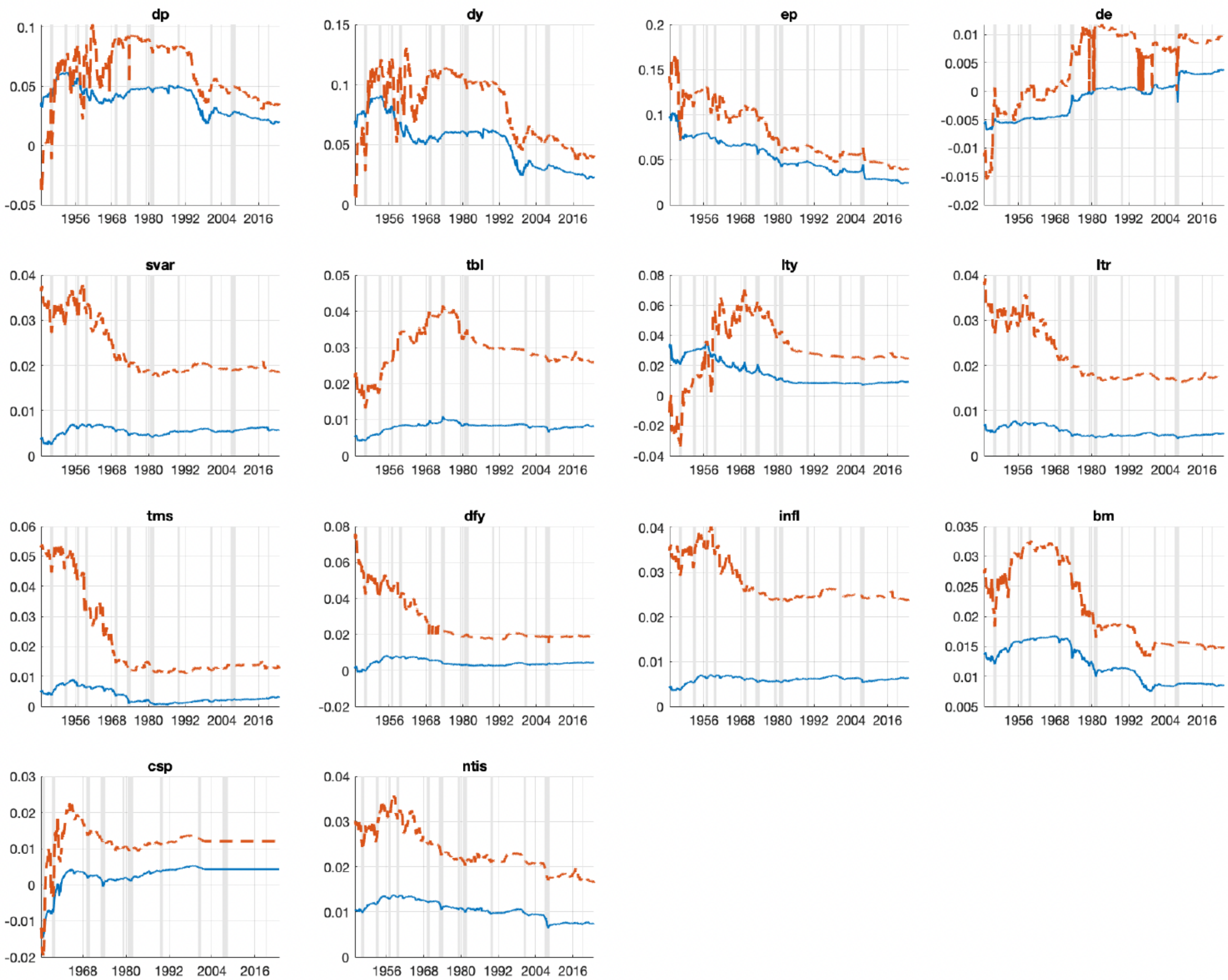
The 17 alternative predictor variables are from Cao et al. (2023): aggregate call order imbalance (**acib**), aggregate equity put option order imbalance (**apib**), aggregate implied volatility spread (**ivs**), cboe vix (**vix**), variance risk premium (**vrp**), index call option order imbalance (**icib**), index put option order imbalance (**ipib**), bw sentiment (**bw sentiment**), pls sentiment (**pls sentiment**), hlw disagreement (**hlw disp**), PC1 of Goyal and Welch (2008) (**gw pc 1**), PC2 of Goyal and Welch (2008) (**gw pc 2**), PC3 of Goyal and Welch (2008) (**gw pc 3**), consumer sentiment (**michigan sentiment**), manager sentiment (**manager sentiment**), and aggregate purchase of DOTM SPX index put options (**pnbo**). The **Diff** column shows noticeable differences between the two models across these variables' CE gains.

Table 19: Average Estimated Coefficients for OLS and OLSCE Approach

	$\bar{\alpha}$ OLS	$\bar{\alpha}$ OLSCE	$\bar{\beta}$ OLS	$\bar{\beta}$ OLSCE
acib	−0.003	−0.018	−0.051	−0.083
apib	−0.001	−0.003	−0.049	0.031
ivs	0.009	−0.009	0.625	0.206
vix	0.012	−0.044	−0.044	0.200
vrp	−0.002	−0.018	0.043	0.040
icib	0.005	−0.003	−0.033	−0.098
ipib	0.008	−0.002	−0.021	−0.029
bw sentiment	0.001	−0.018	−0.007	−0.001
pls sentiment	−0.002	−0.034	−0.010	−0.022
gm sentiment	0.013	0.012	−0.311	−0.605
hlw disp	0.003	−0.003	−0.016	−0.013
gw pc 1	0.002	−0.009	−0.004	−0.004
gw pc 2	0.004	−0.009	−0.007	0.003
gw pc 3	0.002	−0.013	0.005	0.009
michigan sentiment	−0.024	0.006	0.000	0.000
manager sentiment	0.007	−0.004	−0.014	−0.007
pnbo	0.001	−0.004	−0.009	0.001

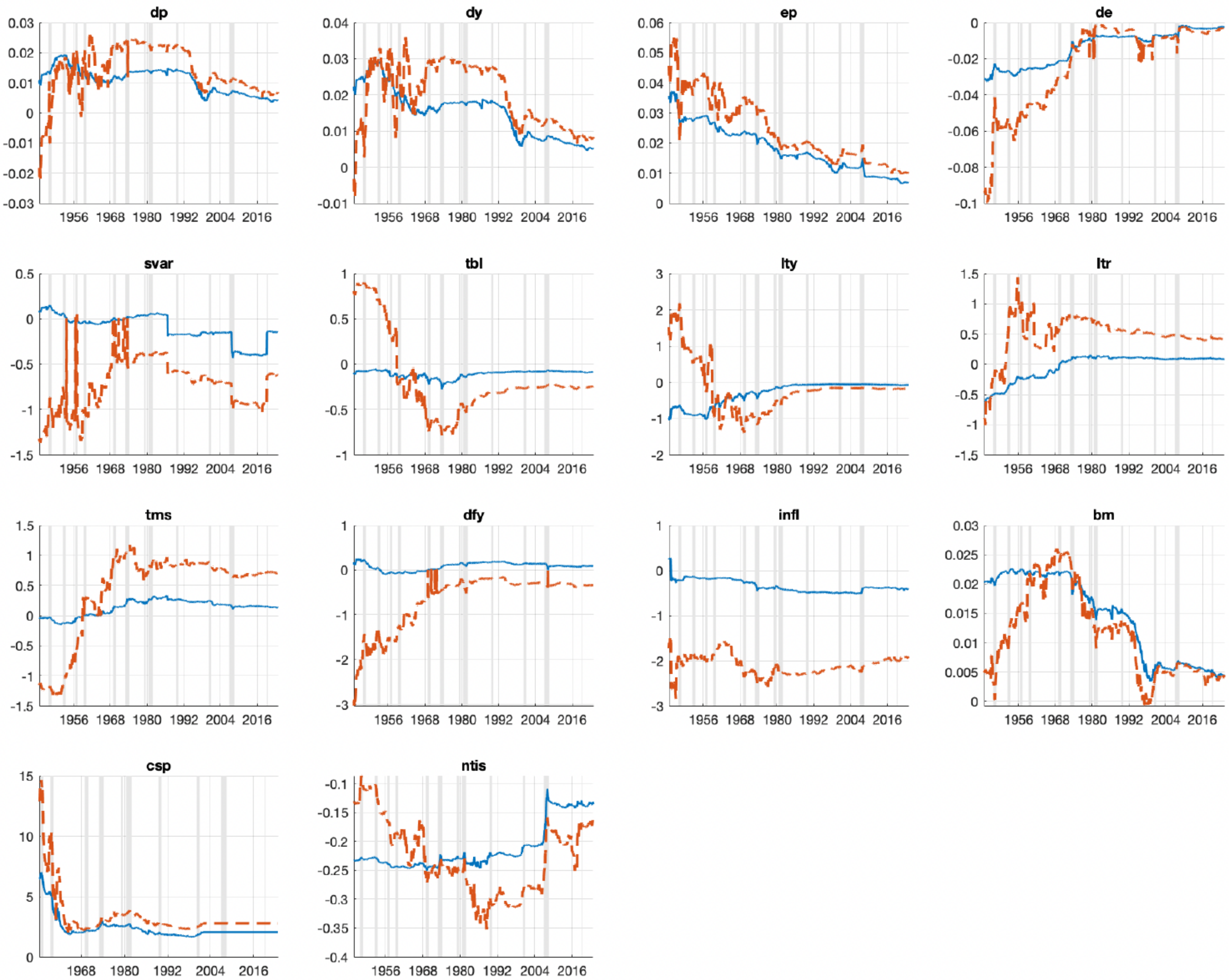
This table reports the time-series average of the coefficients α and β is estimated for an investor who has mean-variance (MV) utility function with risk aversion coefficient $\gamma = 3$. It is notable that the magnitude of the average estimated loadings on the predictor variable, $\bar{\beta}$ is much larger when using our CE method, compared to the OLS method. The 17 alternative predictor variables are from Cao et al. (2023): aggregate call order imbalance (**acib**), aggregate equity put option order imbalance (**apib**), aggregate implied volatility spread (**ivs**), cboe vix (**vix**), variance risk premium (**vrp**), index call option order imbalance (**icib**), index put option order imbalance (**ipib**), bw sentiment (**bw sentiment**), pls sentiment (**pls sentiment**), hlw disagreement (**hlw disp**), PC1 of Goyal and Welch (2008) (**gw pc 1**), PC2 of Goyal and Welch (2008) (**gw pc 2**), PC3 of Goyal and Welch (2008) (**gw pc 3**), consumer sentiment (**michigan sentiment**), manager sentiment (**manager sentiment**), and aggregate purchase of DOTM SPX index put options (**pnbo**).

Figure 12: Time-series of α estimates with respect to each predictor variable



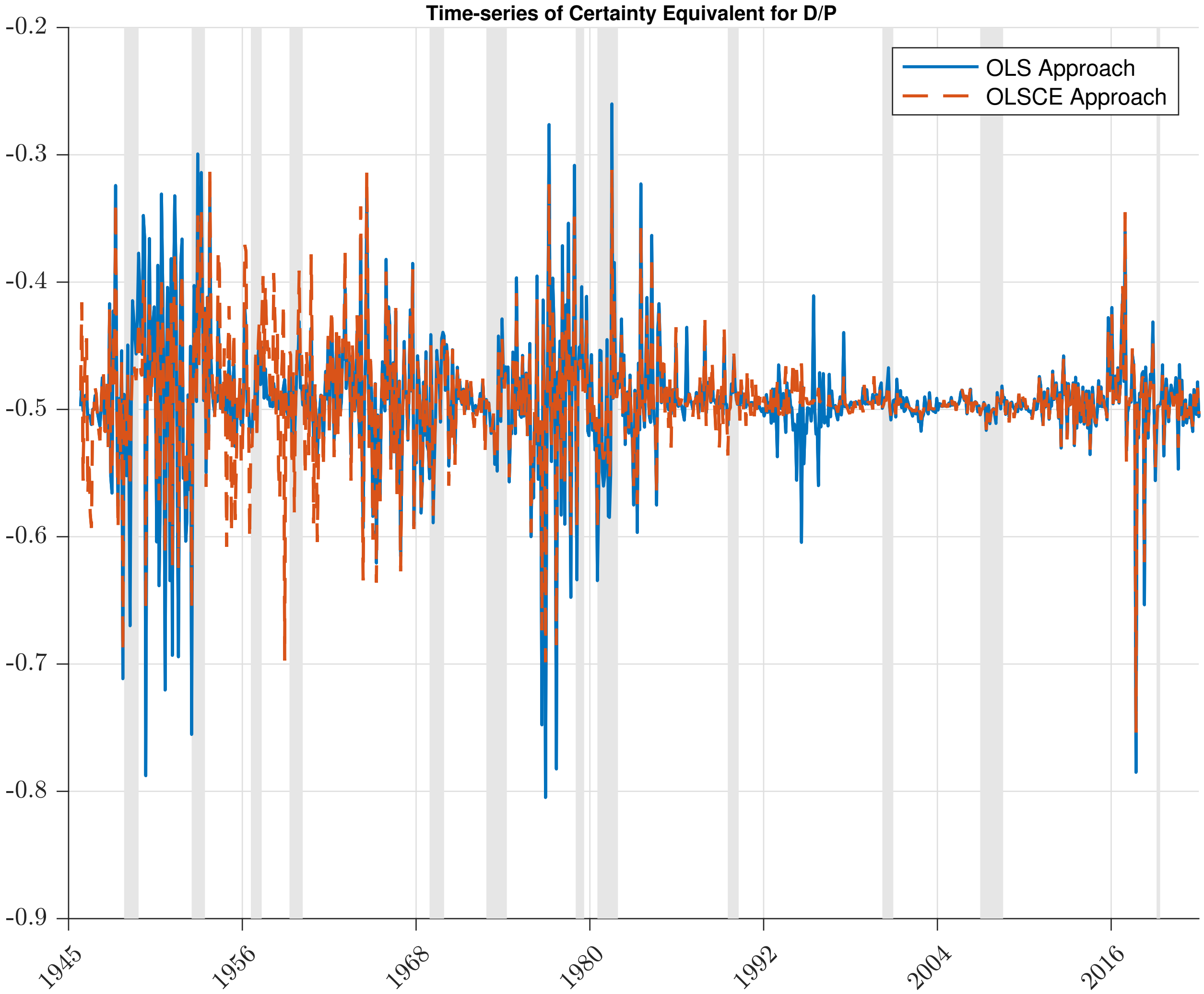
This figure plots the rolling estimates of each estimate of α corresponding to a different predictor variable, for the OLS approach (the solid blue line) vs. our approach (the dashed orange line) capturing the benefit of aligning in-sample loss function and OOS evaluation metric. The α estimated using OLSCE approach adjusts to the fluctuations in the business cycle (with different weighting for each period from the embedded stock market variance component) faster than the OLS approach (with equal weight to all observations). The grey areas indicate NBER-defined recessions. The out-of-sample monthly forecasts are obtained with an expanding window beginning with an initial window of 20 years for the sample period of January 1927 to December 2023.

Figure 13: Time-series of β coefficient estimates for each predictor variable



This figure plots the rolling estimates of each predictor variable's β coefficient, for the OLS approach (the solid blue line) vs. our approach (the dashed orange line) capturing the benefit of aligning in-sample loss function and OOS evaluation metric. The β estimated using OLSCE approach adjusts to the fluctuations in the business cycle (with different weighting for each period from the embedded stock market variance component) faster than the OLS approach (with equal weight to all observations). The grey areas indicate NBER-defined recessions. The out-of-sample monthly forecasts are obtained with an expanding window beginning with an initial window of 20 years for the sample period of January 1927 to December 2023.

Figure 14: Time-series of OOS CRRA Utility of Optimal Portfolios Constructed using D/P as the Predictor Variable



This figure plots the time-series of period-by-period CRRA utility of OOS optimal portfolio returns. The predictor is **dp** (Dividend to Price ratio). The OLS approach (the solid blue line) delivers an extremely volatile time-series of utility gains while the OLSCE approach (the dashed orange line) delivers more stable utilities throughout the OOS period. When CRRA utility is used, investors benefit from using our economic regularization in comparison to the standard OLS approach. Our OLSCE approach also delivers higher CE gains over the simple historical average approach. The grey areas indicate NBER-defined recessions. The OOS monthly forecasts are obtained with an expanding window beginning with an initial window of 20 years for the sample period of January 1927 to December 2023.

Chapter 3

What is the Implied Upper Bound of the Stochastic Discount Factor?

3.1 Introduction

Our understanding of the cross-section of expected stock returns is based on a generally accepted set of test assets. This set of test assets is taken as a set of primitives that are used to build tradeable risk factors (or anomaly portfolios). The initial search for factors comes from either economic theory or data mining. More recently, there has been a growing concern regarding whether the large and growing set of factors (i.e. the "Factor Zoo" coined by Cochrane (2011)) is primarily a consequence of data mining efforts over the past several decades (see Chen et al. (2024); McLean and Pontiff (2016) for recent and prominent examples). It remains unclear whether we have identified a set of test assets that spans the Stochastic Discount Factor (SDF). This paper (i) documents new evidence that shows how the traditional set of test assets fails to span the SDF; and (ii) introduces an alternative approach to building test assets based on hedge fund portfolios that gets us closer to spanning the SDF. Figure 15 illustrates how the current set of anomalies makes up only a small portion of the relevant space of test assets needed to span the SDF. This stylized representation showcases how the extant set of anomalies fails to span the SDF, thereby motivating the research question in this paper.

The traditional approach to generating test assets is best exemplified by Fama and French (1993). In essence, they show considerable variation in stock returns along two firm characteristics: Size and Book-to-Market (B/M). As a consequence, they sort stocks into quintiles along these two dimensions thereby generating 25 Size-B/M portfolios (i.e. test assets). With this set of test assets they create tradable long/short (L/S) factors that buy (sell) the quantile with the highest (lowest) return: Size (SMB) and B/M (HML). Since Fama and

French (1993), the empirical asset pricing literature has proposed 100+ factors to explain the cross-sectional variation in stock returns. However, McLean and Pontiff (2016) note how the out-of-sample (OOS) return predictability of the most notable factors has fallen dramatically over time (58% lower post publication). If these factors in fact span the SDF, then it follows that more than half of the return predictability in stock returns has disappeared over time as many of these anomalies have been traded away.

The initial motivating evidence indicating that the current set of test assets fails to span the SDF is shown in Figures 16 and 17, in which first moments are emphasized in Figure 16 whereas Figure 17 emphasizes second moments. Figure 16 Panel A displays the rolling ten-year Sharpe ratio (SR) of an equal-weighted combination portfolio of the Jensen et al. (2023) 13 factor themes. There is a steady decline in the risk-adjusted returns of anomalies since the 2000 Technology Bubble. Panel B of Figure 16 displays the rolling ten-year information ratio (IR) of daily CAPM alphas of an equal-weighted combination portfolio of the 13 factor themes versus the value-weighted index excess return. Similar to Panel A, there is also a notable downward trend in IRs over the same period. Together, both panels of Figure 16 reveal the basic insight of this paper that builds on the evidence from McLean and Pontiff (2016) documenting how the anomaly returns have declined steadily over time (as evidenced in Panel A). If the anomaly portfolios indeed span the SDF then the logical implication is that risk premia have declined over time across all major assets (including the market portfolio). However, the market's risk premium has increased over time *relative* to the anomaly portfolios, as evidenced by Panel B of Figure 16. Therefore, the anomaly portfolios do not span the SDF.

Figure 17 suggest that the anomaly portfolios have become less relevant over time. Panel A displays the realized volatility of the value-weighted market index (black line) in comparison to the realized volatility of an equal-weighted combination portfolio of the Jensen et al. (2023) 13 anomaly portfolios (red line). Panel B displays the 20-year rolling correlation between the two realized volatilities with a linear trendline. Realized volatility is useful because it indicates periods when risk premia are elevated. That is, when realized volatility of the aggregate market spikes upward, it is reasonable to assume that risk premia are elevated. If upward spikes in realized volatility of the market capture times when risk premia are elevated we should expect to see similar time-variation in the realized volatility of the anomalies portfolio. Surprisingly, we get a very different characterization of time-varying risk premia, with respect to the anomaly portfolios, beginning with the 2000 Tech Bubble (i.e. the correlation drops to 0.28 from 0.82 during this one period). While both panels indicate elevated risk premia during the first half (e.g. the 1929 Stock Market Crash and World War II), the set of anomalies suggest that the most recent periods of market turbulence (e.g. 2008 Financial Crisis) are pretty unexceptional historically in comparison to the 2000 Tech Bubble which generates realized volatility on the same order of magnitude as the 1929 Stock Market Crash and Great Depression. This leads to a significantly negative trendline in the correlation between the two time series of realized volatility. All together, the key takeaway is that not only are the anomalies telling us in their first moments (Figure 16) that they have lost their relevance since the 2000 Tech Bubble but we see the exact same story in terms of their second moments (Figure 17).

Ultimately, if return predictability exists, it is a direct reflection of changes in investors' subjective beliefs. So long as investor behavior has not changed over the past century we should expect predictable reactions (i.e. predictable changes in investors' subjective beliefs) to new and important sources of unforeseen events, manifested in large spikes in realized volatility. Hence, if the current set of test assets indeed spanned the SDF, then we should not see the significant departure in realized volatility between the market and the anomalies beginning with the 2000 Tech Bubble documented in Figure 17. This evidence suggest the current set of test assets does not adequately characterize risk premia (i.e. span the SDF) and it is incumbent on us to search for a new set of test assets.

Instead of using traditional test assets to span the SDF, we can use hedge fund portfolios.[32] Identifying expected returns indirectly via hedge funds can be viewed as analogous to something that is entirely unrelated to asset pricing that is often more sucessful than following a direct approach: looking for where the birds are to find fish. While this example may seem far afield from asset pricing, this observation offers us a useful analogy for identifying (unobservable) expected returns indirectly as opposed to the traditional method in empirical asset pricing of directly approximating expected returns with test assets based on firm characteristics (Fama and French, 1993). This indirect approach is implicitly based on the modus operandi of a hedge fund which is to harvest return predictability. Given that hedge funds are affected in a very real sense by their ability (or inability) to benefit from return predictability, they offer a promising alternative as test assets that span the SDF.

I propose the top three deciles of hedge fund returns (by month) as a parsimonious representation (i.e. factor model) of hedge fund test assets. I call this the implied upper bound of the SDF.[33] This upper bound can be viewed as analogous to the approach by Shiller (1981) in the following sense. Shiller (1981) constructed the implied price of the aggregate index based on perfect foresight of future dividends of the aggregate index. In our setting, the upper bound of the SDF can be viewed as equivalent to the perfect foresight SDF. The perfect foresight SDF is simply based on those factors that performed the best, ex-post, in any given period. While this implied upper bound of the SDF includes unpriced risk (i.e. noise), it avoids the issue of having to take a stance on the factors necessary to build expected returns. It is for this reason that this paper emphasizes the time-series dynamics of the implied upper bound which should mitigate the issue of unpriced risk as it is differenced out over time.

I argue that the implied upper bound of the SDF is useful for two main reasons. First, it provides an objective way of evaluating test assets with respect to whether they help

[32]This unconventional choice of test assets is in part motivated by the recent evidence of Koijen et al. (2023) that documents hedge funds have the largest impact on equity prices per dollar of wealth. I discuss the limitations of using hedge fund portfolios as test assets in Section 3.4.

[33]It is important to note that the implied upper bound of the SDF can be built using any set of test assets. Later in the paper I build an implied upper bound of the SDF based on a factor model comprised of the leading five principal components extracted from the Jensen et al. (2023) anomaly portfolios.

us get closer to spanning the SDF. Second, it forces us to take more seriously the entire factor structure and its dynamics. In doing so, it has the potential to add a more nuanced understanding to the strength of the factor structure and one of the most important economic priors we have in asset pricing (i.e. the absence of near arbitrage opportunities).

In this paper, I apply the methodology of the implied upper bound to characterize expected stock returns. First, I determine what the implied upper bound of the SDF is based on the Jensen et al. (2023) factor theme anomaly portfolios and I examine its factor structure and dynamics over the sample period. Second, I determine what the implied upper bound of the SDF is based on hedge fund decile portfolios and investigate its factor structure and dynamics over its respective sample period. This approach which takes the implied upper bound of the SDF seriously allows us to avoid the endless debate as to whether something is priced and instead focus our attention on how informative the test assets are.

In the cross-section of anomaly portfolios, I find that while the leading five principal components (PCs) are able to explain most of the variation in expected returns, the higher order PCs appear to also be a robust feature which is an important departure from the existing literature (see Kozak et al. (2018, 2020b) for important studies who document the strong factor structure in the cross-section of stock returns). In a pseudo OOS analysis the PC1-5 maximum squared SR is 36% (24%) of the anomaly test asset's maximum squared SR during the IS (OOS) period. This is evidence that the PC factor model gets closer to the mean-variance frontier of anomaly portfolios during the IS period compared to the OOS period. In other words, the higher order PCs that are commonly excluded from factor models are a robust feature in explaining stock returns during both the IS and OOS period. This is in sharp contrast to previous studies that suggest the higher order PCs are no longer robust during the OOS period. The second interesting finding regarding the cross-section of anomaly portfolios is that the factor structure is largely dominated by only three of the 13 factor themes (Value, Low Leverage and Seasonality explain nearly half of the SDF's variance) with the Value Factor Theme contributing 23% to the SDF's variance on its own.

Importantly, this new evidence tells us that many well-known factors (e.g. Investment, Profitability, etc.) play a much more limited role in contributing to the maximum squared SR of the economy.

The time-series of the SDF's variance (based on anomaly portfolios) displays interesting dynamics. First, the variance of the SDF spikes upward during market dislocations (e.g. 2008 Financial Crisis).[34] The maximum squared SR has significant variation ranging between 1.67 to 46.62. This evidence is consistent with the original motivation in Figure 17 regarding elevated levels of risk premia during periods of market turbulence. The second finding which is perhaps more suprising is that the SDF's variance based on the factor structure is notably different than that based on the underlying test assets. In particular, the factor structure is largely a product of one single market dislocation: the 2000 Technology Bubble. More specifically, the maximum squared SR during the Tech Bubble is twice as large as that experienced during the onset of the Russia-Ukraine war in early 2022 and nearly three times as large as that experienced during the 2008 Financial Crisis. We learn from this evidence that our understanding with respect to the factor structure of stock returns based on the anomaly portfolios is in large part shaped by the return variation from two decades ago during the 2000 Tech Bubble. Put differently, the more recent periods of market turbulence (e.g. 2008 Financial Crisis and 2020 COVID-19 Crash) that created historic investment opportunities are significantly under-represented by the test assets that we often assume characterize investment opportunities.

Importantly, the evidence in this paper regarding the disproportionate impact of the 2000 Tech Bubble is consistent with both Jensen et al. (2023) and Chinco et al. (2021). In particular, one of the most salient features in Figure 2 of Jensen et al. (2023), documenting the anomalies portfolio cumulative CAPM alpha, is the significant outperformance relative

[34] I refer to market dislocations as any period that is characterized by market turbulence (i.e. an upward spike in realized volatility of the aggregate market index) in which market prices deviate more from their fundamental values. The most recent market dislocations include the onset of the Russia-Ukraine War in early 2022, the COVID-19 Stock Market Crash in March 2020, the 2008 Financial Crisis and the 2000 Technology Bubble.

to the market around the Tech Bubble. Roughly one third of the gains over their sample period from 1990 to 2020 are due to this market dislocation. The outsized influence on CAPM alphas from this event is entirely consistent with my findings regarding its dominance in the factor structure of the anomaly portfolios. Moreover, Figure 1 of Chinco et al. (2021) displays the evolution of the prior variance (i.e. the ex-ante probability of witnessing a tradable anomaly dubbed the "anomaly base rate") over their sample period from 1978 to 2015 which *peaks* during the Tech Bubble. This evidence further reinforces the central message that our understanding of the cross-section of stock returns is shaped by this one period. A factor structure that is dominated by the Tech Bubble from two decades ago (and is to a large extent unaffected by the more recent bouts of market stress including the 2008 Financial Crisis and COVID-19 Crash in March 2020) illustrates that the anomaly portfolios have lost their relevance as test assets. All together, the skewed time-series and cross-sectional evidence of the SDF's Variance indicates that the anomaly test assets do not span the SDF.

When we instead use hedge fund portfolios as test assets we end up getting a very different picture of the SDF's variance both cross-sectionally and in the time-series. With respect to the cross-section of hedge fund portfolios, there is evidence of persistence in the top decile hedge funds whereby the OOS SR is 5.43 compared to an IS SR of 5.27, in sharp contrast to the anomaly test assets whose superior IS risk-adjusted returns disappear during the OOS period. This evidence documents that near arbitrage opportunities are persistent in contrast to Kozak et al. (2018). That is, we can no longer conclude that near arbitrage opportunities do not exist, since we have a set of test assets that show that they are robust during both the IS and OOS periods. It turns out that the evidence informing our economic prior regarding the absence of near arbitrage opportunities, is a product of the fact that the anomaly portfolios no longer span the SDF. Moreover, while the OOS maximum squared SR of the anomaly test assets is one fifth the size of its IS counterpart, the hedge fund portfolio maximum squared SR more than doubles in the OOS period. Most importantly, using hedge fund portfolios as test assets gets us closer to spanning the SDF by producing a SR that is 3.4x higher than the traditional anomaly portfolios.

Consistent with anomaly portfolios as test assets, the time-variation of the SDF based on hedge fund portfolios also spikes upward during market dislocations. The key difference is that the Tech Bubble no longer dominates the factor structure of hedge fund portfolios. That is, the factor structure changes significantly across all of the most recent market dislocations. With that evidence we learn that near arbitrage opportunities occur from a more diverse set of periods characterized by market turbulence which is again consistent with the initial motivating piece of evidence in Figures 16 and 17 regarding time-varying risk premia. More specifically, time-varying risk premia reflect changes in investors' subjective beliefs, which are largely affected by new and important sources of exogenous events. Figure 17 makes clear that we have witnessed many of these events since the 2000 Tech Bubble that the anomaly portfolios fail to adequately capture in comparison to hedge fund portfolios. In essence, hedge fund portfolios better characterize the investment opportunity set relative to the existing set of anomaly test assets.

This paper contributes to two strands of the literature. First, I contribute to the active literature that studies the factor structure of the cross-section of expected stock returns (see Bryzgalova et al. (2023a,b); Daniel et al. (2020); Feng et al. (2020); Giglio et al. (2023c); Kelly et al. (2019); Kozak and Nagel (2023); Kozak et al. (2018, 2020b); Lettau and Pelger (2020) for recent and important examples). I focus on two important papers which are most closely related to this study that have shaped how we think in empirical asset pricing. In a seminal paper, Kozak et al. (2018) provides robust evidence that we can use reduced form models from PCs of the covariance matrix of returns given their superior ability to capture the cross-sectional variation within stock returns. Moreover, they derive a closed-form expression (see Equation 4 of their paper) that shows the variance of the SDF is a function of the factor structure. All together, their empirical evidence coupled with their analytical result leads to the conclusion that near arbitrage opportunities do not exist in the cross-section of stock returns. In a subsequent and equally influential paper, Kozak et al. (2020b) use a Bayesian approach to build an SDF from a larger set of 50 L/S anomaly portfolios. Given their earlier work, they adopt the novel prior that first and second moments should be linked and in fact find that the leading PCs do a good job explaining this higher-dimensional setting of anomalies. I build on their work by showing that while the larger set of anomalies also exhibits a strong factor structure over the full sample period (i.e. unconditionally), the factor structure of anomaly portfolios is highly *unstable* over time. Moreover, I show that their conclusion with respect to ignoring the lower ranked PCs is no longer warranted (during market dislocations) when we instead use hedge fund portfolios as test assets. Hence, this new evidence motivates us to adopt a more nuanced prior regarding the cross-section of stock returns. More specifically, we need to relax the strong prior regarding the absence of near arbitrage opportunities during periods of market stress.

Second, I contribute to the literature that builds (or chooses) test assets which help characterize the SDF (see Ahn et al. (2009); Bryzgalova et al. (2023c); Giglio et al. (2023c) for important recent examples). In a pathbreaking paper, Bryzgalova et al. (2023c) revisits the important question as to whether we have identified portfolios that characterize the investment opportunity set (test assets) and offers an entirely novel approach called "Asset Pricing Trees (AP-Trees)" that can accomodate complex interactions and higher dimensionality among groups of stocks that conventional portfolio sorts are unable to incorporate. The AP-Trees method achieves a SR that is three times higher OOS compared to simple L/S factors which indicates that their approach gets us much closer to spanning the SDF compared to the traditional anomaly portfolios. I complement their study by documenting novel evidence that the SDF is not spanned by traditional test assets, thereby lending further justification regarding the introduction of alternative methods to construct test assets. In addition, I contribute to this literature by introducing an alternative approach to generating test assets (i.e. hedge fund portfolios) that are able to generate a SR that is 3.4x higher than the traditional anomaly portfolios.

The article proceeds as follows. Section 3.2 details the methodology and data used in the paper. The main results are presented in Section 3.3. Section 3.4 documents the findings regarding the new set of hedge fund test assets. Section 3.5 concludes.

3.2 Methodology, Data and Summary Statistics

This section details the construction of the SDF's upper bound, the data used in this paper and descriptive statistics which provide an overview of the data. The methodology follows Kozak et al. (2018) with respect to its application to anomaly portfolio excess returns. The main departure is with respect to applying Equation 18 to hedge fund portfolios.

I estimate the minimum-variance SDF as in Hansen and Jagannathan (1991),

$$M = 1 - \mu' \Gamma^{-1} (R - \mu) \quad \Rightarrow \quad \mathbb{E}[M] = 1 \text{ and } \operatorname{Var}[M] = \mu' \Gamma^{-1} \mu, \tag{18}$$

where the mean of portfolio excess returns is represented by $\mu = \mathbb{E}[R]$ and Γ represents the variance-covariance matrix of portfolio excess returns.

3.2.1 The Upper Bound of the SDF

One of the most important results in asset pricing is from Hansen and Jagannathan (1991) who derive the famous Hansen-Jagannathan Bound (HJB),

$$\frac{\sigma[M]}{\mathbb{E}[M]} \geq \text{Maximum SR in the Economy} \tag{19}$$

That is, with the SDF normalized to have mean equal to one, the variance of the SDF is equal to the maximum squared SR of the test assets.

Kozak et al. (2018) note how a factor can be both important in explaining the variance of returns and unimportant in explaining expected returns (i.e. the pricing performance of the factor). The pricing performance of a given factor depends on the covariance of a test assets' returns with a given factor. In essence, there is an important distinction between explaining return variances versus explaining expected returns. Said differently, there is some return variation that is not priced and should be viewed as noise (Daniel et al., 2020).

If the set of test assets does not span the SDF, these covariances are less informative. Most importantly, expected returns are inherently unobservable and depend crucially on the set of test assets. It is for this very reason that I rely instead on a factor's ability to explain return variances which avoids the empirical issue as to which part of return variances concerns expected returns. Hence, I call the implied upper bound of the SDF the reduced-form factor model implied by a set of test assets (e.g. the implied upper bound of the SDF can be the PC1-5 factor model whereby the leading five PCs are extracted from the variance-covariance matrix of anomaly excess returns). That is, the implied upper bound of the SDF is based on maximizing the explained variance from a set of test assets,

$$M^* \approx 1 - f'\Sigma^{-1}\left(F - f\right), \tag{20}$$

where F are the excess returns from the leading K principal components with $f = \mathbb{E}F_k$.

By taking the upper bound of the SDF (and its time series dynamics) seriously, I am making the following implicit assumption: the proportion of the SDF's variance that is simply noise is relatively constant over time (i.e. homoskedastic). This assumption seems relatively mild since the proportion of the SDF's variance attributable to risk premia should be time-varying since it reflects some combination of investors' risk preferences and belief distortions. Hence, the time-varying nature of the SDF's variance should reflect time-variation in risk premia.

With the upper bound of the SDF defined, I now delineate how I estimate factor models with anomaly portfolios versus hedge fund portfolios.

3.2.2 Anomaly Portfolios

I begin with a large, comprehensive and widely accepted set of anomaly factor excess returns (i.e. test assets). Equation 18 is first applied to the original set of 13 factor themes which produces the SDF based on the test assets. As in Kozak et al. (2018), I also conduct Principal Component Analysis (PCA) to the anomaly portfolios to produce factor models ranging from the first leading PC to a five factor model comprised of the leading five PCs. With that, I estimate the factor-based SDF as before with the anomaly portfolios substituted for the PC factors.

3.2.3 Hedge Fund Portfolios

The standard approach to forming decile portfolio returns is to generate the decile portfolios over the entire sample. By computing the deciles cross-sectionally, it avoids the issue that explaining the set of test assets' returns is tautological. That is, we are explaining expected returns cross-sectionally, by construction. However, it is not obvious whether the cross-sectional explanatory power will have the same power along the time-series dimension. Given the paper's scope of interest (i.e. to estimate the upper bound of the SDF) I deviate from the convention by computing deciles each month which maximizes the ability to explain the set of hedge fund returns both cross-sectionally and in the time-series:[35]

$$D_i = \begin{cases} 1 & \text{if hedge fund return is in decile } i = 1 \\ 2 & \text{if hedge fund return is in decile } i = 2 \\ \vdots & \\ 10 & \text{if hedge fund return is in decile } i = 10 \end{cases} \tag{21}$$

With that, the implied upper bound based on this new set of test assets (hedge fund

[35] I choose to form decile portfolios (over quintile or tertile portfolios) to reflect the minority of hedge fund managers with skill.

decile portfolios), I compute the implied upper bound in the following way,

$$M^* \approx 1 - d'\Sigma^{-1}(D - d), \tag{22}$$

where D are the excess returns from the top three deciles (D_8, D_9, and D_{10}) with $d = \mathbb{E}D_i$.

The decile by month approach is highly related to conducting PCA which maximizes the variance of the original set of test assets for a given number of PC factors. More specifically, PCA involves an eigenvalue decomposition of the variance-covariance matrix of portfolio excess returns which yields N orthogonal factors. Subsequent to this, the rotated test assets are then sorted from highest to lowest based on their eigenvalue (variance). Given the relatively short time series of hedge fund data relative to the number of hedge funds ($T << N$), the covariance matrix is unstable which is why we cannot apply PCA to the space of hedge fund returns. It is for this reason that I apply a similar method to PCA that is feasible (i.e. decile by month hedge fund portfolios) to maximize the variance explained both cross-sectionally and in the time-series of hedge fund returns.

3.2.4 Data

I consider a large publicly-available data set of factor excess returns from Jensen et al. (2023) ("JKP") who construct 153 factors spanning 93 countries from January 1926 to December 2023.[36] I have chosen to restrict the factor themes to the U.S. to make it more comparable to previous studies that have studied the factor structure within the cross-section of stock returns (Kozak et al., 2018, 2020b). In Section 3.3, the time series is restricted to start from January 1952 to ensure that all 13 factor themes are represented in the SDF.

In addition to the large set of anomaly returns, I use individual monthly hedge fund excess returns from the Lipper Trading Advisor Selection System Hedge Fund Commercial Database (hereafter TASS), accessed in September 2023 with a sample period from Jan-

[36] I have downloaded the daily U.S. capped value weighted factors across all 13 themes from https://jkpfactors.com/factor-returns.

uary 1996 to June 2023 comprised of 315,646 hedge fund month observations. The TASS data represents the most widely studied hedge fund database (Joenväärä et al., 2021). I have cleaned the TASS data to address the known biases present in hedge fund commercial databases (Pedersen, 2015).[37] Hedge fund returns are net of fees and are in excess of the monthly risk-free rate.

Lastly, I have downloaded the 25 Size-B/M portfolios and Fama and French (1992) MKT, SMB and HML ("FF3") factors at a daily frequency from Ken French's website.[38] In addition, I have downloaded the daily Daniel et al. (2020) ("DMRS") five efficient factors that hedge out unpriced risk with respect to the original Fama and French (2015) ("FF5") from Kent Daniel's website.[39] The DMRS factor sample period begins in July 1963 and ends in March 2023.

3.2.5 Summary Statistics

Table 1 presents the descriptive statistics (mean, volatility, SR and skewness) of the two main datasets: L/S anomaly portfolios (Panel A) and hedge fund decile portfolios (Panel B). The mean and volatility of returns are expressed in percent and have been annualized.

The average anomaly returns range from 0.25% (Low Leverage) to 4.30% (Momentum) per annum. The spread in factor theme volatilities is consideraly larger ranging from 1.72% (Seasonality) to 10.98% (Low Risk) per annum. It is somewhat surprising that the low risk anomaly has the highest volatility of the 13 factor themes. The average SR is 0.57 with a minimum and maximum of 0.03 (Low Leverage) and 1.19 (Debt Issuance). The variation in SRs is primarily driven by the variation in volatility across factor themes. Ten of the 13 Factor themes have mild negative skewness with the remaining three (Profitability, Short

[37] I have used the 3-step unsmoothing methodology proposed by Couts et al. (2020) to address the serial correlation present at both the strategy- and individual-level of hedge fund returns. The cleaned data has removed hedge funds whose primary strategy is labelled "Fund of Funds" given the high degree of autocorrelation in their returns. I have also removed any hedge funds whose primary strategy is labelled "Other" or "Undefined" consistent with Couts et al. (2020).

[38] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[39] http://www.kentdaniel.net/data.php

Term Reversal and Value) having more significant positive skewness between 0.56 to 0.94. The average annualized hedge fund decile excess returns range from -92.9% to 107.0%, with an average of 6.1%. The SRs (Skewness) range between -3.91 (-3.18) to 5.18 (2.01), with an average SR (Skewness) of 0.53 (-0.73). Interestingly, the average SR is roughly the same for both the anomaly portfolios and hedge fund decile portfolios (0.57 vs. 0.53, respectively). In contrast, the average hedge fund portfolio is more negatively skewed compared to the average anomaly portfolio (-0.73 vs. -0.18, respectively). This is unsurprising given that the anomaly portfolios are highly diversified both within and across factor themes, whereas hedge fund returns are notorious for their more concentrated market downside risks (Kelly and Jiang, 2012) which is evident in the bottom six deciles.

3.3 Main Results

In this section, I start by showing that the more comprehensive set (cross-sectionally and in the time-series) of anomaly returns from Jensen et al. (2023) ends up following a strong factor structure over the full sample period from January 1952 to December 2023. After confirming this well-established stylized fact in the empirical asset pricing literature, I document a new property of the factor structure. That is, the factor structure of anomaly returns is *highly time-varying* and has particularly interesting dynamics around the Technology Bubble.

3.3.1 Establishing a Baseline: Revisiting Kozak et al. (2018)

Table 21 shows that the first five PCs are able to explain most of the cross-sectional variation in expected returns of the anomaly portfolios. Table 21 revisits the analysis conducted in Table II of Kozak et al. (2018) with a set of anomaly portfolios that has been extended both in the time series (1952 to 2023 versus 1966 to 2015) and the cross-section (Jensen et al. (2023) 153 characteristics clustered into 13 factor themes versus 15 L/S anomaly strategies from Novy-Marx and Velikov (2016)). In keeping with Kozak et al. (2018), for the purposes of comparison, we evaluate the anomaly returns alongside the Fama and French (1993) 25 Size-B/M portfolios. The upper panel displays the annualized mean returns of the factor themes along with their alphas. The alphas are estimated from time series regressions of the L/S anomaly return on the leading PCs extracted from the underlying anomaly test assets. The last five columns correspond to successively larger factor models beginning with one PC (PC1) and ending with five PCs (PC1-5). The results are consistent with Kozak et al. (2018) findings that the first PC is inadequate in explaining anomaly returns (e.g. momentum which has the largest mean return of 4.30% has an alpha of 4.2% per year). With five PCs the pricing error for momentum has shrunk to 0.1% per year.

The bottom panel of Table 21 displays the (ex-post) maximum squared SR of the factor theme portfolios (9.31) and the respective PC factors. It is interesting to note how the maximum squared SR of this updated set of anomalies is more than twice the size as the value reported in Kozak et al. (2018) of 4.23. This simply reflects the more volatile period in equity markets since their sample ended in 2015, beginning with the COVID-19 pandemic. Aside from the more volatile SDF in this updated sample, the results are consistent with Kozak et al. (2018): the maximum squared SR of the PC factors is well below that of the test assets (PC1-5 model has a maximum squared SR of 1.76). All five of the factor models reject the null hypothesis of no pricing errors at the 1% level of statistical significance. In spite of the large HJ-distance, the PC factor model outperforms the traditional Fama and French (1993) FF3 factor model (maximum squared SR of 0.71). Interestingly, it is not able to beat the DMRS factor model (maximum squared SR of 1.99) based on the anomaly set of test assets which helps showcase the importance of removing unpriced risk present in test assets. The best performing model is the PC1-5 model based on the Fama and French (1993) 25 Size-B/M test assets that has a maximum squared SR of 2.07. All together, this updated evidence lends further support to the results in Kozak et al. (2018) that we can successfully build reduced-form factor models from PCA.

The last part of this subsection provides the final piece of evidence that confirms the baseline result first established by Kozak et al. (2018): the large HJ-distance observed in the bottom panel of Table 21 does not reflect persistent near-arbitrage opportunities. Figure 18 conducts the same pseudo OOS analysis from Figure 4 of Kozak et al. (2018) with our updated sample. That is, I divide the sample into two equal halves with the first (second) half treated as the IS (OOS) period. The two panels reflect scatterplots of the IS versus OOS SRs of the anomaly portfolios (Panel A) and the 25 Size-B/M Portfolios (Panel B). The OOS period begins December 31, 1987. Consistent with the findings in Kozak et al. (2018), near arbitrage opportunities are *not* persistent over time based on this analysis (i.e. nearly all of the test assets in both panels are well below the 45 degree line). The OOS SRs are, on average, 41% (61%) as large as the IS SRs for the anomaly (25 Size-B/M) portfolios. This poor OOS evidence is consistent with the data mining concerns regarding anomalies (McLean and Pontiff, 2016). With that (in addition to further robustness checks), Kozak et al. (2018) reach the conclusion that the higher order PCs contributon to the higher IS SR is not a robust feature of expected returns. In essence, expected returns (i.e. the SDF) can be nicely characterized by a reduced-form factor model with a handful of the leading PCs (e.g. a five factor model).

Before reaching the conclusion that the lower ranked PCs are of no consequence to the cross-section of expected stock returns (i.e. near arbitrage opportunities do not exist), it seems like the most natural exercise would be to conduct the analysis in the bottom panel of Table 21 for the two halves of the sample (IS vs. OOS) and determine if we get a similarly large HJ-distance for the IS and OOS periods. Table 22 performs this analysis. The upper (bottom) panel is based on estimating the SDF during the IS (OOS) period for both the anomaly and the 25 Size-B/M portfolios. There are two important takeaways from this exercise. First, the maximum squared SR is notably higher in the IS period compared to the OOS period for both sets of test assets. However, the second and more interesting finding is that the HJ-distance is relatively larger during the OOS period compared to the IS period for *both* sets of test assets. A factor model with the five leading PCs gets closer to the mean-variance frontier during the IS period compared to the OOS period. More specifically, the PC1-5 maximum squared SR is 36% (24%) of the anomaly test assets' maximum squared SR during the IS (OOS) period. Moreover, the PC1-5 maximum squared SR is 58% (24%) of the 25 Size-B/M test assets' maximum squared SR during the IS (OOS) period. By performing this subsample analysis with respect to the maximum squared SR, I reach a different conclusion to Kozak et al. (2018): The higher order PCs appear to be a robust feature of the cross-section of expected stock returns (i.e. the higher order PCs appear to play an even larger role in the OOS period compared to the IS period).

After reporting the surprising findings of Table 22 that at first glance seem contradictory to the results in Figure 18, two natural questions arise: (i) What does the SDF's variance (i.e. maximum squared SR) look like over time?; and (ii) What does the SDF's factor structure look like over time? The next set of empirical results are guided by these two questions which will help reconcile the contradictory findings between a lower OOS SR (compared to IS SR) and a weaker OOS factor structure (compared to the IS factor structure). Answers to these questions will shed more light on this finding regarding the factor structure and in doing so, will provide new insights on the nature of the underlying test assets.

3.3.2 On the Dynamics of the SDF's Variance

Figure 19 displays the one-year rolling variance of the SDF based on the JKP anomaly portfolios (blue line) and the five leading PCs extracted from the test assets (red line). The variance is estimated over the full sample period from January 1952 to December 2023. There are two key takeaways from this figure. First, the SDF (based on the underlying test assets) reveals an important property that is intuitive from an asset pricing perspective: The SDF's variance is highly time-varying and spikes upward during market dislocations (e.g. the most recent episodes include the 2000 Technology Bubble, 2008 Financial Crisis, 2020 COVID-19 Stock Market Crash and 2022 Russia-Ukraine War). This is intuitive since we should expect the maximum squared SR to be larger when risk premia are larger. By definition, risk premia are larger during market dislocations (i.e when market prices deviate more from their fair values). The maximum squared SR experiences a wide range from 1.67 (February 1978) to 46.62 (April 2023). The ranking (with respect to the maximum squared SRs) of the most recent market dislocations from highest to lowest is: (i) 2022 Russia-Ukraine War (46.62), (ii) 2008 Financial Crisis (46.33), (iii) 2000 Technology Bubble (39.62); and (iv) 2020 COVID-19 Stock Market Crash (30.14). This exhibit showcases the heightened level of stock market volatility over this more recent period.

The second finding from Figure 19 illustrates the notably different dynamics of the SDF based on the PC1-5 factor model.[40] While this SDF is also time-varying,it peaks during the Technology Bubble of 2000 (one-year variance estimate reaches a maximum value of 20.38 in April 2001) and reaches a minimum value of 0.27 (December 1985). Most notably, the Tech Bubble plays a disproportionate role in characterizing the SDF's variance relative to any other market dislocation. The variance estimate during the Tech Bubble is roughly twice as large as that experienced during the 2022 Russia-Ukraine War and nearly three times as large as that experienced during the 2008 Financial Crisis. In essence, the factor structure (and its properties) is in large part a product of the stock market dynamics during the Tech Bubble.

To help illustrate the dynamics of the entire factor structure over the sample period, Figure 20 displays a heatmap of the scree plot over time whereby PCA is estimated via a rolling window of the past 252 trading days. The warmer shades (yellow and orange) represent larger eigenvalues (stronger factor structure) whereas the cooler shades (blues and purples) represent smaller eigenvalues (weaker factor structure). This figure confirms the second finding of Figure 19. That is, the Tech Bubble is the brightest spot on this heat map. This evidence is consistent with Chinco et al. (2021); Jensen et al. (2023). In particular, Figure 2 of Jensen et al. (2023) shows the cumulative CAPM alpha of an average of this set of anomaly portfolios. The most salient feature of this plot is the significant outperformance

[40]The correlation with the SDF (based on test assets) is 0.75 leaving a significant amount unexplained by the reduced form factor model.

relative to the market around the Tech Bubble. Roughly one third of the gains over this period are due to this market dislocation. Moreover, Figure 1 of Chinco et al. (2021) displays the evolution of the prior variance over time (i.e. the ex-ante probability of witnessing a tradable anomaly dubbed the "anomaly base rate"). This anomaly base rate peaks during the Tech Bubble. This is entirely consistent with the evidence presented in this section showing the SDF's variance (based on the PC1-5 factor model) peaking during the Tech Bubble. Interestingly, the heat map reveals that all of the eigenvalues shift upward indicated by the warmer shades along the entire vertical axis. This happens to a lesser degree during the other market dislocations. All together, the factor structure is more *unstable* during market dislocations which coincides with the periods that experience the upward spikes in the variance of the SDF.

Lettau and Pelger (2020), who improve on PCA by adding regularization with respect to estimates of the test assets risk premia ("RP-PCA"), highlight the instability of the factor structure with respect to individual stocks. Moreover, they document the high generalized correlations (in excess of 0.9) of the leading PCA factors and conclude that the factor structure (based on portfolios) is relatively stable over time. However, the evidence presented in Figure 20 makes clear how the factor structure is highly time-varying (i.e. experiences periods of instability) based on this set of anomaly portfolios. Moreover, this new evidence is also consistent with Kozak et al. (2018) Equation 4 which decomposes the SDF's variance providing an important insight regarding how the SDF's Variance is a function of the strength of the factor structure (i.e. a weaker factor structure coupled with a large spread in average returns implies a higher Variance of the SDF).

This section began by highlighting the seemingly contradictory evidence between both a relatively weaker factor structure and lower SR in the second half of the sample compared to the first half ("OOS period" versus "IS period," respectively). We can reconcile this evidence in light of the discrepancy in the SDF's variance estimated via test assets versus PC factors in Figure 19. That is, the time-variation in the PC1-5 SDF Variance is significantly different in the first half (IS) of the sample period in comparison to the latter half (OOS) of the sample. In particular, the average ten-year rolling correlation between the two estimates of the SDF Variance is 0.22 in the IS period compared to 0.74 during the OOS period. This evidence coupled with the insight of Kozak et al. (2018) Equation 4 tells us that we should expect higher SRs when the cross-sectional variation in average returns does not coincide with the leading PCs. In spite of the higher degree of comovement, the proportion of the SDF's variance explained by the PC factor model is lower in the OOS period as evidenced by Table 22.

All together, this section has documented the time-varying dynamics of the SDF's variance (based on the test assets) which spikes upward during market dislocations. More importantly, it is now clear how the SDF's variance (based on the factor structure of the underlying anomaly portfolios) displays notably different dynamics whereby the Tech Bubble has a disproportionate impact on the SDF's variance. Hence, we have a more nuanced understanding as to what time period is driving the dynamics of the SDF's variance. With that, the next logical step is to reveal how the underlying L/S anomaly portfolios contribute to the SDF's variance both cross-sectionally and in the time-series.

3.3.3 On the Composition of the SDF's Variance

Figure 21 displays the percent contribution of each respective factor theme to the SDF's Variance. More specifically, each factor theme is calculated in the following way,

$$\text{\% Contribution to SDF's Variance} = 1 - \frac{\text{Max SR}^2_{\text{w/o factor}}}{\text{Max SR}^2_{\text{w/ all factors}}} \tag{23}$$

By leaving the factor theme out one at a time and re-estimating the variance of the SDF, Figure 21 captures the cross-sectional impact of each factor theme on the maximum squared SR. It turns out that the Value Factor Theme plays a disproportionate role in the maximum squared SR (22.9% contribution to the SDF's variance). Moreover, three out of the 13 factor themes (Value, Low Leverage and Seasonality) explain nearly half of the variance 47.5%. Perhaps surprisingly, Investment, Profitability and Size have almost no influence on the SDF's variance (i.e. a combined contribution of 0.2%). It is important to note that this analysis by construction does not take into account any interaction effects between the factor themes (the interactions account for 18.2% of the SDF's variance). All together, this evidence highlights how the SDF's Variance is in large part a product of only a handful of the factor themes.

After having established the cross-sectional impact of each of the factor themes on the SDF's variance, I now display how each of the factor theme's percent contribution to the SDF varies over time. Figure 22 displays the rolling one year variance of the difference between the SDF with all factors and the SDF with the respective factor removed. Given the outsized role of Value on the SDF's Variance, the Value Factor Theme's dynamics are highly similar to the SDF's variance based on the test assets with a similar ranking in peaks of the variance, with respect to market dislocations. In particular, value has the largest effect on the SDF's variance during the 2008 Financial Crisis. Most importantly, this figure provides a window into the underlying reason why the Tech Bubble has such a outsized role on the maximum squared SR. Six of the 13 factor themes (Debt Issuance, Low Leverage, Momentum, Profit Growth, Short-Term Reversal and Size) witness their peak contribution to the SDF's variance during the Tech Bubble. Of the remaining seven factor themes, the Tech Bubble is either second or third in relative peaks of its contribution to the SDF's variance. It is this commonality across factor themes during a turbulent period in the stock market that results in the Tech Bubble playing such a large role in the factor structure of the anomaly test assets.

The evidence presented in this section highlights the need for a new set (and perhaps approach) to test assets if we are going to adequately characterize the cross-sectional variation in expected stock returns. In particular, the current set of test assets is dominated by one period (Tech Bubble) and one factor (Value). The next section takes up this challenge to get us one step closer to spanning the SDF by proposing hedge fund portfolios as a new set of test assets.

3.4 Expanding the Scope of Test Assets

Identifying test assets that help explain the cross-section of expected stock returns that remain robust OOS is a challenging enterprise. One of the main ideas of this paper is that we can potentially get closer to spanning the SDF if we identify expected returns *indirectly*. That is, if we are going to catch fish (i.e. expected returns) we shouldn't look for fish but instead we should look for the birds. In our setting, those birds are hedge funds. This section conducts the same empirical analysis as Section 3.3 with respect to hedge fund (in comparison to the Jensen et al. (2023) anomaly) portfolios as test assets.

3.4.1 Underlying Motivation to Use Hedge Fund Portfolios as Test Assets

In principle, hedge funds can be viewed as machines that harvest return predictability from markets. Given that this is arguably the modus operandi of any hedge fund it seems only natural that we might be more successful in identifying the maximum squared SR (i.e. spanning the SDF) among a group of managed portfolios comprised of hedge funds. The main obstacle that we face is that we don't get to observe the complete universe of existing hedge funds. Instead, we rely on a subset of the hedge fund universe that is reported in hedge fund commercial databases.

It is strictly voluntary whether or not a hedge fund chooses to report to a commercial database. The primary motivation for hedge funds to report (i.e. market) their returns to these databases is to raise more capital. Hence, the hedge funds with the best and worst performance will likely not be reflected in these databases. In fact, Barth et al. (2023); Brown et al. (2024) note how most large hedge funds with assets greater than US $ 1 billion do not report to commercial databases (i.e. in excess of US$ 2 trillion of combined AUM). Most importantly, the unobserved right tail of the distribution of hedge funds has superior performance (i.e. higher alphas) to those reported in commercial databases. In light of this fact, I aim to proxy for the very best performing hedge funds by constructing an *upper bound* of hedge fund returns by combining the top three deciles each month of hedge fund returns reported in TASS. This group of hedge fund returns corresponds to average SRs ranging from 2.48 to 5.18 (see Table 1).

3.4.2 On the Robustness of Hedge Fund Portfolios

In general, there are several issues with using hedge fund portfolios as test assets in empirical asset pricing. First, unlike asset pricing factors, hedge funds cannot be shorted. Second, hedge fund returns often involve leverage. Third, hedge funds have limited capacity and as a group are too small relative to the broader stock market. Fourth, hedge fund skill is in short supply and as a result it is difficult to achieve meaninful separation in hedge fund returns both ex-ante and ex-post. The scope of interest for this study is primarily focused on documenting the dynamics of the SDF's variance. While these limitations almost surely affect the level of the SDF's variance (maximum squared SR), they should not meaningfully alter the dynamics of the SDF's variance identified in this section. The only issue that poses a potential challenge to uncovering the SDF's dynamics concerns the issue of separation among hedge fund managers with respect to skill, which I turn to next.

It has been well documented in the hedge fund literature that a minority of hedge fund managers possess skill. Chen et al. (2017) estimates that 9% of hedge funds have superior performance which is consistent with Giglio et al. (2021) who also note the difficulty in predicting hedge fund performance given the small minority of hedge fund managers who possess skill. With that, it rules out the natural choice to form hedge fund test assets by hedge fund strategy since there are too few hedge fund managers within a given strategy that are able to achieve superior returns (i.e. hedge fund strategies produce similar returns). By construction, the methodology I propose (hedge fund deciles by month) addresses the lack of separation problem. Ultimately, the top decile hedge fund returns are intended to capture the most successful group of unobserved hedge funds that do not report to commercial databases (consistent with the findings of Barth et al. (2023); Brown et al. (2024)).

In a recent study, Koijen et al. (2023) note how hedge funds are the most price-sensitive (i.e. elastic) among institutional investors. In particular, they estimate the impact of hedge funds on stocks to be $3.58 per dollar of wealth. In short, they conclude that hedge funds have the greatest impact on equity prices per dollar of wealth. This new evidence lends support to further investigating hedge funds as a potentially new set of test assets in spite of the limitations mentioned in the previous paragraph. Moreover, the scope of this study is in constructing an upper bound of the SDF. Hence, this study explicitly acknowldeges that the use of these new test assets is to construct an *upper bound*. Subsequent studies can later identify whether this upper bound is in fact the least upper bound (supremum) of the SDF implied by alternative sets of test assets.

Overall, I argue that hedge fund portfolios as test assets offer us a net benefit to achieving a more accurate representation of the investment opportunity set. Ultimately, the proof will be in the empirical results generated from using them as a new set of test assets.

3.4.3 Pricing Performance: Magnitude and Persistence

The first set of results reported in this section begins with Table 23. As in Section 3.3, this table follows the same analysis as Table 21 with the differences being: (i) the sample is now from January 1996 to June 2023 at a monthly frequency (to coincide with the available hedge fund data); and (ii) the maximum squared SR of hedge fund portfolios (in red) is reported in the bottom panel. The main results of the table are largely unchanged. That is, the upper panel again shows that the first five PCs are able to explain most of the cross-sectional variation in expected returns of the anomaly portfolios. In the bottom panel, the maximum squared SR of all anomalies rises considerably (from 9.31 to 24.52) thereby reflecting the shorter time window, sampled at a monthly frequency. That said, there remains a large distance between the maximum squared SR of the PC1-5 factor model and the test assets which is also confirmed with the 25 Size-B/M portfolios. The evidence to take note of in this table is that the maximum squared SR of hedge fund decile portfolios is larger (26.25) than any of the other underlying sets of test assets. Importantly, there is not too wide a gap betwen the Jensen et al. (2023) anomalies' maximum squared SR and that implied by hedge fund portfolios, which suggest that there is considerable overlap between the risk premia exposure of both sets of test assets. This initial piece of evidence is encouraging that hedge fund portfolios are worthwhile to investigate further as a new set of test assets.

Figure 23 conducts the same pseudo OOS analysis as in Figure 18 now with the 25 Size-B/M portfolios replaced by hedge fund decile portfolios. Given that the sample period now begins in January 1996, the OOS period has a start date of September 2009. As before, near arbitrage opportunities are *not* persistent over time with respect to the anomaly portfolios. More specifically, the OOS SRs are, on average, 43% as large as the IS SRs for the anomaly portfolios. In sharp contrast to the factor theme anomaly test assets, there is evidence of persistence in the near arbitrage opportunities of hedge fund portfolio test assets. In particular, the highest decile hedge funds (D10) is above the 45 degree line (i.e. an OOS SR of 5.43 compared to an IS SR of 5.27). This is evidence that near arbitrage opportunities are not as transient as previously considered (Kozak et al., 2018). It is important to note that while the remaining nine decile portfolios fall below the 45 degree line, this is to be expected given that there is a small minority of hedge fund managers with skill (i.e. less than 10% based on the estimate of Chen et al. (2017)).

As mentioned previously, the most important exercise involves the same pseudo OOS analysis with respect to the maximum squared SR (Table 22) in order to determine whether or not the lower ranked PCs (i.e. near arbitrage opportunities) matter to expected returns. Table 24 conducts this exercise with respect to anomaly versus hedge fund portfolios. As before the maximum squared SR of the anomalies is significantly lower in the OOS period (i.e. one fifth the size of the IS period). This is consisent with the results from Figure 23. All of this changes when we now look at the maximum squared SR with respect to hedge fund portfolios: the maximum squared SR more than doubles in the OOS period (48.98) compared to the IS period (22.46). Most importantly, hedge fund portfolios get us closer to spanning the SDF by generating a SR that is 3.4x higher than the traditional anomaly portfolios during the OOS period. In contrast to our existing understanding of the cross-section of expected returns, this piece of evidence confirms two new important facts: near arbitrage opportunities are both *larger* and more *persistent* than previously documented. With that, it becomes less clear that we can rule out near arbitrage opportunities in markets. Having established the cross-sectional results of the hedge fund portfolios, we now turn towards the time-series dynamics of this new set of test assets.

3.4.4 Variance Dynamics of the SDF's Upper Bound (Implied by Hedge Fund Portfolios)

Figure 24 displays the one-year rolling variance of the SDF based on the hedge fund decile portfolios (blue line) versus the upper bound (red line). Recall that the upper bound is simply a three-factor model comprised of hedge fund deciles eight to ten (D8-D10). Consistent with the SDF estimated via the anomaly portfolios, the variance is highly time-varying with upward spikes during market dislocations (e.g. the most recent episodes include the 2000 Technology Bubble, 2008 Financial Crisis, 2020 COVID-19 Stock Market Crash and 2022 Russia-Ukraine War). However, the noteworthy departure is that we now see significant factor structure dynamics across each of the most recent market dislocations in contrast to Figure 19 which exhibited a disproportionate impact of the Technology Bubble with respect to the SDF's Variance based on PC1-5 factor model. Recall that decile portfolios (especially the upper bound) are highly related to portfolios estimated via PCA. In other words, the decile portfolios exhibit a more balanced representation of different market dislocations.

Figure 25 better showcases the factor structure dynamics (of hedge fund portfolios) by displaying a heatmap of the scree plot over time estimated via a five-year rolling window with monthly observations. Recall, the warmer shades (yellow and orange) represent larger eigenvalues (stronger factor structure) whereas the cooler shades (blues and purples) represent smaller eigenvalues (weaker factor structure). The key takeaway from this figure is that the factor structure of hedge fund test assets shows that each of the most recent market dislocations has a significant impact on the factor structure in contrast to the previous heatmap which showed the Tech Bubble dominated the factor structure. This is what we should expect of test assets that span the SDF. That is, we should expect risk premia (including near arbitrage opportunities) to be large during periods of market turbulence. The evidence on factor structure highlights the third new fact regarding expected returns: near arbitrage opportunities occur from a more diverse set of periods characterized by market turbulence. That is, the factor structure is no longer the product of just the Tech Bubble.

3.4.5 Decomposing the Upper Bound of Hedge Fund Returns

After having presented both the cross-sectional and time-series results with respect to hedge fund test assets, we are left asking what exactly do these portfolios represent. Table 25 reports the results of univariate time series regressions of the hedge fund upper bound on the Jensen et al. (2023) factor theme anomalies (Panel A) and hedge fund strategies (Panel B). This table documents two interesting findings. First, the hedge fund upper bound has significant loadings across a number of factor theme anomaly portfolios. The only anomalies which the upper bound does not load on are: Momentum, Profit Growth and Short Term Reversal. The Low Risk Anomaly has the highest R^2 of 11%. When we include all 13 factor themes, the R^2 rises to 16%. Importantly, the hedge fund upper bound has highly significant alphas across all of the factor theme anomaly portfolios which is unchanged when we include them all in a multivariate regression.

Second, in Panel B of Table 25 we see that again the upper bound loads significantly on a wide array of hedge fund strategies. In particular only three of the 11 strategies are not significant: Fixed Income Arbitrage, Managed Futures and the Options Strategy. The L/S Equity Hedge has the highest R^2 of 14%, which is reassuring given that we are interested in spanning the cross-section of expected stock returns. When we include all 11 hedge fund strategies, the R^2 rises to 25%. Importantly, the hedge fund upper bound has highly significant alphas across all of the hedge fund strategies which is left unchanged when we include them all in a multivariate regression. All together, this evidence confirms that the hedge fund upper bound portfolio returns is comprised of a wide array of exposures with respect to factor theme anomalies and hedge fund strategies. In spite of these exposures, the vast majority of the upper bound's return variation is left unexplained. This is a necessary feature of a set of test assets that arguably gets us closer to spanning the SDF.

3.5 Conclusion

I document that a large and comprehensive set of test assets does not span the SDF. In particular, I show that the factor structure of the test assets is dominated in the time-series by the Tech Bubble and is dominated in the cross-section by only three (out of 13) factor themes: Value, Low Leverage and Seasonality.

After having established that the existing set of anomalies fails to span the SDF, I offer an alternative set of test assets: hedge fund portfolios. By adopting hedge fund portfolios as test assets, I show that near arbitrage opportunities are larger, more persistent and are the product of a wider array of periods characterized by market turbulence. All together, this new evidence forces us to revist our current prior regarding the absence of near arbitrage opportunities. While this asset pricing prior is likely satisfied on an unconditional basis, it is far less clear whether it is not violated during market dislocations when the factor structure is highly unstable. Updating our economic prior to reflect the dynamics of the factor structure will undoubtedly lead to better OOS prediction in the important and emerging field of economic regularization.

Table 20: Descriptive Statistics of Anomaly and Hedge Fund Portfolios
The table presents the summary statistics for the factor theme anomaly portfolios (Panel A) and hedge fund decile portfolios (Panel B). The sample periods reflect the full sample that is available for each respective dataset. Anomaly (Hedge Fund) portfolio return statistics are annualized from their daily (monthly) returns. The mean and volatility of returns are expressed in percent.

Panel A: Anomaly Portfolios (Jan. 1952 to Dec. 2023)

	Mean Return	Volatility	Sharpe Ratio	Skewness
Accruals	2.90	3.10	0.93	−0.08
Debt Issuance	2.46	2.07	1.19	−0.07
Investment	3.00	5.97	0.50	−0.34
Low Leverage	0.25	7.79	0.03	−0.05
Low Risk	1.08	10.98	0.10	−0.15
Momentum	4.30	8.57	0.50	−1.44
Profit Growth	2.04	3.10	0.66	−1.07
Profitability	2.17	5.43	0.40	0.56
Quality	3.24	4.56	0.71	−0.86
Seasonality	1.68	1.72	0.98	−0.28
Short Term Reversal	2.05	3.45	0.59	0.95
Size	2.18	5.95	0.37	−0.10
Value	3.67	8.49	0.43	0.64

Panel B: Hedge Fund Decile Portfolios (Jan. 1996 to Jun. 2023)

	Mean Return	Volatility	Sharpe Ratio	Skewness
D1	−92.90	23.75	−3.91	−3.18
D2	−36.63	16.42	−2.23	−2.77
D3	−19.51	13.60	−1.43	−2.46
D4	−8.53	12.12	−0.70	−2.01
D5	0.93	11.43	0.08	−1.31
D6	9.96	11.37	0.88	−0.48
D7	19.84	11.78	1.68	0.31
D8	31.51	12.72	2.48	1.01
D9	49.24	14.94	3.30	1.55
D10	106.96	20.63	5.18	2.01

Table 21: Revisiting Anomaly Portfolios with Principal Component Factors
This table revisits the analysis conducted in Table II of Kozak et al. (2018) with the following modifications. The long-short strategy daily returns have been replaced by a larger and more comprehensive set of long-short strategy daily returns from Jensen et al. (2023) that span 13 themes comprised of 153 factors. In addition, I have added the Daniel et al. (2020) ("DMRS") factors, indicated by an asterisk (*), that hedge out unpriced risk. All table entries have been annualized. The sample period for all of the analysis, except for the DMRS factors, is January 1952 to December 2023. The DMRS factors uses data from July 1963 to March 2023.

		PC Factor-Model Alphas				
	Mean Return	PC1	PC1-2	PC1-3	PC1-4	PC1-5
Accruals	2.90	2.88	2.97	2.65	2.66	3.60
Debt Issuance	2.46	2.51	2.43	2.27	2.25	1.90
Investment	3.00	1.83	2.10	1.62	1.50	1.66
Low Leverage	0.25	1.95	1.73	1.61	1.80	1.94
Low Risk	1.08	−1.23	−2.14	−2.35	−2.03	−0.93
Momentum	4.30	4.24	1.73	0.94	0.76	0.11
Profit Growth	2.04	2.31	1.73	1.84	1.84	1.35
Profitability	2.17	1.50	0.89	1.92	1.90	0.28
Quality	3.24	3.39	2.49	3.04	3.21	1.37
Seasonality	1.68	1.57	1.46	1.36	1.38	1.18
Short Term Reversal	2.05	1.91	2.05	2.17	2.26	2.96
Size	2.18	2.11	3.16	2.13	2.29	−0.51
Value	3.67	1.88	2.78	2.99	2.84	1.90

		PC Factors' Max. Squared SR				
	Max SR^2	PC1	PC1-2	PC1-3	PC1-4	PC1-5
All Anomalies	9.31	0.07	0.22	0.32	0.33	1.76
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
For Comparison:						
25 Size-B/M	3.83	0.89	1.00	1.22	1.89	2.07
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
MKT, SMB and HML	0.71	n/a	n/a	n/a	n/a	n/a
MKT*, SMB*, HML*	1.99	n/a	n/a	n/a	n/a	n/a

Table 22: Anomaly Portfolios with Principal Component Factors (IS versus OOS)
This table repeats the empirical work in Table 21 except now for the two equal halves as in Figure 18 whereby the first half (upper panel) represents the IS period (January 1952 to December 1987) and the second half (bottom panel) represents the OOS period (January 1988 to December 2023). All table entries have been annualized.

		PC Factors' Max. Squared SR (IS)				
	Max SR^2	PC1	PC1-2	PC1-3	PC1-4	PC1-5
All Anomalies	20.77	0.25	1.73	2.19	2.24	7.58
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
25 Size-B/M	8.13	1.59	1.59	4.65	4.70	4.70
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

		PC Factors' Max. Squared SR (OOS)				
	Max SR^2	PC1	PC1-2	PC1-3	PC1-4	PC1-5
All Anomalies	6.69	0.04	0.10	0.11	0.12	1.61
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
25 Size-B/M	4.76	0.60	0.67	0.67	1.03	1.14
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Table 23: Anomaly Portfolios with PC Factors (Jan. 1996 to Jun. 2023)
This table repeats the empirical work in Table 21 except for a shorter sample period from January 1996 to June 2023. The shorter time series is used here to accomodate the more limited hedge fund data. As previously, the long-short strategy daily returns are from Jensen et al. (2023). The Daniel et al. (2020) ("DMRS") factors are indicated by an asterisk (*). All table entries have been annualized.

		PC Factor-Model Alphas				
	Mean Return	PC1	PC1-2	PC1-3	PC1-4	PC1-5
Accruals	1.15	1.34	1.42	1.38	1.38	2.15
Debt Issuance	2.15	2.23	2.16	2.12	2.11	1.40
Investment	2.99	1.51	1.84	1.70	1.74	1.50
Low Leverage	0.07	2.42	2.06	2.09	1.87	1.91
Low Risk	1.66	−1.15	−2.13	−2.11	−2.54	−0.25
Momentum	4.12	3.87	1.34	1.17	1.35	0.09
Profit Growth	0.57	0.87	0.30	0.35	0.36	−0.38
Profitability	3.59	2.40	2.00	2.18	2.25	0.09
Quality	3.74	3.78	2.91	3.05	2.80	0.79
Seasonality	1.42	1.27	1.14	1.12	1.10	1.03
Short Term Reversal	1.43	1.31	1.40	1.43	1.40	2.23
Size	1.44	1.54	2.49	2.33	1.88	−0.73
Value	3.84	1.48	2.53	2.53	2.72	1.18

		PC Factors' Max. Squared SR				
	Max SR^2	PC1	PC1-2	PC1-3	PC1-4	PC1-5
All Anomalies	24.52	0.27	0.58	0.59	0.68	6.81
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
For Comparison:						
25 Size-B/M	19.70	1.99	2.08	2.09	3.54	4.15
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
MKT, SMB and HML	1.46	n/a	n/a	n/a	n/a	n/a
MKT*, SMB*, HML*, RMW* and CMA*	8.23	n/a	n/a	n/a	n/a	n/a
Hedge Fund Decile Portfolios	26.25	n/a	n/a	n/a	n/a	n/a

Table 24: HF and Anomaly Portfolios with Principal Component Factors (IS versus OOS) This table repeats the empirical work in Table 22 except now for the hedge fund sample period whereby the first half (upper panel) represents the IS period (January 1996 to September 2009) and the second half (bottom panel) represents the OOS period (October 2009 to June 2023). All table entries have been annualized.

		PC Factors' Max. Squared SR (IS)				
	Max SR^2	PC1	PC1-2	PC1-3	PC1-4	PC1-5
All Anomalies	22.28	0.08	0.14	0.21	2.78	3.79
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Hedge Fund Decile Portfolios	22.46	n/a	n/a	n/a	n/a	n/a

		PC Factors' Max. Squared SR (OOS)				
	Max SR^2	PC1	PC1-2	PC1-3	PC1-4	PC1-5
All Anomalies	4.22	0.10	0.14	0.16	0.71	0.72
χ^2 p-value for zero pricing errors		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Hedge Fund Decile Portfolios	48.98	n/a	n/a	n/a	n/a	n/a

Table 25: Decomposition of Hedge Fund Upper Bound Portfolio Returns
The table presents univariate regressions of the hedge fund upper bound portfolio returns (equal weighted across D8-D10) spanned on factor theme anomaly portfolios (Panel A) and hedge fund strategy returns (Panel B). Coefficient estimates (in decimal form) of the intercept and coefficient are reported along with their respective Newey and West (1987) t-statistics and R^2. The sample period is monthly returns from January 1996 to June 2023. Asterisks denote the levels of statistical significance of the coefficient: 10% level (*), 5% level (**) and 1% level (***).

Panel A: Anomaly Portfolios

	Intercept	t-stat	Coefficient	t-stat	R^2
Accruals**	0.155	14.21	1.743	2.27	0.03
Debt Issuance**	0.152	15.01	2.521	1.96	0.04
Investment*	0.158	13.09	−0.605	−1.79	0.02
Low Leverage***	0.157	14.10	0.748	2.58	0.08
Low Risk***	0.157	14.59	−0.783	−3.16	0.11
Momentum	0.157	13.75	−0.203	−0.57	0.01
Profit Growth	0.156	13.47	0.026	0.03	0.00
Profitability***	0.159	13.58	−1.040	−3.17	0.07
Quality*	0.158	13.27	−0.691	−1.66	0.01
Seasonality*	0.158	13.36	−1.389	−1.81	0.01
Short Term Reversal	0.157	13.11	−0.798	−1.44	0.01
Size***	0.156	13.68	0.951	2.96	0.04
Value**	0.158	13.55	−0.578	−2.03	0.05
All Anomalies	0.151	15.52	n/a	n/a	0.16

Panel B: Hedge Fund Strategy Returns

	Intercept	t-stat	Coefficient	t-stat	R^2
Convertible Arbitrage***	0.154	13.66	0.674	4.94	0.06
Dedicated Short Bias***	0.160	11.70	−0.271	−3.96	0.06
Emerging Markets***	0.154	13.82	0.401	3.43	0.08
Equity Market Neutral**	0.153	13.84	0.876	2.37	0.03
Event Driven***	0.152	14.04	0.727	2.97	0.06
Fixed Income Arbitrage	0.155	12.92	0.448	1.03	0.01
Global Macro**	0.155	13.14	0.396	2.21	0.02
L/S Equity Hedge***	0.152	15.75	0.771	4.90	0.14
Managed Futures	0.134	12.96	0.015	0.17	0.00
Multi Strategy***	0.153	14.24	0.770	2.64	0.06
Options Strategy	0.135	17.46	−0.020	−0.52	0.00
All Strategies	0.120	11.13	n/a	n/a	0.25

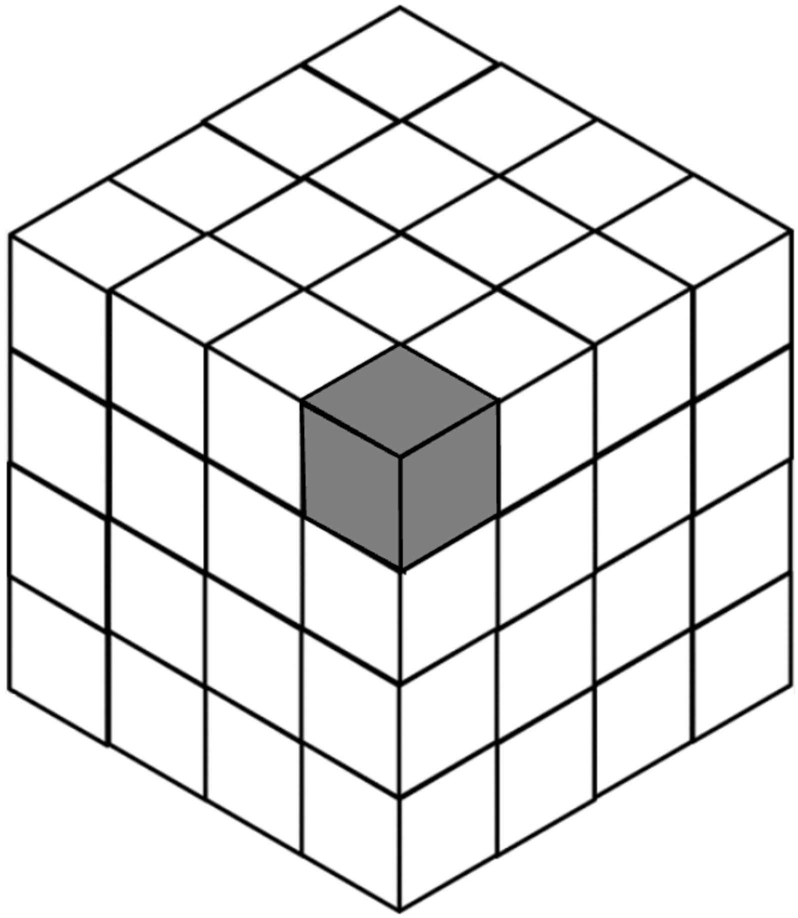
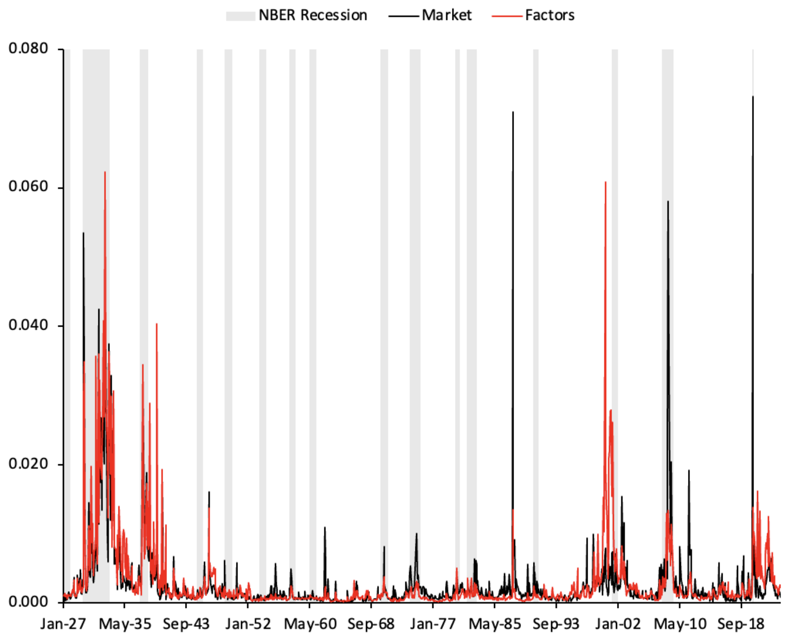


Figure 15: The Space of Relevant Test Assets
The figure displays a cube partitioned into smaller cubes to represent the set of relevant test assets required to span the SDF. Each cube represents a test asset that possesses a unique source of return variation required to span the SDF. The grey shaded cube represents the extant set of anomaly portfolios used to span the SDF.

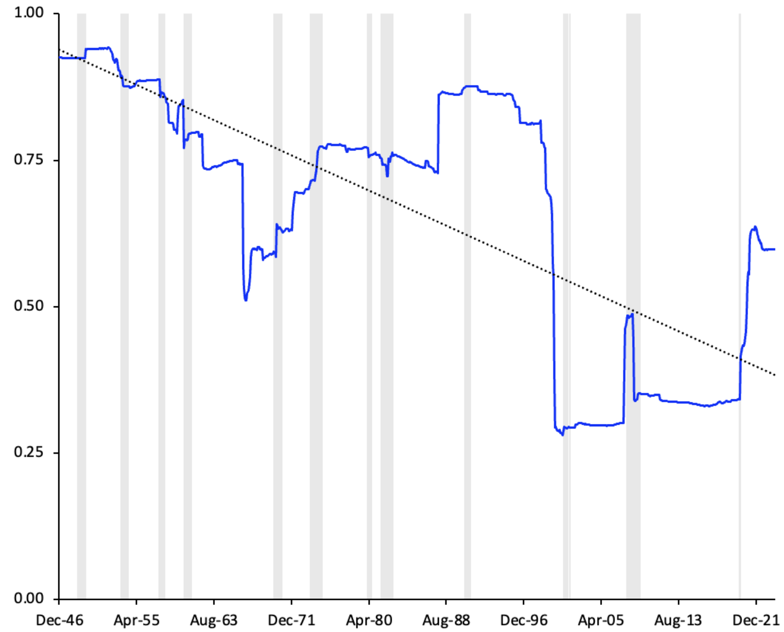
(a) Sharpe Ratio (SR)

(b) Information Ratio (IR)

Figure 16: Anomaly Portfolios (versus the Market) over Time
The figure displays the time-varying dynamics of Jensen et al. (2023) anomaly portfolios compared to the CRSP value-weighted index. Panel A displays the ten-year rolling Sharpe ratio (SR) of an equal-weighted combination portfolio of factor themes. Panel B shows the ten-year rolling information ratio (IR) of daily CAPM alphas (with Newey and West (1987) standard errors) of an equal-weighted combination portfolio of factor themes versus the value-weighted index excess return. The sample period is from January 1952 to December 2023 to ensure that all 13 factor themes are represented in the combination portfolio.

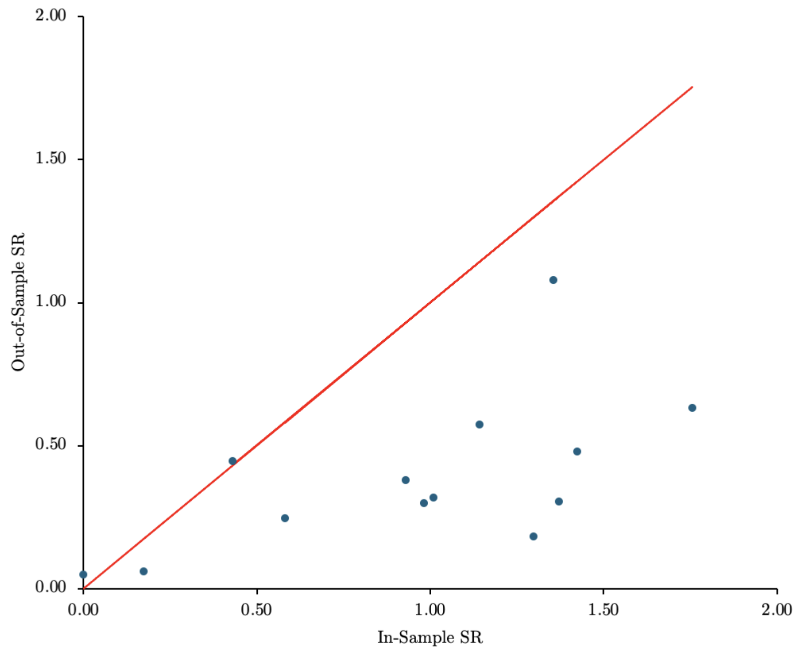


(a) Realized Volatility

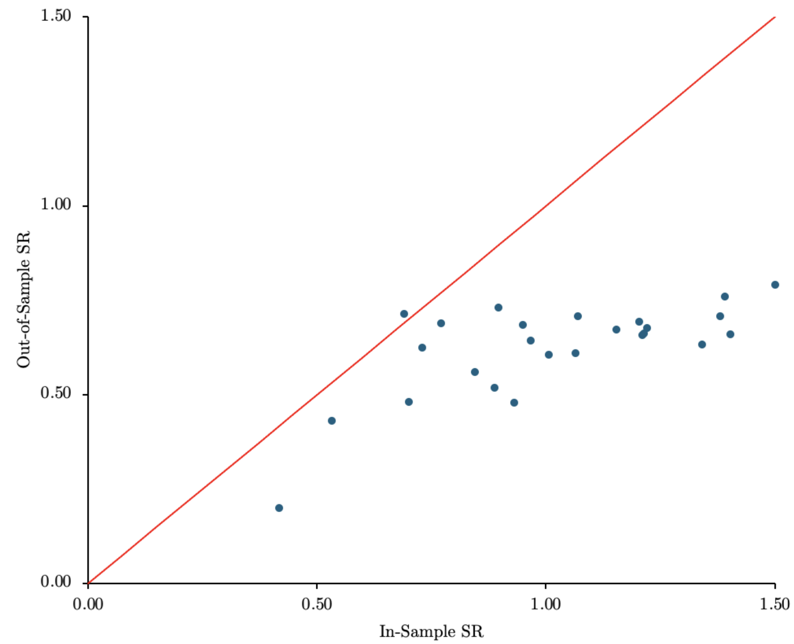


(b) Rolling 20-Year Correlation

Figure 17: Realized Volatility of Aggregate Market vs. Anomalies
Panel A displays the monthly realized volatility (based on daily returns) of the aggregate value-weighted market index (black line) versus the monthly realized volatility of an equal-weighted combination portfolio of the realized volatility from the Jensen et al. (2023) 13 Anomaly Portfolios (red line). Panel B displays the rolling 20-year correlation between the two realized volatilities along with the respective trendline. The combination portfolio is constructed in the following way. First, the anomaly returns are scaled to have 1/13th the annualized volatility of the aggregate market index. Second, the scaled anomaly returns are each squared and then summed into a combination portfolio at a daily frequency. The last step computes the monthly realized volatility based on the combination of the scaled daily squared anomaly excess returns. NBER recessions are indicated by the grey shaded regions. The sample period is from January 1927 to December 2023.



(a) Anomaly Portfolios



(b) 25 Size-B/M Portfolios

Figure 18: IS and OOS Sharpe Ratios (Anomaly vs. Size-B/M Portfolios)
The figure applies the pseudo OOS analysis from Figure 4 of Kozak et al. (2018). That is, I divide the full sample into two equal halves and treat the first half as the IS period whereas the second half of the sample is the OOS period with a start date of December 31, 1987. Panel A performs this analysis on the 13 factor themes from Jensen et al. (2023) whereas Panel B uses the 25 Size-B/M portfolios. The full sample period uses daily data from January 1952 to December 2023.

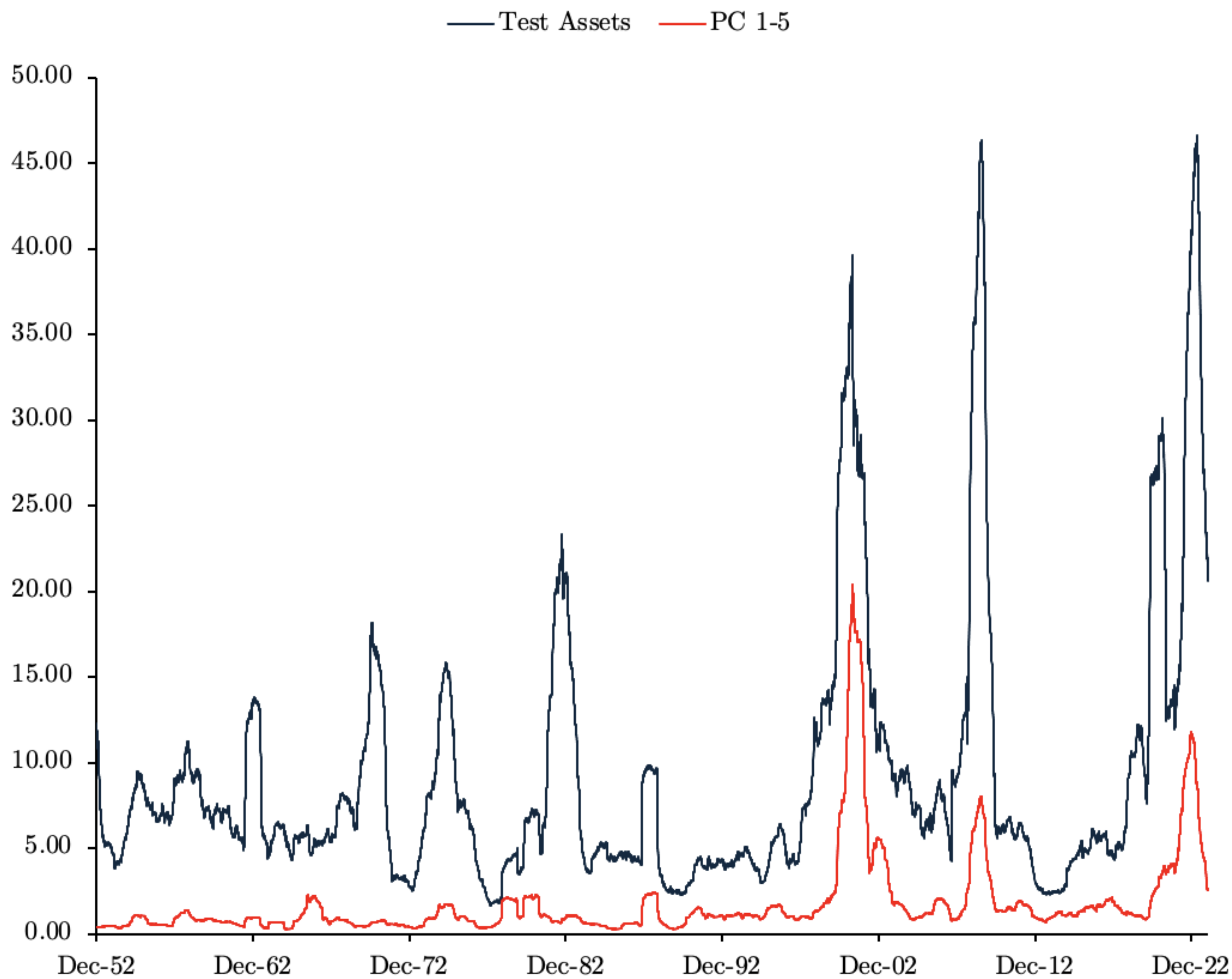


Figure 19: SDF Variance Over Time (Anomaly versus PC Portfolios)
The figure displays the one-year rolling variance estimates of the SDF based on the test assets (dark blue line) and the five leading PCs (PC 1-5) extracted from the test assets (red line). The test assets correspond to the 13 factor theme portfolios from Jensen et al. (2023). The sample period uses daily data from January 1952 until December 2023.

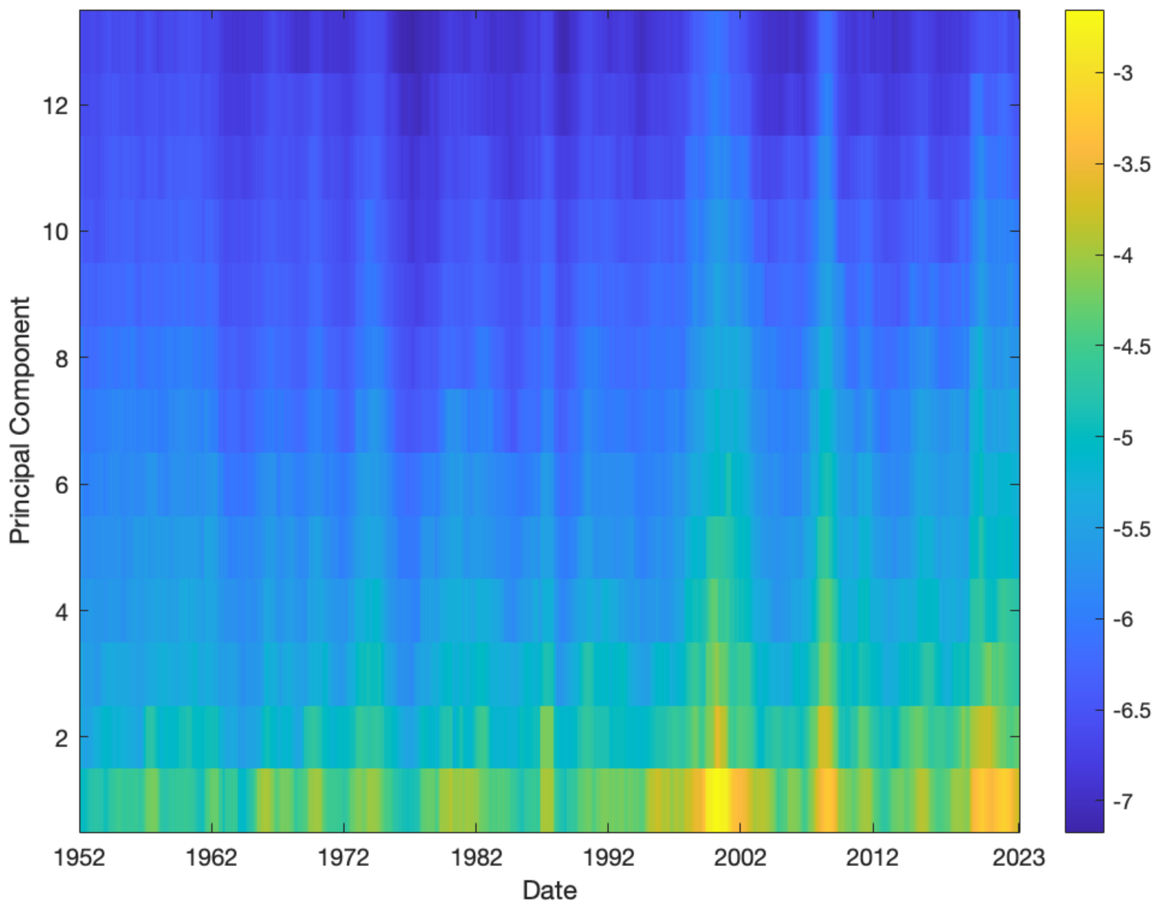


Figure 20: Factor Structure Over Time (Anomaly Portfolios)
The figure displays the time-varying factor structure of the anomaly portfolios. The colorbar in the two-dimensional view of the scree plot over time represents the magnitude of the eigenvalues on a logarithmic scale. The warmer shades (yellow and orange) represent larger eigenvalues (i.e. stronger factor structure). The cooler shades (blue and purple) represent smaller eigenvalues (i.e. weaker factor structure). Higher values of the vertical axis should correspond to decreasing importance of the higher order PCs (warmer to cooler shades). The factor structure is estimated via rolling PCA over the prior year. The sample period uses daily data from January 1952 until December 2023.

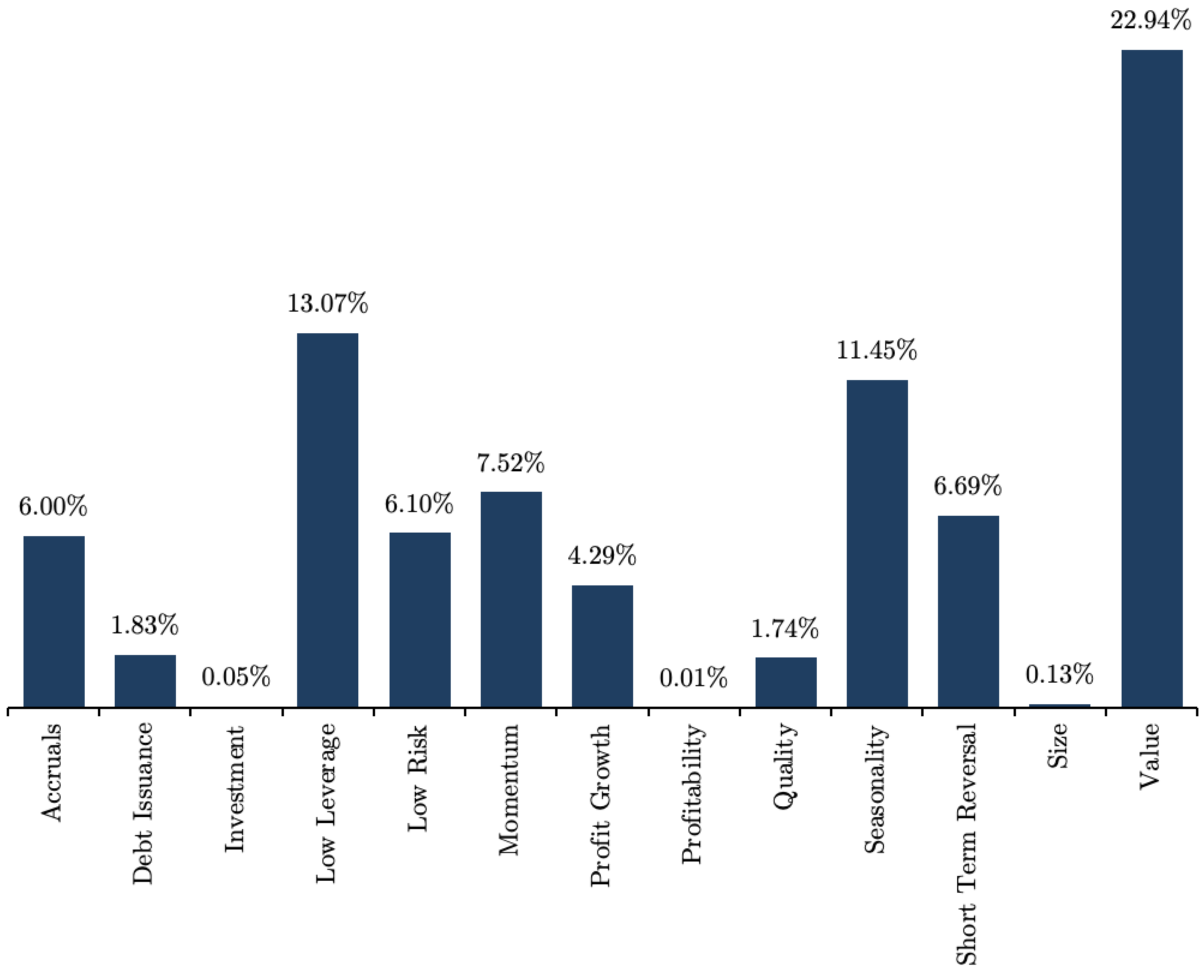


Figure 21: Factor Theme (%) of SDF Variance (Anomaly Portfolios)
The figure displays the percentage of each respective factor theme's contribution of the total SDF variance (maximum SR). That is, each factor theme percentage is calculated as $1 - \frac{\text{Max SR}^2_{\text{w/o factor}}}{\text{Max SR}^2_{\text{w/ all factors}}}$. The sample period uses daily data from January 1952 until December 2023.

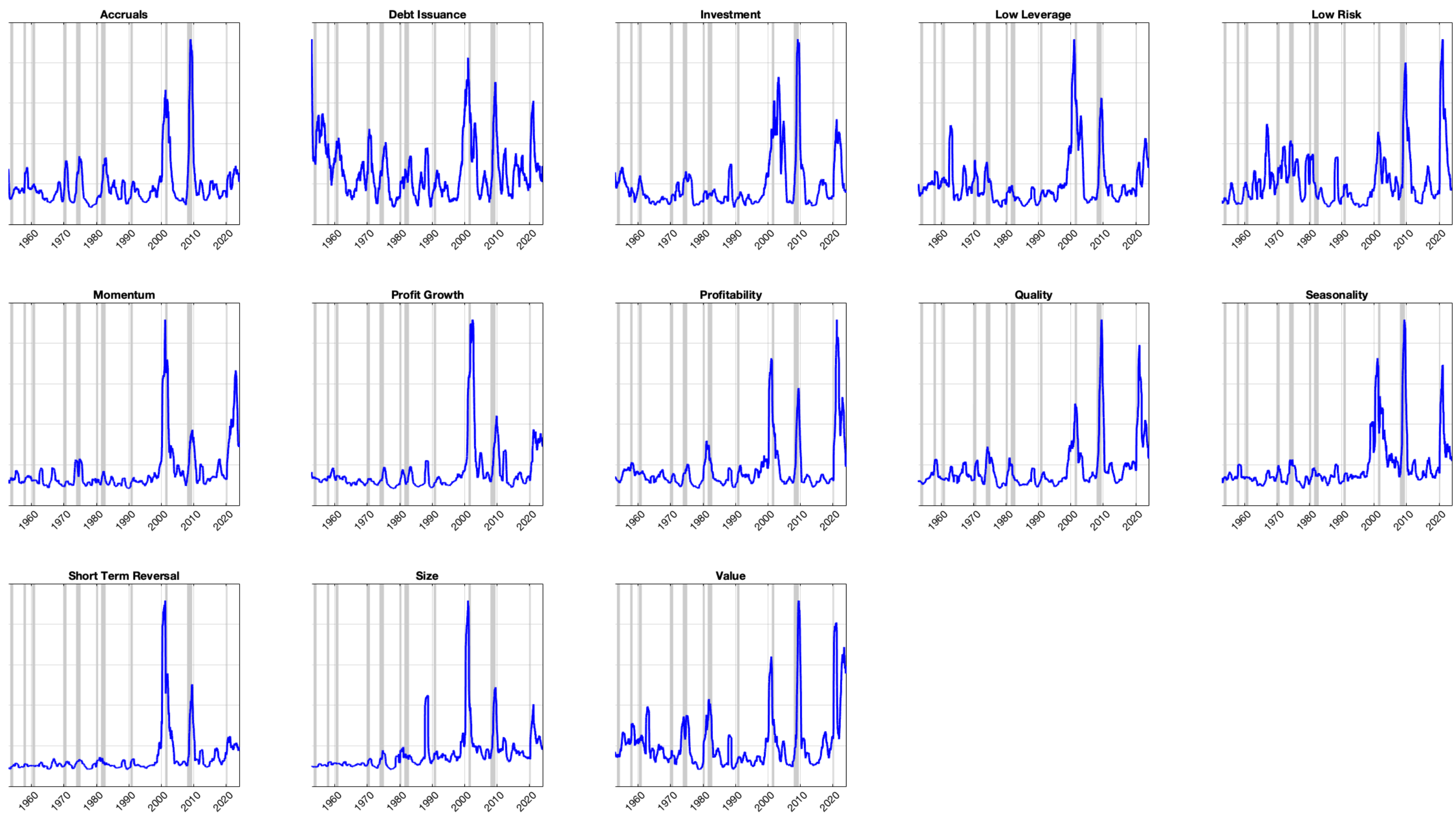
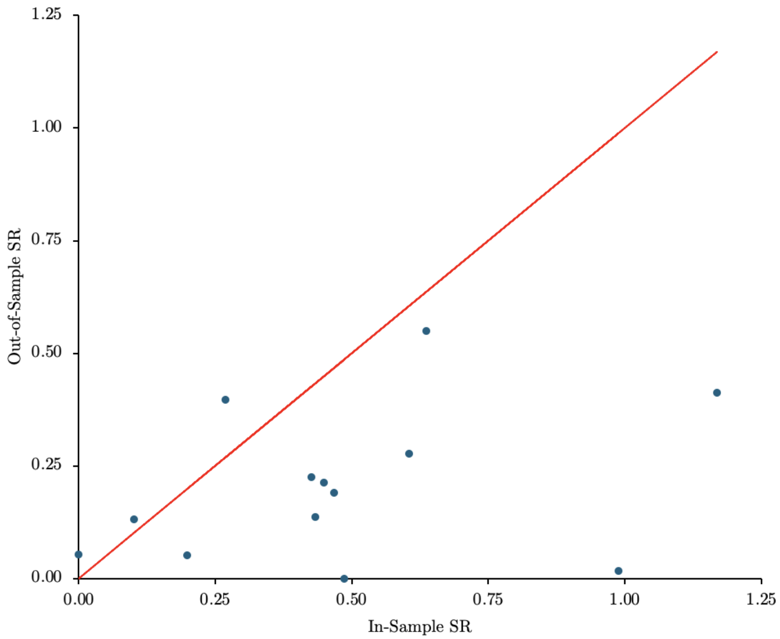
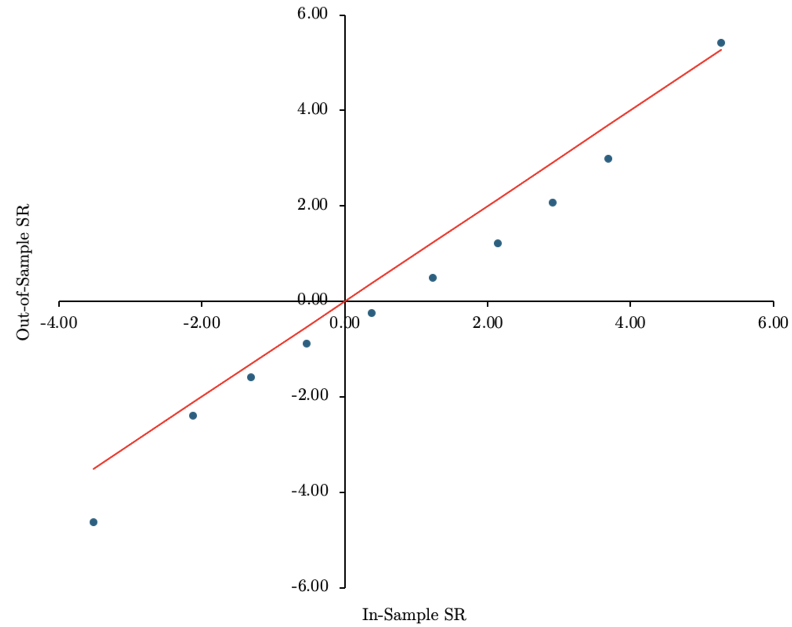


Figure 22: Factor Theme Contribution to SDF Variance Over Time (Anomaly Portfolios)
The figure displays the rolling one-year variance of the difference between the SDF with all factors and the SDF with the respective factor removed. Each respective factor theme is displayed as its own subfigure. NBER shaded recessions are the grey shaded bars. The sample period uses daily data from January 1952 until December 2023.

Business)



(a) Anomaly Portfolios



(b) Hedge Fund Portfolios

Figure 23: IS and OOS Sharpe Ratios (Anomaly vs. Hedge Fund Portfolios)
The figure applies the pseudo OOS analysis as Figure 18 over a smaller sample period to account for the more limited time series of hedge fund data. The OOS period begins September 2009. Panel A performs this analysis on the 13 factor themes from Jensen et al. (2023) whereas Panel B now uses the ten hedge fund decile portfolios. The full sample period uses monthly data from January 1996 to June 2023.

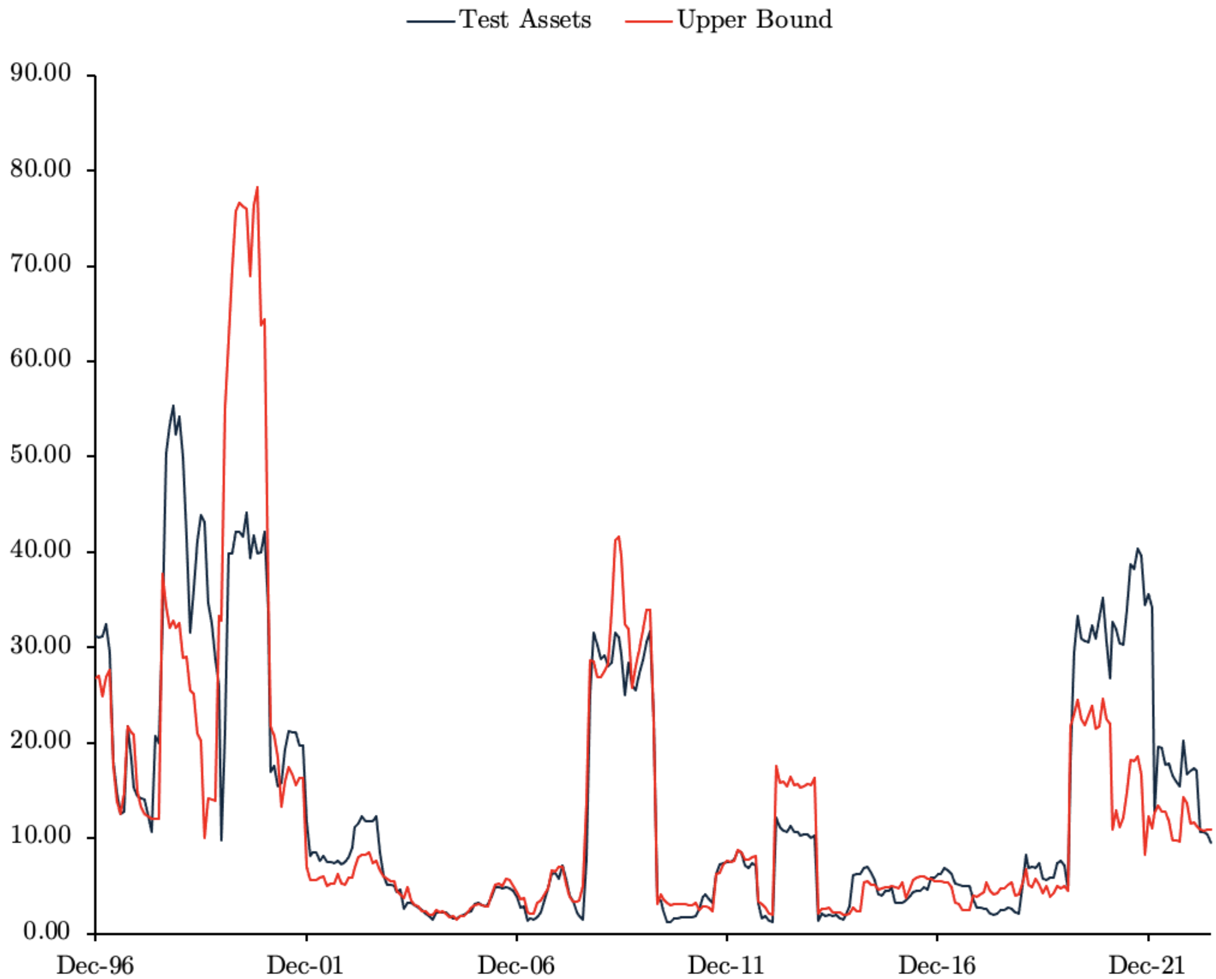


Figure 24: SDF Variance Over Time (Hedge Fund Deciles versus Upper Bound Portfolios) The figure displays the one-year rolling variance estimates of the SDF based on the test assets (dark blue line) and the upper bound (Deciles 8-10) extracted from the test assets (red line). The test assets correspond to the ten hedge fund decile portfolios. The sample period uses monthly data from January 1996 until June 2023.

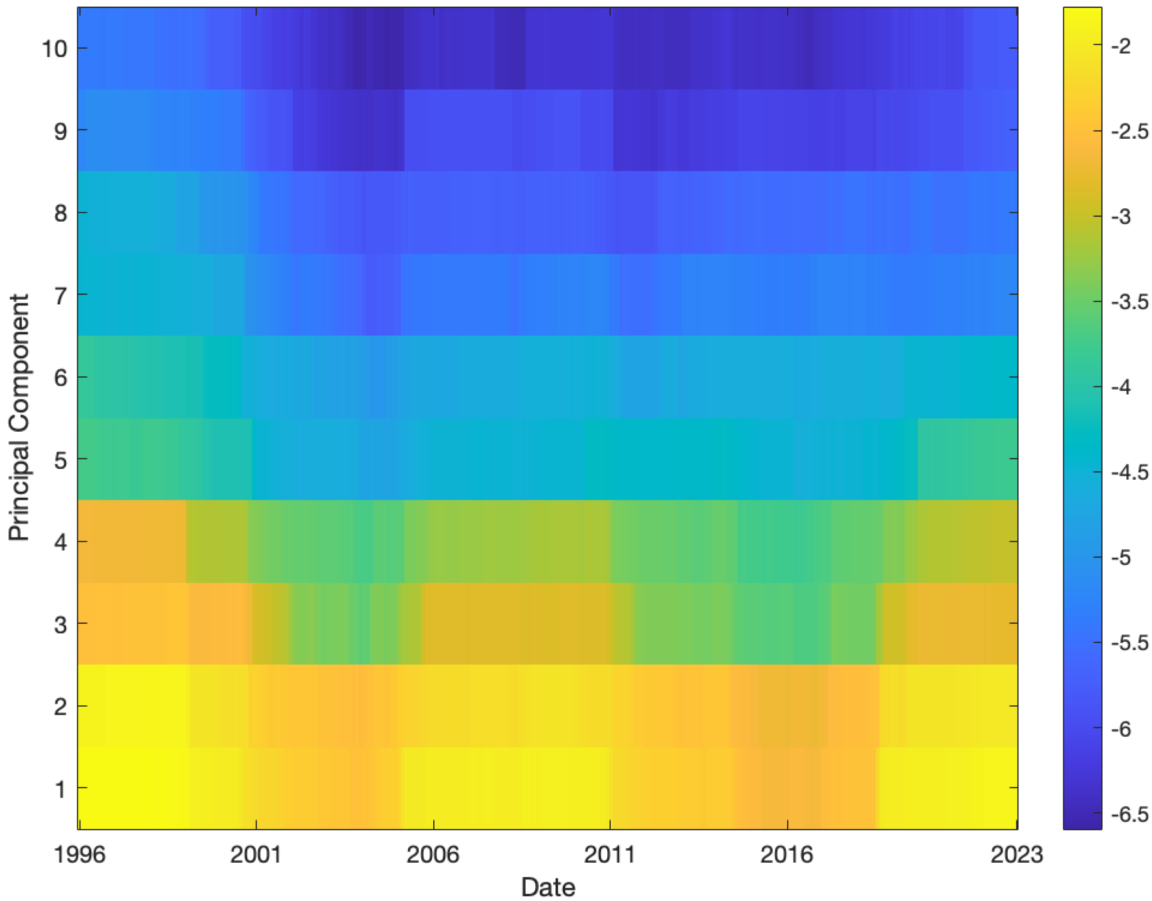


Figure 25: Factor Structure Over Time (Hedge Fund Portfolios)
The figure displays the time-varying factor structure of the hedge fund portfolios. The colorbar in the two-dimensional view of the scree plot over time represents the magnitude of the eigenvalues on a logarithmic scale. The warmer shades (yellow and orange) represent larger eigenvalues (i.e. stronger factor structure). The cooler shades (blue and purple) represent smaller eigenvalues (i.e. weaker factor structure). Higher values of the vertical axis should correspond to decreasing importance of the higher order PCs (warmer to cooler shades). The factor structure is estimated via rolling PCA over the last five years. The sample period uses monthly data from January 1996 until June 2023.

Appendices

Data Cleaning Methodology

I collect panel data, at a monthly frequency, from the TASS data from May 2004 to December 2022. This represents 22,269 unique hedge funds of which 3,790 survived (18,479 died) over the sample period, which in total is a sample size of 1,598,051 hedge fund month observations. The primary categories of the hedge fund strategies are: (i) Convertible Arbitrage, (ii) Dedicated Short Bias, (iii) Emerging Markets, (iv) Equity Market Neutral, (v) Event-Driven, (vi) Fixed Income Arbitrage, (vii) Fund of Funds, (viii) Global Macro, (ix) Long-Short Equity Hedge, (x) Managed Futures, (xi) Multi-Strategy, (xii) Options Strategy, (xiii) Other and (xiv) Undefined. No dedicated Short Bias hedge fund survives over the sample period. In addition, nearly two thirds (64%) of all hedge funds are represented by three categories: Fund of Funds (31.5%), Long-Short Equity (20.9%) and Multi-Strategy (11.5%). This over-representation declines to 55% of those that survive. Table A1 provides the unique hedge fund counts for the cleaned database by primary category for all, survived and dead hedge funds. The strategy composition of the cleaned database is consistent with the original database (i.e. top three categories represent 73.1% vs. 64%). To provide more colour, Figure A1 displays the total number of hedge funds (and the percentage that survive) over the sample period for the cleaned database. The number of hedge funds is hump-shaped with its peak at 2,425 unique funds in April 2011 with a steady decline afterwards. The number of survived hedge funds over the entire sample period represents 7% of the total number of reporting hedge funds, thereby highlighting the significant issue of attrition in hedge fund commercial databases.

I perform several screens on the data that are standard in the literature. I first exclude all observations with missing returns or AUM. I restrict the sample to those funds that have at least US$5 mm in AUM at some point during the sample period and report net-of-fee returns with a minimum of 36 consecutive monthly observations consistent with Couts et al.

(2020). The AUM restriction mitigates the impact of small funds. All funds with AUM whose base currency is foreign-denominated is converted to USD using end-of-month spot rates from Bloomberg. 51% of my sample consists of USD denominated hedge funds. Also consistent with Couts et al. (2020), I exclude hedge funds whose primary strategy is classified as "Undefined" or "Other." I remove backfill bias by adopting the method proposed by Jorion and Schwarz (2019) to identify backfilled observations. Survivorship bias is addressed by using both "dead" and "alive" hedge funds included in the TASS database. Lastly, I address the smoothing bias present in hedge fund returns by implementing the recent 3-step methodology proposed by Couts et al. (2020) which builds off of the 1-step methodology introduced by Getmansky et al. (2004). The 3-step methodology removes autocorrelation in returns at the fund- and strategy-level. I follow the convention in the hedge fund literature by estimating an MA(2) process which is used to unsmooth reported hedge fund returns.

Tables

Table A1: Hedge Fund Counts
The table presents the unique number of hedge funds over the sample period from May 2004 to December 2022 from the cleaned TASS Database. The data is reported by primary category for all, survived and dead hedge funds.

Hedge Fund Category	All (#)	All (%)	Survived (#)	Survived (%)	Dead (#)	Dead (%)
Convertible Arbitrage	72	1.77	6	0.74	66	2.02
Dedicated Short Bias	10	0.25	0	0.00	10	0.31
Emerging Markets	257	6.30	47	5.80	210	6.43
Equity Market Neutral	180	4.41	25	3.08	155	4.74
Event Driven	247	6.06	55	6.78	192	5.88
Fixed Income Arb	109	2.67	17	2.10	92	2.82
Fund of Funds	1,500	36.78	236	29.10	1,264	38.69
Global Macro	186	4.56	38	4.69	148	4.53
L/S Equity Hedge	1,013	24.84	247	30.46	766	23.45
Managed Futures	10	0.25	6	0.74	4	0.12
Multi Strat.	468	11.48	130	16.03	338	10.35
Options Strategy	26	0.64	4	0.49	22	0.67
All Hedge Funds	4,078	100.00	811	100.00	3,267	100.00

Table A2: Asset Pricing Factor Model Alphas of Hedge Fund Returns and Hedge Fund Replicators
The table presents the alphas with their Newey and West (1987) t-statistics (in parentheses) from monthly time series regressions of hedge fund returns and their substitutes (HFRI 500 Index, TASS equal-weighted portfolio, Fung and Hsieh (2001) Short Straddle (PTFSSTK), OTM Put-Writing Strategy and Short VIX) on four models: (i) the CAPM (market factor from Fama and French (1992)), (ii) FF3 (Fama and French, 1992), (iii) FF5 (Fama and French, 2015), and (iv) FF5 (Fama and French, 2015) + MOM (Carhart, 1997). The alphas are expressed in decimal form. The sample period is from January 2005 to December 2022.

	HFRI 500	TASS	Short Straddle	Short Put	**Short VIX**
CAPM	0.001	−0.002	0.029	0.003	**0.021**
	(1.08)	(−1.76)	(2.24)	(2.07)	**(2.80)**
FF3	0.001	−0.002	0.030	0.003	**0.022**
	(1.07)	(−1.74)	(2.38)	(2.16)	**(3.24)**
FF5	0.001	−0.002	0.028	0.003	**0.023**
	(1.50)	(−1.48)	(2.06)	(1.90)	**(3.31)**
FF5 + MOM	0.001	−0.002	0.026	0.003	**0.022**
	(1.40)	(−1.51)	(1.87)	(1.88)	**(3.22)**

Table A3: Panel Regressions (Full Sample) of Short VIX One-Factor Model
The table reports the results from the panel regressions of the following Equation: $r_t^i = \alpha + a^i + \beta \text{Benchmark}_t + \varepsilon_t^i$, for each hedge fund strategy with hedge fund fixed effects, where a^i is a hedge fund-specific fixed effect, α is unexplained return variation, Benchmark refers to the monthly returns from the ShortVIX$_t$ in Panel A and OTM Short Put$_t$ in Panel B, and β is the coefficient of interest that measures how well the benchmark explains returns. The sample period is May 2004 to December 2022 at a monthly frequency. Coefficient estimates (expressed in decimal form) of α and β along with their respective t-statistics and sample size are also reported. Standard errors are clustered at the hedge fund level. Asterisks denote the levels of statistical significance of beta: 10% level (*), 5% level (**) and 1% level (***).

Panel A: Short VIX as the Hedge Fund Benchmark

Hedge Fund Strategy	$\hat{\beta}$	t-stat	$\hat{\alpha}$	t-stat	R^2
Convertible Arbitrage***	0.180	6.99	−0.013	−4.16	0.21
Dedicated Short Bias***	−0.338	−4.48	0.006	0.87	0.35
Emerging Markets***	0.283	39.74	−0.003	−1.76	0.26
Equity Market Neutral***	0.080	10.01	0.002	1.25	0.12
Event Driven***	0.183	29.41	0.002	1.90	0.38
Fixed Income Arb***	0.084	17.10	0.003	2.35	0.21
Fund of Funds***	0.109	73.17	0.000	0.08	0.08
Global Macro***	0.044	4.83	0.006	2.77	0.04
L/S Equity Hedge***	0.225	59.83	−0.001	−1.80	0.33
Managed Futures	0.008	0.44	0.027	10.65	0.06
Multi Strat.***	0.108	29.73	0.001	0.29	0.17
Options Strategy***	0.050	2.67	0.002	0.58	0.12
All Strategies***	0.149	77.41	0.000	0.84	0.15

Panel B: OTM Short Put as the Hedge Fund Benchmark

Hedge Fund Strategy	$\hat{\beta}$	t-stat	$\hat{\alpha}$	t-stat	R^2
Convertible Arbitrage***	0.685	7.58	−0.013	−4.37	0.27
Dedicated Short Bias***	−0.987	−8.29	0.002	0.48	0.39
Emerging Markets***	0.988	40.01	−0.001	−0.56	0.31
Equity Market Neutral***	0.189	7.14	0.004	2.43	0.09
Event Driven***	0.584	27.69	0.004	4.80	0.38
Fixed Income Arb***	0.290	15.16	0.004	2.81	0.24
Fund of Funds***	0.330	65.20	0.002	3.04	0.07
Global Macro***	0.161	6.11	0.007	2.91	0.05
L/S Equity Hedge***	0.710	52.65	0.002	2.78	0.33
Managed Futures	−0.025	−0.26	0.026	9.36	0.05
Multi Strat.***	0.380	24.28	0.001	0.59	0.18
Options Strategy***	0.199	2.76	0.001	0.31	0.10
All Strategies***	0.481	70.31	0.002	5.24	0.15

Figures

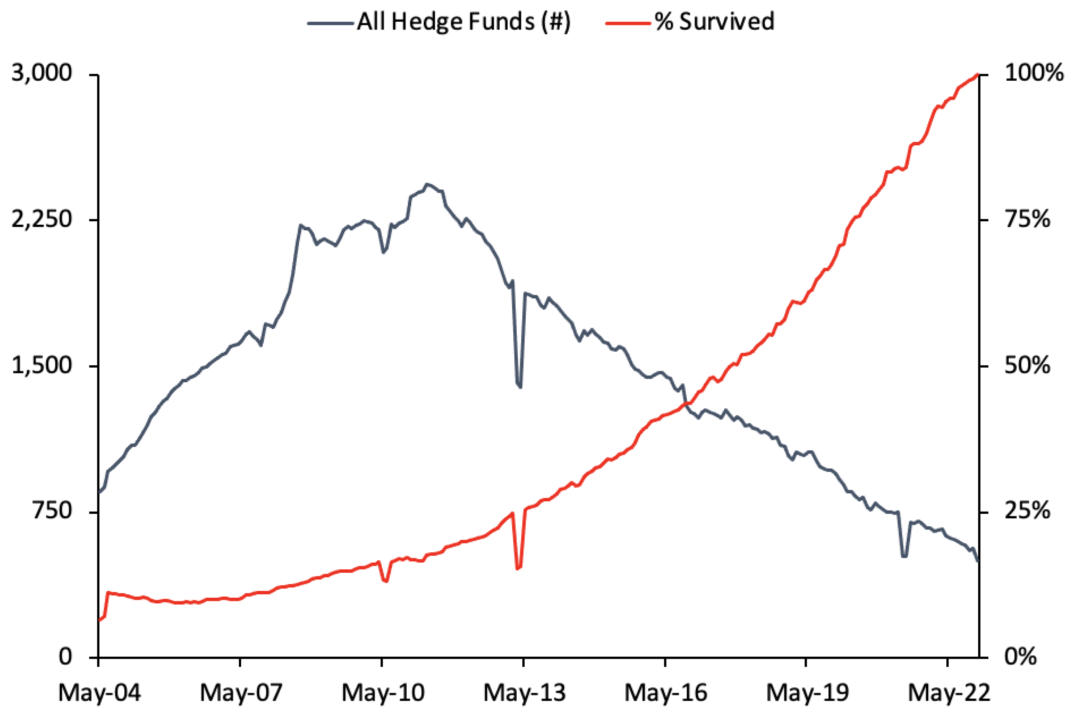
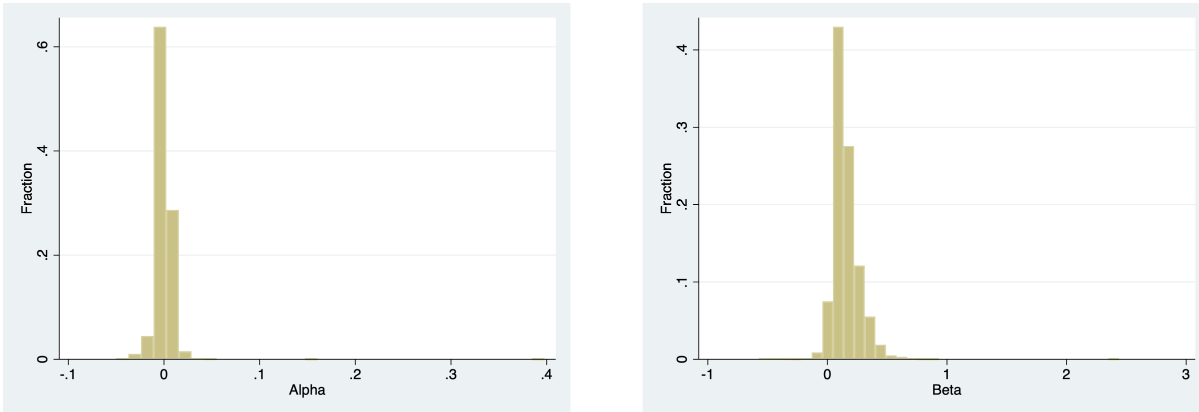


Figure A1: Hedge Fund Counts over Time
The figure displays the total number of hedge funds (blue line) reported in the cleaned TASS Database. In addition, the percent of survived hedge funds is also displayed (red line). The sample period is May 2004 to December 2022.

Figure A2: Rolling 36-Month Correlations
The figure displays the rolling 36-month correlations between the Short VIX futures strategy returns and the HFRI 500 Composite Index returns. The sample period is from December 2007 until December 2022.



(a) $\hat{\alpha}$ (b) $\hat{\beta}$

Figure A3: Edge Parameter Estimate Histograms
The figure shows the histograms of the estimates for the respective parameter that comprises the Edge measure. Panel A and B display the histogram of the alpha and beta estimates, respectively. Each histogram is based on the average parameter estimate, by hedge fund, over the OOS period (May 2008 to December 2022).

Derivation of Optimal Coefficients

Here we derive the estimator that modifies the OLS estimator by adding a penalization term based on maximizing an investor's CE.

$$\begin{aligned}\min_{\alpha,\beta} \sum_{t=1}^{T-1} \left(R_{t+1} - R_t^f - \alpha - \beta X_t\right)^2 \\ - \lambda \left[\bar{R} - \frac{\gamma}{2} \frac{1}{T-2} \sum_{t=1}^{T-1} \left(w_t R_{t+1} + (1 - w_t) R_t^f - \bar{R}\right)^2\right],\end{aligned} \tag{24}$$

where $w_t = \frac{1}{\gamma} \frac{\mathbb{E}_t\left[R_{t+1} - R_t^f\right]}{\sigma^2_{t+1|t}} = \frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}$ and $\bar{R} = \frac{1}{T-1} \sum_{t=1}^{T-1} w_t R_{t+1} + (1 - w_t) R_t^f$. Substituting the expressions for w_t and $\bar{R}$ into the above minimization problem gives us,

$$\begin{aligned}\min_{\alpha,\beta} \sum_{t=1}^{T-1} \left(R_{t+1} - R_t^f - \alpha - \beta X_t\right)^2 - \lambda \Bigg\{ \frac{1}{T-1} \sum_{t=1}^{T-1} \left(\frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}} R_{t+1} + \left(1 - \frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}\right) R_t^f \right) \\ - \frac{\gamma}{2} \frac{1}{T-2} \sum_{t=1}^{T-1} \Bigg\{ \frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}} R_{t+1} + \left(1 - \frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}\right) R_t^f \\ - \left(\frac{1}{T-1} \sum_{t=1}^{T-1} \left[\frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}} R_{t+1} + \left(1 - \frac{1}{\gamma} \frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}\right) R_t^f \right] \right) \Bigg\}^2 \Bigg\}\end{aligned}$$

Taking the first order condition with respect to β yields,

$$
\begin{aligned}
\frac{\partial}{\partial \beta} &= -2X_t\left(R_{t+1} - R_t^f - \alpha - \beta X_t\right) - \lambda \frac{X_t R_{t+1}}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)} + \lambda \frac{X_t R_t^f}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)} \\
&+ \left[\lambda \frac{X_t R_{t+1}}{(T-2)\left(\gamma \sigma^2_{t+1|t}\right)} - \lambda \frac{X_t R_t^f}{(T-2)\left(\gamma \sigma^2_{t+1|t}\right)} - \lambda \frac{X_t R_{t+1}}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)} + \lambda \frac{X_t R_t^f}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)}\right] \\
&\cdot \left[\frac{1}{\gamma}\frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}} R_{t+1} + \left(1 - \frac{1}{\gamma}\frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}\right) R_t^f - \left(\frac{1}{T-1}\sum_{t=1}^{T-1}\left[\frac{1}{\gamma}\frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}} R_{t+1} + \left(1 - \frac{1}{\gamma}\frac{\alpha + \beta X_t}{\sigma^2_{t+1|t}}\right) R_t^f\right]\right)\right] \\
&= -2X_t\left(R_{t+1} - R_t^f - \alpha - \beta X_t\right) - \lambda \frac{X_t R_{t+1}}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)} + \lambda \frac{X_t R_t^f}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)}, \quad \text{for} \ \ t = 1, ..., T-1
\end{aligned}
$$

Setting the FOC equal to zero,

$$
\begin{aligned}
-2X_t\left(R_{t+1} - R_t^f - \alpha - \beta X_t\right) - \lambda \frac{X_t R_{t+1}}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)} + \lambda \frac{X_t R_t^f}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)} &= 0 \\
-2X_t\left(R_{t+1} - R_t^f\right) + 2X_t\alpha + 2X_t^2\beta - \frac{\lambda X_t}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)}\left(R_{t+1} - R_t^f\right) &= 0 \\
2X_t^2\beta = 2X_t\left(R_{t+1} - R_t^f\right) - 2X_t\alpha + \frac{\lambda X_t}{(T-1)\left(\gamma \sigma^2_{t+1|t}\right)}&\left(R_{t+1} - R_t^f\right)
\end{aligned}
\tag{25}
$$

In matrix notation, the new estimator is:

$$
\hat{\beta} = \left(X'X\right)^{-1} X'Y + \underbrace{\frac{\lambda \Sigma^{-1}\left(X'X\right)^{-1} X'Y}{2\left(T-1\right)\gamma} + \alpha X^{-1}}_{\text{Bias}}, \ \ \text{where} \ \ Y = R_{t+1} - R_t^f \tag{26}
$$

References

Agarwal, Vikas, Wei Jiang, Yueha Tang, and Baozhong Yang, 2013, Uncovering hedge fund skill from the portfolio holdings they hide, *The Journal of Finance* 68, 739–783.

Agarwal, Vikas, and Narayan Y. Naik, 2000, Multi-period performance persistence analysis of hedge funds, *The Journal of Financial and Quantitative Analysis* 35, 327–342.

Agarwal, Vikas, Stefan Ruenzi, and Florian Weigert, 2023, Unobserved performance of hedge funds, Working Paper.

Agarwal, Vikas, Stefan Ruenzi, and Florian Weigert, 2024, Unobserved performance of hedge funds, *The Journal of Finance* 79, 3203–3259.

Ahn, Dong-Hyun, Jennifer Conrad, and Robert F. Dittmar, 2009, Basis assets, *The Review of Financial Studies* 22, 5133–5174.

Anatolyev, Stanislav, and Anna Mikusheva, 2022, Factor models with many assets: Strong factors, weak factors, and the two-pass procedure, *Journal of Econometrics* 229, 103–126.

Ardia, David, Laurent Barras, Patrick Gagliardini, and Olivier Scaillet, 2024, Is it alpha or beta? decomposing hedge fund returns when models are misspecified, *Journal of Financial Economics* 154, Forthcoming.

Avramov, Doron, Robert Kosowski, Narayan Y. Naik, and Melvyn Teo, 2011, Hedge funds, managerial skill, and macroeconomic variables, *Journal of Financial Economics* 99, 672–692.

Back, Kerry, Alan D. Crane, and Kevin Crotty, 2018, Skewness consequences and seeking alpha, *The Review of Financial Studies* 31, 4720—-4761.

Barth, Daniel, Juha Joenväärä, Mikko Kauppila, and Russ Wermers, 2023, A bias-free assessment of the hedge fund industry: A new evaluation of total assets, alphas, and the flow-performance relation, Working Paper.

Bollen, Nicolas P.B., Juha Joenväärä, and Mikko Kauppila, 2024, Decreasing returns to scale has eroded hedge fund performance persistence, Critical Finance Review, Forthcoming.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected stock returns and variance risk premia, *Review of Financial Studies* 22, 4463–4492.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric portfolio policies: Exploring characteristics in the cross-section of equity returns, *Review of Financial Studies* 22, 3411–3447.

Brown, Gregory, Christian Lundblad, and William Volckmann, 2024, What do we know about institutional-quality hedge funds?, Institute for Private Capital.

Brown, Stephen J., William N. Goetzmann, and Roger G. Ibbotson, 1999, Offshore hedge funds: Survival and performance, 1989–95, *The Journal of Business* 72, 91–117.

Bryzgalova, Svetlana, 2015, Spurious factors in linear asset pricing models, Technical Report, Stanford University.

Bryzgalova, Svetlana, Victor DeMiguel, Sicong Li, and Markus Pelger, 2023a, Asset-pricing factors with economic targets, Working Paper.

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard, 2023b, Bayesian solutions for the factor zoo: We just ran two quadrillion models, *The Journal of Finance* 78, 487–557.

Bryzgalova, Svetlana, Markus Pelger, and Jason Zhu, 2023c, Forest through the trees: Building cross-sections of stock returns, Journal of Finance, Forthcoming.

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Campbell, John Y., and L. M. Viceira, 2002, Strategic asset allocation: Portfolio choice for long-term investors, *Oxford University Press* .

Cao, Jie, Gang Li, Xintong Zhan, and Guofu Zhou, 2023, Betting against the crowd: Option trading and market risk premium, *Working Paper* .

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

Cederburg, Scott, Travis L. Johnson, and Michael S. O'Doherty, 2023, On the economic significance of stock return predictability, *Review of Finance* 27, 619–657.

Cenesizoglu, Tolga, and Allan Timmermann, 2012, Do return prediction models add economic value?, *Journal of Banking & Finance* 36, 2974–2987.

Chen, Yong, Michael Cliff, and Haibei Zhao, 2017, Hedge funds: The good, the bad, and the lucky, *The Journal of Financial and Quantitative Analysis* 52, 1081–1109.

Chen, Yong, Sophia Z. Li, Yushan Tang, and Guofu Zhou, 2024, Anomalies as new hedge fund factors, Working Paper.

Chen, Yong, and Bing Liang, 2007, Do market timing hedge funds time the market?, *The Journal of Financial and Quantitative Analysis* 42, 827–856.

Chen, Yong, and Hanjiang Zhang, 2023, Do hedge funds hedge? evidence from risk gap, Working Paper.

Cheng, Ing-Haw, 2019, The vix premium, *The Review of Financial Studies* 32, 180–227.

Chinco, Alex, Andreas Neuhierl, and Michael Weber, 2021, Estimating the anomaly base rate, *Journal of Financial Economics* 140, 101–126.

Christoffersen, Peter, and Kris Jacobs, 2004, The importance of the loss function in option valuation, *Journal of Financial Economics* 72, 291–318.

Cochrane, John H., 2011, Presidential address: Discount rates, *The Journal of Finance* 66, 1047–1108.

Couts, Spencer J., Andrei S. Gonçalves, and Andrea Rossi, 2020, Unsmoothing returns of illiquid funds, Working Paper.

Daniel, Kent, Lira Mota, Simon Rottke, and Tano Santos, 2020, The cross-section of risk and returns, *The Review of Financial Studies* 33, 1927–1979.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* 22, 1915–1953.

Dybvig, Philip, and Stephen Ross, 1985, The analytics of performance measurement using a security market line, *The Journal of Finance* 40, 401–416.

Edwards, Franklin R., and Mustafa Onur Caglayan, 2001, Hedge fund performance and manager skill, *The Journal of Futures Markets* 21, 1003–1028.

Engle, Robert F., 1993, A comment on hendry and clements on the limitations of comparing mean square forecast errors, *Journal of Forecasting* 12, 642–644.

Faias, Jose Afonso, and Pedro Santa-Clara, 2017, Optimal option portfolio strategies: Deepening the puzzle of index option mispricing, *Journal of Financial and Quantitative Analysis* 52, 277–303.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

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

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

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

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *The Journal of Finance* 75, 1327–1370.

Fung, William, and David A. Hsieh, 2001, The risk in hedge fund strategies: Theory and evidence from trend followers, *The Review of Financial Studies* 14, 313–341.

Fung, William, and David A. Hsieh, 2002, Asset-based style factors for hedge funds, *Financial Analysts Journal* 58, 16–27.

Fung, William, and David A. Hsieh, 2004, Hedge fund benchmarks: A risk-based approach, *Financial Analysts Journal* 60, 65–80.

Getmansky, Mila, Andrew W. Lo, and Igor Makarov, 2004, An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529–609.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *The Review of Financial Studies* 34, 3456–3496.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, 2023a, Prediction when factors are weak, Working Paper.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, 2023b, Test assets and weak factors, The Journal of Finance, Forthcoming.

Giglio, Stefano, Dacheng Xiu, and Dake Zhang, 2023c, Test assets and weak factors, Journal of Finance, Forthcoming.

Gormsen, Niels J., and Christian Skov Jensen, 2022, Higher-moment risk, Working Paper.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Goyal, Amit, Ivo Welch, and Athanasse Zafirov, 2024, A comprehensive 2022 look at the empirical performance of equity premium prediction, *Review of Financial Studies* .

Griffin, John M., and Jin Xu, 2009, How smart are the smart guys? a unique view from hedge fund stock holdings, *The Review of Financial Studies* 22, 2531–2570.

Grinblatt, Mark, Gergana Jostova, Lubomir Petrasek, and Alexander Philipov, 2020, Style and skill: Hedge funds, mutual funds, and momentum, *Management Science* 66, 5505–5531.

Han, Bing, and Gang Li, 2020, Aggregate implied volatility spread and stock market returns, *Management Science* .

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Hoerl, Arthur E., and Robert W. Kennard, 1970, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics* 12, 55–67.

Huang, Dashan, and Guofu Zhou, 2017, Upper bounds on return predictability, *Journal of Financial and Quantitative Analysis* 52, 401–425.

Jagannathan, Ravi, Alexey Malakhov, and Dmitry Novikov, 2010, Do hot hands exist among hedge fund managers? an empirical evaluation, *The Journal of Finance* 65, 217–255.

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, 2023, Is there a replication crisis in finance?, *The Journal of Finance* 78, 2465–2518.

Jiang, George J., Bing Liang, and Huacheng Zhang, 2021, Hedge fund manager skill and style-shifting, *Management Science* 68, 2284–2307.

Joenväärä, Juha, Mikko Kauppila, Robert Kosowski, and Pekka Tolonen, 2021, Hedge fund performance: Are stylized facts sensitive to which database one uses?, *Critical Finance Review* 10, 271–327.

Johnson, Travis L., 2019, A fresh look at return predictability using a more efficient estimator, *Review of Asset Pricing Studies* 9, 1–46.

Jorion, Philippe, and Christopher Schwarz, 2019, The fix is in: Properly backing out backfill bias, *The Review of Financial Studies* 32, 5048–5099.

Jurek, Jakub W., and Erik Stafford, 2015, The cost of capital for alternative investments, *The Journal of Finance* 70, 2185–2226.

Kan, Raymond, and Chu Zhang, 1999, Two-pass tests of asset pricing models with useless factors, *The Journal of Finance* 54, 203–235.

Kelly, Bryan, Semyon Malamud, and Kangying Zhou, 2024, The virtue of complexity in return prediction, *Journal of Finance* 79, 459–503.

Kelly, Bryan T., and Hao Jiang, 2012, Tail risk and hedge fund returns, Working Paper.

Kelly, Bryan T., Seth Pruitt, and Yinan Su, 2019, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kleibergen, Frank, 2009, Tests of risk premia in linear factor models, *Journal of Econometrics* 149, 149–173.

Koijen, Ralph S.J., Tobias J. Moskowitz, Lasse H. Pedersen, and Evert B. Vrugt, 2018, Carry, *Journal of Financial Economics* 127, 197–225.

Koijen, Ralph S.J., Robert J. Richmond, and Motohiro Yogo, 2023, Which investors matter for equity valuations and expected returns?, *The Review of Economic Studies* 91, 2387–2424.

Kozak, Serhiy, and Stefan Nagel, 2023, When do cross-sectional asset pricing factors span the stochastic discount factor?, Working Paper.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2018, Interpreting factor models, *The Journal of Finance* 73, 1183–1223.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020a, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020b, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Lettau, Martin, and Markus Pelger, 2020, Estimating latent asset-pricing factors, *Journal of Econometrics* 218, 1–31.

Li, Jiahan, and Ilias Tsiakas, 2017, Equity premium prediction: The role of economic and statistical constraints, *Journal of Financial Markets* 36, 56–75.

MacLean, Leonard C, Edward O. Thorp, and William T. Ziemba, 2011, *The Kelly Capital Growth Investment Criterion: Theory and Practice* (World Scientific Handbook in Financial Economics Series: Volume 3).

McLean, R. David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

Mei, Xiaoling, and Francisco J Nogales, 2018, Portfolio selection with proportional transaction costs and predictability, *Journal of Banking & Finance* 94, 131–151.

Nagel, Stefan, 2021, *Machine Learning in Asset Pricing* (Princeton University Press).

Neely, Christopher J, David E Rapach, Jun Tu, and Guofu Zhou, 2014, Forecasting the equity risk premium: the role of technical indicators, *Management science* 60, 1772–1791.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Novy-Marx, Robert, and Mihail Velikov, 2016, A taxonomy of anomalies and their trading costs, *The Review of Financial Studies* 29, 104–147.

Pasquariello, Paolo, 2014, Financial market dislocations, *The Review of Financial Studies* 27, 1868–1914.

Pedersen, Lasse Heje, 2015, *Efficiently Inefficient: How Smart Money Invests and Market Prices are Determined* (Princeton University Press).

Pettenuzzo, Davide, Allan Timmermann, and Rossen Valkanov, 2014, Forecasting stock returns under economic constraints, *Journal of Financial Economics* 114, 517–553.

Rapach, David, and Guofu Zhou, 2013, Forecasting stock returns, *Handbook of Economic Forecasting* 2, 328–383.

Rapach, David E, Matthew C Ringgenberg, and Guofu Zhou, 2016, Short interest and aggregate stock returns, *Journal of Financial Economics* 121, 46–65.

Shiller, Robert, 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–436.

Tibshirani, Robert, 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 267–288.

Titman, Sheridan, and Cristian Tiu, 2011, Do the best hedge funds hedge?, *The Review of Financial Studies* 24, 123–168.

Treynor, Jack, and Kay Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131–136.