Essays in Derivatives Markets

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

submitted by

Mathis Rudolf Werner Mörke

from

Germany

Approved on the application of

Prof. Dr. Manuel Ammann

and

Prof. Paul Söderlind, PhD

Dissertation no. 5301

Difo-Druck GmbH, Untersiemau 2023

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

St.Gallen, October 24, 2022

The President:

Prof. Dr. Bernhard Ehrenzeller

Für Mama und Papa.
Danke.

Acknowledgements

This dissertation has been written with the continuous and great support of a number of people I am highly indebted to. At this point, I wish to express my gratitude to them.

First, I would like to thank my supervisor Prof. Dr. Manuel Ammann whose ever open door, his great empathy, support and guidance have essentially shaped this thesis. Thank you, Manuel. Further, I would like to thank Prof. Paul Söderlind, PhD. for valuable comments and inputs as my co-supervisor.

I would also like to thank my colleagues and friends from the Swiss Institute of Banking and Finance, namely Alexander Cochardt, Stephan Heller, Sebastian Fischer and Simon Straumann. Without their endless discussions, creative inputs and never-stopping motivation this thesis would not have become what it is. Especially, I want to thank my co-author Florian Weigert who became a friend during the time of my PhD. Thank you, Florian.

Finally, I would like to thank my family and Kathrin for their continuous support, love, and encouragement. Thank you.

St.Gallen, October 2022
Mathis Rudolf Werner Mörke

Contents

Abstract **vi**

Zusammenfassung **vii**

Liquidity Provision to Leveraged ETFs and Equity Options Rebalancing Flows **1**

1 Introduction . . . 3
2 Data and Measurements . . . 10
2.1 Data Sources . . . 10
2.2 Measuring Gamma Hedging Pressure . . . 12
2.3 Measuring Leveraged ETF Rebalancing Pressure . . . 14
2.4 Summary Statistics . . . 15
2.5 Abnormal Order Flow . . . 17
3 Empirical Results . . . 17
3.1 Main Results . . . 17
3.2 Discretionary Rebalancing Times . . . 22
3.3 Reversal . . . 24
3.4 Liquidity Provision . . . 26
3.5 Trading Strategies . . . 28
3.6 The Evolution of the Imbalance and its Price Impact . . . 31
4 Conclusion . . . 32
Appendix A Detailed Summary Statistics on Γ^{HP} and Ω^{LETF} . . . 35
Appendix B Errors in Measuring Γ^{HP} – A Simulation Study . . . 36
Appendix C Controlling for Alternative Explanations . . . 37
Appendix D Liquidity Provision: Evidence from Spreads . . . 39
Appendix E Performance of Trading Strategies Over Time . . . 40
Bibliography . . . 41

Online Appendix A Cleaning of High-Frequency Data 47
Online Appendix B Summary Statistics of the Underlying 49
Online Appendix C Summary Statistics of Leveraged ETFs 51
Online Appendix D Correlation of Exchange Specific Γ^{HP} and Total Γ^{HP} . 52
Online Appendix E Different ADV Measures 53
Online Appendix F Different Sets of Delta-Hedgers 54
Online Appendix G Γ^{HP} based on Older Positions 56
Online Appendix H Scaled Returns 60
Online Appendix I Mechanical Flows 61
Online Appendix J Effects Across Industries 64
Online Appendix Bibliography 65

Credit Variance Risk Premiums **66**
1 Introduction 68
2 Methodology 72
2.1 Credit Default Index Swaptions 72
2.2 Variance Swap Contracts 75
3 Data 77
4 Empirical Results 80
4.1 Preliminary Data Analysis 80
4.2 Credit Variance Risk Premiums 83
4.3 Credit Variance Risk Premiums across Different Market Regimes . 89
4.4 Structural Model Variables as Drivers 90
4.5 Risk-adjusted Returns 92
4.6 Comovement with Equity Variance and Variance-of-variance Risk Premiums 96
4.7 Predictive Power of Credit Implied Volatility 97
5 Conclusion 99
Appendix A Synthesizing Variance Swap Contracts 102
Appendix B Spread vs. Price Realized Volatility 105
Appendix C Statistical Significance 106
Appendix D Term Structure of Credit Variance Risk Premiums 107
Appendix E Transaction Costs 109
Bibliography 111

Commodity Tail Risk **116**
1 Introduction . . . 118
2 Data and Methodology . . . 121
2.1 Data . . . 121
2.2 Tail Risk . . . 122
2.3 Futures Returns . . . 124
2.4 Explanatory Variables . . . 125
3 Empirical Results . . . 128
3.1 Summary Statistics . . . 128
3.2 Determinants of Tail Risk . . . 132
3.3 Portfolio Sorts . . . 141
4 Conclusion . . . 143
Appendix A CRB Data Handling . . . 145
Appendix B Change in Tail Risk as Cross-sectional Predictor . . . 147
Bibliography . . . 149

Option Return Predictability with Machine Learning and Big Data **152**
1 Introduction . . . 154
2 Related Literature . . . 159
3 Methodology and Performance Evaluation . . . 161
4 Data and Variable Definitions . . . 164
4.1 Option Returns . . . 165
4.2 Summary Statistics . . . 166
4.3 Option and Equity Characteristics . . . 167
5 Predicting Option Returns . . . 169
5.1 Predictability Comparison . . . 169
5.2 Machine Learning Portfolios . . . 179
5.3 Which Characteristics Matter? . . . 185
5.4 Impact of the Information Set . . . 190
6 Sources of Option Return Predictability . . . 193
6.1 Presence of Informed Investors . . . 194
6.2 Option Demand by Types of Market Participants . . . 197
6.3 Option Mispricing . . . 199
7 Conclusion . . . 200
Bibliography . . . 202

Online Appendix A Methods Used . . . 212
Online Appendix B Estimation Details . . . 213
Online Appendix C Option-Based Characteristics . . . 215
Online Appendix D Classification of Characteristics . . . 227
Online Appendix E Additional Summary Statistics . . . 237
Online Appendix F Model Comparison . . . 242
Online Appendix G Predicting All vs. Only Short-Term ATM Options . . . 246
Online Appendix H Machine Learning Portfolios . . . 249
Online Appendix I Which Characteristics Matter? . . . 257
Online Appendix J Impact of the Information Set . . . 264
Online Appendix K Sources of Option Return Predictability . . . 266
Online Appendix Bibliography . . . 269

An Autoencoder Based Factor Model for Option Returns 274
1 Introduction . . . 276
2 Methodolgy and Performance Evaluation . . . 280
2.1 IPCA . . . 280
2.2 Autoencoder . . . 281
2.3 Conditional Autoencoder . . . 281
2.4 Model Set . . . 283
2.5 Estimation Details . . . 284
2.6 Performance Evaluation . . . 286
3 Data and Variable Definitions . . . 286
3.1 Option Filters . . . 286
3.2 Option Returns . . . 287
3.3 Summary Statistics . . . 288
3.4 Option Characteristics . . . 290
4 Empirics . . . 290
4.1 Statistical Performance Evaluation . . . 290
4.2 Economic Performance Evaluation . . . 296
4.3 Risk Premia vs. Mispricing . . . 301
4.4 Characteristics Importance . . . 302
5 Conclusion . . . 304
Appendix A Hyperparameters . . . 306
Appendix B Option Characteristics . . . 307

Appendix C DM Forecast Comparisons 308
C.1 DM Forecast Comparison for R^2_{total} for Individual Option Contracts 308
C.2 DM Forecast Comparison for R^2_{pred} for Individual Option Contracts 309
C.3 DM Forecast Comparison for R^2_{total} for Managed Portfolios 310
C.4 DM Forecast Comparison for R^2_{pred} for Managed Portfolios 311
Appendix D Performance Assessment for 2003 – 2017 312
D.1 Out-of-sample R^2s 312
D.2 Out-of-Sample Annualized Sharpe Ratios of Long-Short Portfolios 313
D.3 Performance of Out-Of-Sample Long-Short Portfolios 314
Appendix E Performance Attribution via Shapley Value 315
E.1 Methodology 315
E.2 Performance Attribution of Out-Of-Sample Long-Short Value-Weighted Portfolios 317
Bibliography 318

Curriculum Vitae **322**

Abstract

This dissertation consists of five independent papers.

Rebalancing of leveraged ETFs and delta-hedging of equity options are two distinct and economically significant sources of order flow and liquidity demands. *Liquidity Provision to Leveraged ETFs and Equity Options Rebalancing Flows* finds that delta-hedging effects are persistent, those stemming from leveraged ETFs are decreasing significantly over time. These dynamics arise from different intermediation structures, generating heterogeneous levels of information asymmetry. While leveraged ETF providers generate perfectly predictable flows, option delta-hedgers have flexibility in deciding the strategic timing of their rebalancing, resulting in less predictable flows.

Credit Variance Risk Premiums studies the pricing of variance risk in credit markets by employing a unique data set of credit swaptions. Returns of credit variance swaps are negative and economically large, irrespective of the credit rating class. Shorting credit variance swaps yields annualized Sharpe ratios well above their counterparts in other asset classes. The returns remain highly statistically significant when accounting for transaction costs and cannot be explained by established risk-factors and structural model variables.

Commodity Tail Risks investigates the cross-section of tail risks in commodity markets. Left and right tail risks play an equally important role in commodity markets. For commodity producers, negative price jumps (left tail risk) might have devastating consequences. For commodity consumers though, e.g., companies that process commodities, right tail risks matter more. Both left and right tail risk implied by option markets are large. Moreover, both risks are priced in the cross-section of commodity futures returns.

The option market has witnessed significant growth in recent years. Whereas risk factors in equity markets have been studied in detail, the same does not apply for option returns. *Option Return Predictability with Machine Learning and Big Data* finds that option returns are highly predictable. Allowing for nonlinearities significantly increases the out-of-sample performance of option and stock characteristics in predicting future option returns. Although option-based characteristics are the most important standalone predictors, stock-based measures offer substantial incremental predictive power.

Most of the factor models assume a linear structure. There are, nonetheless, no obvious theoretical or intuitive justifications for this assumption. *An Autoencoder Based Factor Model for Option Returns* adopts a novel latent factor model, which allows for nonlinearities, and applies it to index options. The model excels in explaining variation in risk compensation and out-of-sample trading strategies.

Zusammenfassung

Die vorliegende Dissertation besteht aus fünf unabhängigen Studien.

Portfolioumschichtungen von gehebelten ETFs und das Delta-Hedging von Market Maker in Optionsmärkten sind zwei unterschiedliche, aber bedeutsame Quellen für Liquiditätsanforderungen auf dem amerikanischen Aktienmarkt. Die erste Studie zeigt, dass Delta-Hedging Effekte beständig sind, während die von gehebelten börsengehandelten Fonds herrührenden Effekte im Laufe der Zeit abnehmen. Diese Dynamik ergibt sich aus unterschiedlichen Intermediationsstrukturen, die ein heterogenes Maß an Informationsasymmetrie erzeugen.

Die zweite Studie untersucht die Varianzrisikoprämie auf Kreditmärkten unter Verwendung eines neuartigen Datensatzes von Optionen auf Kreditausfallversicherungen. Handelsstrategien, die die Varianzrisikoprämie einzunehmen versuchen, erzielen ökonomisch bedeutsame Renditen. Die Renditen bleiben auch unter Berücksichtigung von Transaktionskosten statistisch hoch signifikant und können nicht durch etablierte Risikofaktoren oder strukturelle Modellvariablen erklärt werden.

Die dritte Studie untersucht die Risikoprämien für die Absicherung gegen massive Preiseinbrüche oder -anstiege auf Rohstoffmärkten. Während auf dem Aktienmarkt die Angst vor Preisverfallen überwiegt, spielt auf den Rohstoffmärkten beides eine wichtige Rolle. Für Rohstoffproduzenten können negative Preissprünge verheerende Folgen haben. Für Rohstoffkonsumenten, z. B. Unternehmen, die Rohstoffe verarbeiten, sind jedoch Risiken durch Preisanstiege von grösserer Bedeutung. Beide Risikoprämien, welche mittels Optionen geschätzt werden können, sind auf den Rohstoffmärkten gross und in den Renditen von Rohstofftermingeschäften eingepreist.

Der Optionsmarkt hat in den letzten Jahren ein erhebliches Wachstum erfahren. Während Risikofaktoren auf den Aktienmärkten eingehend untersucht worden sind, gilt dies nicht für Optionsrenditen. Die vierte Studie stellt fest, dass die Prognosekraft bei Optionsrenditen hoch ist. Die Berücksichtigung von Nichtlinearitäten und Interaktionseffekten erhöht die Prognosegüte erheblich. Optionsbasierte Merkmale sind die wichtigsten Prädiktoren, jedoch bieten aktienbasierte Maße eine zusätzliche Vorhersagekraft.

Die meisten Faktorenmodelle bei der Preisgestaltung von Vermögenswerten gehen von einer linearen Struktur aus. Es gibt jedoch keine offensichtlichen theoretischen oder intuitiven Begründungen für diese Annahme. Die fünfte Studie nimmt ein neuartiges latentes Faktormodell an, das Nichtlinearitäten zulässt, und wendet es auf Optionsrenditen für den S&P 500 Index an. Das Modell eignet sich hervorragend zur Erklärung von zukünftigen Schwankungen bei der Risikokompensation und zur Bildung von Handelsstrategien.

Chapter 1

Liquidity Provision to Leveraged ETFs and Equity Options Rebalancing Flows: Evidence from End-of-Day Stock Prices

*Andrea Barbon, Heiner Beckmeyer, Andrea Buraschi, Mathis Mörke**

ABSTRACT

Rebalancing of leveraged ETFs and delta-hedging of equity options are two distinct and economically significant sources of order flow and liquidity demands. We show that they induce significant end-of-day momentum and mean-reversion in stock returns, while the effects dissipate within the next trading day. While gamma effects are persistent throughout our sample, those stemming from leveraged ETFs are decreasing significantly over time. We argue that these dynamics arise from different intermediation structures, generating heterogeneous levels of information asymmetry on rebalancing flows. While leveraged ETF providers generate perfectly predictable flows because of their strict mandate, option delta-hedgers have flexibility in deciding the strategic timing of their rebalancing, resulting in less predictable flows. Consistent with this view, we find that leveraged ETF flows attract more liquidity provision and that their effect on prices is shorter-lived. Our findings suggest that information disclosure is beneficial for market liquidity.

Keywords: Liquidity Provision, Gamma Exposure, Option Delta-Hedging, Leveraged ETF, End-of-Day Momentum

JEL classification: G12, G13, G14, G23.

*We thank Kris Jacobs (discussant), Neil Pearson, Jason Wei (discussant), Dmitriy Muravyev, Mahendrarajah Nimalendran (discussant), Danny Qin (discussant), John J. Shim (discussant), Fabrice Riva (discussant), Alexander Wagner, and Angelo Ranaldo, as well as participants at the FMA Conference on Derivatives and Volatility 2021, the 37th International Conference of the French Finance Association, the EFMA annual meeting 2021, the 5th SAFE Market Microstructure Conference, the NFA annual meeting 2021, the FMA annual meeting 2021, the 13th annual Hedge Fund Research Conference, the MFA annual meeting 2022, internal seminars at Goldman Sachs, Bank of America Merrill Lynch, and Morgan Stanley, and the Brown Bag Seminar of the University of St.Gallen for insightful and constructive comments. This paper has been previously circulated under the title The Role of Leveraged ETFs and Option Market Imbalances on End-of-Day Price Dynamics. Parts of this paper were written while Mathis Moerke visited the Kellogg School of Management whose hospitality is gratefully acknowledged.

1. Introduction

The role of options markets and leveraged ETFs on the dynamics of the underlying stock prices has recently received significant attention, attracting negative press coverage for potentially contributing to market volatility during already turbulent times. The trading activity in options was blamed to increase the violent stock swings during the February-March 2020 Covid-19 selloff. The Wall Street Journal wrote:

> "Investors searching for clues on what drove the back-to-back drops in the stock market are pointing to the options market as a contributor, saying hedging activity by traders may have exacerbated the decline."
>
> Wall Street Journal, Feb. 27, 2020[1]

In this paper, we investigate the effects of portfolio rebalancing on the intraday dynamics of stock prices at the end of the trading day, arising from two distinct derivative instruments.

The first channel relates to the options market, where market makers and broker/dealers provide liquidity to clients who want to take positions in stock options. As they have institutional incentives to avoid directional exposures, they typically delta-hedge their positions. Since the option delta changes when the value of the underlying changes, market makers need to regularly update their positions to maintain delta neutrality. The direction of the resulting flows depends on their initial gamma imbalance and the price movement of the underlying asset. Suppose, for instance, that the price of a stock experiences a positive jump, driven by positive unexpected news about future cash flows. If the aggregate gamma of market makers is initially negative, maintaining delta-neutrality requires the purchase of additional shares in the underlying stock. On the contrary, a positive gamma exposure requires selling the underlying asset. Thus, if the aggregate gamma of market makers is significantly negative, delta-hedging can give rise to significant net buying, contributing to end-of-day momentum. Symmetrically, if the aggregate gamma imbalance of market makers were positive, delta-hedging would have a stabilizing effect in the form of an end-of-day reversal.[2]

[1] `https://www.wsj.com/articles/the-invisible-forces-exacerbating-market-swings-11582804802`

[2] Derivative markets are by construction in zero net demand since for each option there is a buyer and a seller. Therefore, the overall dollar value of aggregate gamma for each option is zero across all purchasers and sellers. However, some market participants are likely not involved in delta-hedging – think of retail investors or mutual funds – because of different incentives. This implies that the flows arising by delta-hedging activity may not be trivial.

The second channel relates to leveraged ETFs, that is, synthetic instruments that rely on total return swaps, and whose notional principal is a multiple of the value of a referenced index. Different from a standard ETF, a price appreciation of the underlying index has the compounded effect of increasing both the referenced portfolio and the required notional value of the swap. As a consequence, any price change gives rise to an imbalance between the required and effective notional amount of the swap. The swap counterparty has to manage her risk exposure, thus potentially inducing a large rebalancing of the portfolio of physical assets used to hedge the swap (see Section 2.3). Cheng and Madhavan (2010) argue that the portfolio rebalancing of leveraged ETFs may have an impact on intraday prices.

Figure 1 illustrates the rebalancing effects caused by delta-hedgers and leveraged ETFs towards the end of the trading day. The upper panel shows the intraday return path for Tesla's stock on 13 December 2012. At the beginning of the day, the aggregate gamma was positive and economically significant. During the day, Tesla experienced a negative return equal to -6.62% by 15:30. Based on the information available, the gamma imbalance implied that delta-hedgers needed to trade an amount equal to 102.11% of the average dollar trading volume of Tesla shares in the last half-hour. Indeed, a strong price reversal emerged in the last 30 minutes of the trading day, which is consistent with the large initial positive gamma imbalance.

An interesting example that relates to the role of leveraged ETFs is provided by the dynamics of Apple stock on 24 October 2018, see Figure 1 (middle panel). Apple shares are an important constituent for leveraged ETFs. At the same time, the aggregate gamma of delta-hedgers was close to zero at the beginning of the trading day. By 15:30, Apple shares had dropped by -2.24%. As a consequence, leveraged ETF swap counterparties had to sell large quantities of Apple shares to rebalance their portfolios for an estimated amount equal to 8.85% of the average trading volume in Apple shares. Possibly as a result, the price dropped further by -1.22%.

The timing of option delta-hedging and portfolio rebalancing by leveraged ETFs can be different. Figure 1 (bottom panel) illustrates the potential effect of this heterogeneity on the price dynamics. On 23 June 2016, the gamma imbalance of delta-hedgers on Amazon stock options was large and positive. At the market opening, the price dropped by almost 4%. The implied hedging demand by delta-hedgers required purchasing shares for approximately 50% of the average dollar volume in the last half trading hour in Amazon shares. Consistently, the share price started to mean revert at 15:30, albeit not completely. Leveraged ETF swap counterparties had to rebalance, which caused further

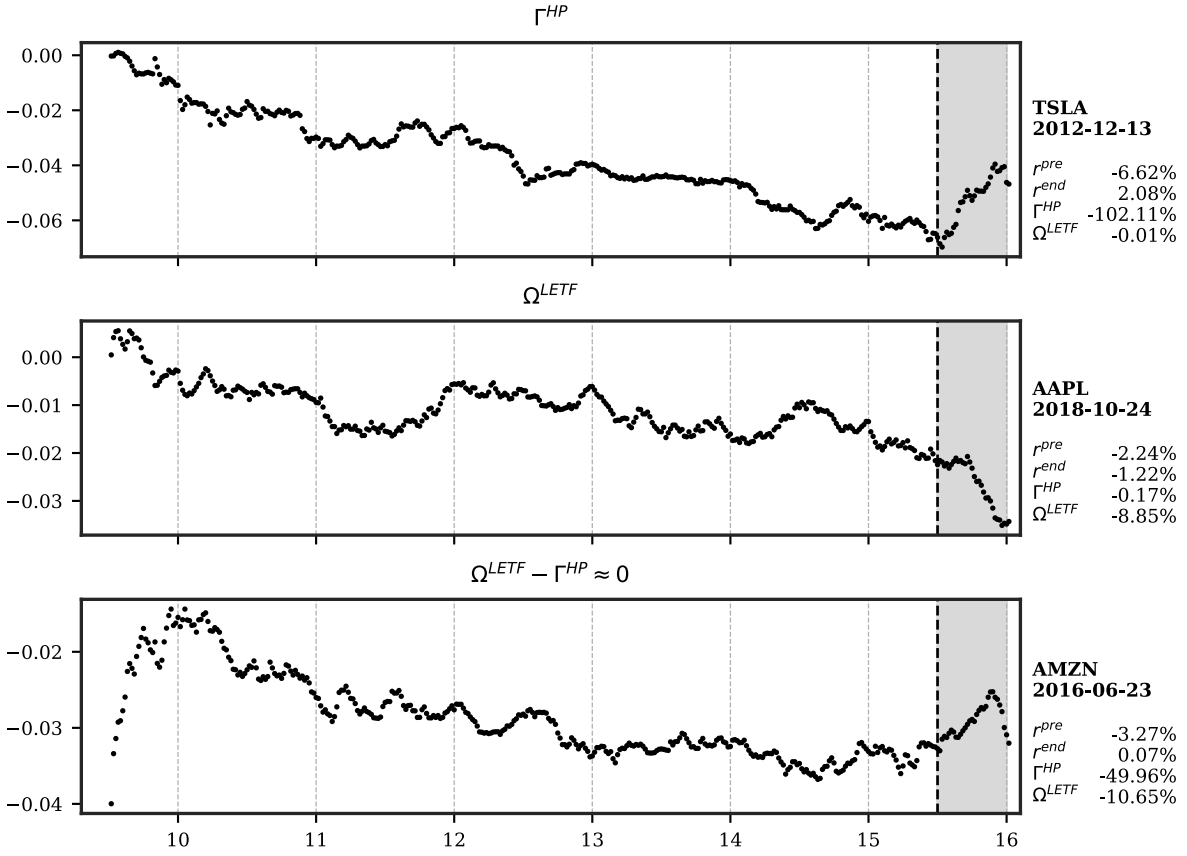


Fig. 1. **Delta-Hedging and Leveraged ETF Rebalancing Effects**
The figure depicts the effects of delta-hedging and leveraged ETF rebalancing on three days in our sample for Tesla (TSLA), Apple (AAPL), and Amazon (AMZN) stock, respectively. r^{pre} denotes the return from the previous day's close price until 15:30. r^{end} denotes the return from 15:30 to close. Γ^{HP} is defined in Equation (5) and is the product of r^{pre} and the aggregate gamma imbalance, Γ^{IB}. Ω^{LETF} is the measure for leveraged ETF rebalancing, defined in Equation (6). Γ^{HP} and Ω^{LETF} are expressed in relative terms to the average dollar trading volume in the last half hour over the last month.

downward pressure as it emerged shortly before the market closing.

We start our analysis by testing whether the above-described price effects have empirical support. To do so, we build a novel dataset merging data from several options exchanges, including the identity of all option counterparties, and the portfolio composition of 72 leveraged ETFs for 24 underlying benchmark ETFs, which represents almost the whole universe of leveraged ETFs on U.S. equity indexes. After computing the gamma imbalance of delta-hedgers and the rebalancing demand of leveraged ETFs, we merge with

NYSE TAQ data to get intraday prices of equity stocks. To the best of our knowledge, this is is the most comprehensive dataset covering options and leveraged ETFs – both in the cross-section and in the time series. Confirming previous findings from the literature (Tuzun, 2014; Shim and Todorov, 2021; Baltussen, Da, Lammers, and Martens, 2021; Ni, Pearson, Poteshman, and White, 2020), we find that both sources of mechanical flows are significantly related to stock returns in the last 30 minutes of the trading day. A standard-deviation increase in Γ depresses end-of-day returns by -113% of the average return in the last thirty minutes of a trading day, while a standard-deviation increase in leveraged ETF rebalancing flows increases end-of-day returns by 430% of the average return in the last half hour. Moreover, the impact of both rebalancing sources is amplified when controlling for the magnitude of the other.

These results are robust to a battery of control variables suggested in the literature, such as the trading volume in the option relative to the equity market (Roll, Schwartz, and Subrahmanyam, 2010), the put/call ratio (Blau, Nguyen, and Whitby, 2014), implied volatility, the riskiness of the underlying stock, and the release of fundamental news.

Next, we turn our attention to the institutional differences between the two channels and their implications. On the one hand, leveraged ETFs are subject to a strict mandate, which requires replicating the underlying index returns at the close multiplied by a leverage factor on a daily basis. This forces leveraged ETF providers to rebalance only a few minutes before the end of the trading day, with little room for discretion. On the other hand, option delta-hedgers enjoy some degree of flexibility with regards to the timing of their rebalancing strategy. Moreover, while leveraged ETF holdings are fully disclosed and constant over a single trading day, the intraday inventory of delta-hedgers potentially changes and is not readily available. These different sets of rules and incentives may generate a heterogeneous degree of information asymmetry regarding flows for the two channels.

An important empirical contribution of our analysis, therefore, relates to the consequences of disclosing order flow information. By exploiting the institutional differences of the two channels, we shed light on the relationship between order flow disclosure and market quality. First, do the transparency of leveraged ETFs and the predictability of their flows lead to enhanced liquidity provision or does it attract predatory trading? How does this affect the speed of reversal of stock prices after the liquidity shocks? Second, how do the effects of the two sources of mechanical order flow evolve over time? Do liquidity providers improve their ability to accommodate shocks over time? And does such a learning process depend on the availability of information on flows?

We begin by investigating how quickly option delta-hedgers actively hedge following a large price shock in the underlying. If the large movement has occurred early during the trading day, we find that the hedging activity is almost immediate. In contrast, the rebalancing activity of leveraged ETFs is unrelated to intraday jumps and takes place almost exclusively at the end-of-day. We note that these results can explain why the regression coefficient on leveraged ETFs flow is almost four times larger than its options-based counterpart. Leveraging on a numerical simulation, we show that the wedge may result from the measurement errors on our proxy for gamma imbalance – arising from the discretion of the rebalancing strategy of delta-hedgers.

Next, we show that the price effects of both channels are on average fully reversed, but with heterogeneous speed. While leveraged ETF-induced price pressure evaporates at the next-day opening, that arising from delta-hedging is significantly longer-lived and requires almost a full trading day to be reversed. We also find that for leveraged ETFs the price impact coefficient is lower for larger flows, while it does not change significantly for gamma-related flows. As in Brøgger (2021), we interpret these findings as evidence of enhanced liquidity provision for the more predictable leveraged ETF-based flows.

Finally, we exploit the length of our sample to analyze the time-series evolution and persistence of the two channels, using rolling-window regressions. The findings clearly indicate that the magnitude of the price effects stemming from leveraged ETFs is declining over time, being barely statistically significant in the last part of the sample. On the contrary, the importance of gamma-induced price effects is persistent throughout the sample and, if anything, has been increasing over time. The high and increasing fragmentation of the U.S. options market potentially causes the difference in the statistical performance over time, as it hinders market participants from constructing a precise measure of the true inventory positions of delta-hedgers.

Overall, these results suggest that liquidity providers can better identify shocks coming from leveraged ETFs, thus incrementally limiting their impact on asset prices over time. On the contrary, the price effects due to gamma imbalances and delta-hedging of options exposures persist throughout the sample, suggesting that liquidity provision is impeded by information asymmetry.

Related Literature

Our work relates to several streams of the literature. The first concerns the role of order flow information disclosure in microstructure dynamics (Madhavan, 2000). A num-

ber of theoretical works suggest that transparency is positively associated with market quality. Forster and George (1992) and Benveniste, Marcus, and Wilhelm (1992) model the effects of anonymity in financial markets and show that spreads and price impact are lower when dealers are partially informed about the direction of liquidity trades. Admati and Pfleiderer (1991) propose a model of sunshine trading, showing that liquidity traders who can pre-announce their order flow enjoy better liquidity provision but increase adverse selection costs for others who cannot. Other models, however, argue in favor of opaqueness. For instance, Madhavan (1995) provides a theoretical setting featuring market segmentation where large institutional investors benefit from non-disclosure of their liquidity trades. The empirical evidence on the subject has not yet settled the issue. Porter and Weaver (1998) study the effects of increased transparency on the Toronto Stock Exchange, concluding that spreads widen after an increase in transparency. Gemmill (1996), analyzes the consequences of an increase in post-trade transparency on the London Stock Exchange, finding no significant impact on market liquidity. By exploiting the heterogeneous impact of two sources of mechanical intraday rebalancing, our results shed new light on the issue, providing evidence that order flow transparency leads to better liquidity provision. We argue that, as the institutional mandate of leveraged ETFs forces them to act under full transparency, they collectively enjoy a positive equilibrium with lower spreads during their rebalancing operations. Conversely, the freedom of options delta-hedgers to partially control their rebalancing strategy generates a less desirable equilibrium with higher transaction costs and reduced price efficiency.

The second stream studies the feedback effects of options trading on the underlying stock price dynamics. The literature generally distinguishes between two channels through which options trading may have an impact on the price of the underlying. Hu (2014) provides evidence that the information found in market makers' initial delta-hedges can significantly affect the price dynamics of the underlying.[3] However, a non-informational channel may also be at work: Ni, Pearson, and Poteshman (2005) and Golez and Jackwerth (2012) document that rebalancing and unwinding of option market makers' delta hedges on or very close to expiration drive the prices of individual stocks and stock index futures towards option strike prices on option expiration dates. Lately, Ni et al. (2020) analyze the effects of Γ-imbalance on absolute returns and the autocorrelation of returns, based on theoretical models that predict a negative relation between

[3]Other studies advocating an informational channel are, among others, Easley, O'Hara, and Srinivas (1998), Pan and Poteshman (2006), Ni, Pan, and Poteshman (2008), Cremers and Weinbaum (2010), Roll et al. (2010), Johnson and So (2012), and Ge, Lin, and Pearson (2016).

stock volatility and Γ-imbalance.[4] Whereas Ni et al. (2020) resort to daily data, Barbon and Buraschi (2020) concentrate on intraday price dynamics. They find that Γ-imbalance is negatively related to intraday volatility and document that Γ-imbalance can affect the frequency and magnitude of flash crashes. Baltussen et al. (2021) show that end-of-day momentum in many futures contracts concentrates on days with negative Γ-exposure of option market makers. Finally, Chordia, Kurov, Muravyev, and Subrahmanyam (2021) propose a risk-based channel. The authors show that net buying pressure in index puts on the International Securities Exchange positively predicts subsequent S&P 500 index returns and trace the predictability to the purchase of protection when uncertainty is high.

A different, but related stream of the literature studies the effects of the inventory of option market makers. Gârleanu, Pedersen, and Poteshman (2009) have provided path-breaking work on how demand pressure affects option prices. A closely related study is by Fournier and Jacobs (2020). Johnson, Liang, and Liu (2016) investigate the forces behind the use of S&P 500 index options and conclude that unspanned crash risk drives much of their demand. Relatedly, Jacobs and Mai (2020) find a tight link between prices and demand in S&P 500 and VIX options. Chen, Joslin, and Ni (2019) infer financial intermediary constraints via deep out-of-the-money index put options. The authors show that a tightening of intermediary constraints is accompanied by option expensiveness and broker-dealer deleveraging.

A third stream focuses on the effects of (leveraged) ETF ownership on the constituent stocks. Ben-David, Franzoni, and Moussawi (2018) show that stocks with higher ETF ownership exhibit higher volatility, as liquidity shocks caused by short-horizon traders in the ETF can be transmitted to the underlying stocks by an arbitrage mechanism. Shum, Hejazi, Haryanto, and Rodier (2016) show that the rebalancing flows of leveraged ETFs amplify end-of-day volatility in the period from 2006 to 2011.

Another part of the literature studies intraday return patterns. We find high-frequency return continuation in the cross-section of stock returns, consistent with the evidence provided by Gao, Han, Zhengzi Li, and Zhou (2018) and Baltussen et al. (2021). Both studies focus on aggregate investment vehicles, such as ETFs and index Futures. Gao et al. (2018) show that their effects are stronger on days with elevated volatility, which are typically also accompanied by higher trading volume. In the cross-section of stocks, we confirm the finding of Komarov (2017) that stocks performing best in the first half

[4]For a theoretical foundation, see among others Frey and Stremme (1997), Frey (1998), Sircar and Papanicolaou (1998), Platen and Schweizer (1998), Wilmott and Schönbucher (2000).

of the day will likely lose in the second half if controlled for market returns. Another study on short-term return reversals is Heston, Korajczyk, and Sadka (2010). The authors show that the returns of a half-hour period have predictive power over the same half-hour periods for up to 40 days in the future when controlling for the impact of the market. They relate this to the usage of trade mechanisms by institutional traders, designed to limit the relative price impact of their orders. More recently, studies link investor heterogeneity on the stock level to cross-sectional intraday and overnight return variations. Lou, Polk, and Skouras (2019) hypothesize that different investor types trade predominately at different times throughout the trading day. Empirically, the authors document high persistence in overnight and intraday return components, which they find not on a single-stock basis, but also for 14 equity strategies such as size, value or profitability. Bogousslavsky (2020) focuses only on intraday returns and finds that a mispricing factor earns positive returns up to the last half hour, consistent with the idea that arbitrageurs trading on mispricing reduce their positions at the end of the trading day.

Finally, the last stream focuses on the U.S. equity market closing auction. Bogousslavsky and Muravyev (2020) show that the share of daily volume in the closing auction has more than doubled from 2010 to 2018. They attribute the increase in trading volume to the rise of indexing and ETFs. Wu and Jegadeesh (2020) examine the price impact, including ts temporary component, of closing auctions. Trading strategies based on market-on-close imbalances generate outsized returns.

2. Data and Measurements

To conduct our empirical analysis we source data from several databases, including information for single stocks, single-name options, and leveraged ETFs. By merging these data sources we obtain a unique dataset allowing us to measure flows coming from the hedging of options and rebalancing of leveraged ETFs to study their potential impact on stock prices.

2.1. Data Sources

Options Market. The first dataset merges option data from five different exchanges: (a) the CBOE C1 exchange, (b) NASDAQ GEMX (GEMX), (c) NASDAQ International Security Exchange (ISE), (d) NASDAQ Options Market (NOM), and (e) NASDAQ PHLX (PHLX). The dataset includes information on signed trading volume, the underlying

stock, and the category of the counterparties engaged in the trade.

Each of the five exchanges provides four categories of volume for each option series: open buy, open sell, close buy, and close sell. Each category is further broken down into different types of market participants: *broker/dealer*, *proprietary*, and *customer*.[5] For each type of market participant, we sum the buy and sell trades to estimate the long and short open interest at the trader-type level. The five exchanges sum up to a substantial proportion of the equity options market, delivering the most comprehensive coverage available at the moment. Nonetheless, we do not cover volume outside of these exchanges and OTC options trading. We also gather daily bid and ask quotes, implied volatility, trading volume, open interest, and Greeks for each option contract from OptionMetrics. As individual stock options are of American type, OptionMetrics uses binomial trees to compute implied volatility and Greeks.

Leveraged ETFs. We obtain information on all leveraged ETFs on U.S. equity indexes from ETFGlobal including the leverage amount, the benchmark index referenced by the fund and the assets under management of each leveraged ETF at the daily frequency. We compute the constituents of the benchmark index of each leveraged ETF and use TAQ data to calculate intraday returns of each selected benchmark referenced by the leveraged ETF.

Table C.1 in the Online Appendix provides an overview of the properties of leveraged ETFs included in our sample. On average, we consider 72 leveraged ETFs for 24 underlying benchmark indices, with a cross-sectional distribution that is fairly stable over time. On average, 45% of the funds we consider are inverse or bear funds. Weighted by the AUM (VW), this number drops to 33%, but fluctuates substantially over time, with a proportion of just 16% at the 10th percentile and 63% at the 90th. The average fund is leveraged by an absolute value of 2.35.

Stock Market. Information on individual equity stocks is obtained from the Center for Research on Security Prices (CRSP) and includes trading volume, shares outstanding, and closing prices. We restrict our analysis to stocks with CRSP share codes 10 and 11, and exchange codes 1, 2, 3, 31, 32, and 33. Information on any type of distribution (e.g. dividends and stock splits) is also obtained from CRSP. We match data from CRSP with our options data via the matching algorithm provided by WRDS.

High-frequency Data on Underlying Assets. Intraday stock price data and transaction volumes are obtained from TAQ. We use standard cleaning procedures and match intraday

[5]In 2009, the type of *professional customer* has been introduced alongside the customer. We merge *professional customers* with *customers*.

trade prices with CRSP to obtain PERMNOs as unique identifiers. More details are given in Online Appendix A. Equipped with intraday prices, we calculate intraday returns relative to the previous day's closing price. Following standard practice in the literature, (see Lou et al., 2019), we assume that corporate events that mechanically impact prices, e.g. dividend payments and stock splits, take place overnight and are realized at the time of the first trade on the target date. If a delisting occurs as reported by CRSP, we assume that the delisting amount is realized at the respective day's close.

To summarize, our sample across optionable stocks and leveraged ETFs comprises 1,882,332 stock-day observations from January 2012 through December 2019.[6]

2.2. Measuring Gamma Hedging Pressure

Let $V(t,S)$ denote the value of an option contract and $\Delta(t,S) = \frac{\partial V(t,S)}{\partial S}$ be the first derivative of the option price with respect to the underlying, whereas $\Gamma(t,S) = \frac{\partial^2 V(t,S)}{\partial S^2}$ measures the change of $\Delta(t,S)$ for changes in S. When $\Gamma(t,S)$ is not zero, $\Delta(t,S)$ changes depending on time to maturity and the level of S and, consequently, any hedging position has to be adjusted periodically. If $\Gamma(t,S)$ is large in absolute terms, $\Delta(t,S)$ is very sensitive to movements in the underlying and it implies a large amount of rebalancing for the market maker to remain delta-neutral.

Since option market makers and broker/dealers have similar hedging incentives, we classify both as "delta-hedgers". Consequently, we categorize proprietary and customers as non-delta-hedgers and refer to them jointly as "end-customers" in the remainder of the paper.

To obtain the gamma imbalance of delta-hedgers, we proceed as follows. Let $OI_{o,t}^{Buy,I}$ be the open interest of investors of type I in long positions in option o at time t, which is related to daily volume as follows (see Ni et al., 2020):

$$OI_{o,t}^{Buy,I} = OI_{o,t-1}^{Buy,I} + Volume_{o,t}^{OpenBuy,I} - Volume_{o,t}^{CloseSell,I} \tag{1}$$

$$OI_{o,t}^{Sell,I} = OI_{o,t-1}^{Sell,I} + Volume_{o,t}^{OpenSell,I} - Volume_{o,t}^{CloseBuy,I}, \tag{2}$$

where $Volume_{o,t}^{OpenBuy,I}$ and $Volume_{o,t}^{OpenSell,I}$ denote the volume from investors type I to open new long and short option positions, and $Volume_{o,t}^{CloseBuy,I}$ and $Volume_{o,t}^{CloseSell,I}$

[6]To appear in the sample, each stock has to be optionable and included in one or more leveraged ETFs.

refer to volumes with which investors of type I closed existing long and short positions, respectively.

Second, we calculate the delta-hedgers' net open interest in option series o at date t's close as

$$netOI_{o,t} = -\left[OI_{o,t}^{Buy,Customer} - OI_{o,t}^{Sell,Customer} + OI_{o,t}^{Buy,Proprietary} - OI_{o,t}^{Sell,Proprietary}\right], \tag{3}$$

where $netOI_{o,t}$ is measured in units of option contracts and $OI_{o,t}^{D,I}$ is the open interest of direction D (either buy or sell) by market participant type I (either customer or proprietary) in option series o at the close of date t. The net open interest of market makers is the opposite of the sum of the remaining market participant types. Assuming that broker/dealers are also delta-hedgers yields Equation (3).

Let $\Gamma_o(t, S)$ denote the gamma of option series o on stock j at day t and underlying price S, expressed in shares of the underlying.[7] To compute the day t delta-hedger dollar gamma imbalance in option series o, we take the product of $\Gamma_o(t, S)$ with the stock price S at 15:30 on the target day and multiply by the contract multiplier $Mult_o$ of o (typically $Mult_o = 100$).[8]

To obtain the aggregated gamma imbalance on an underlying stock j for trading day t, denoted by $\Gamma_{j,t}^{IB}$, we compute the sum over all options on the underlying:

$$\Gamma_{j,t}^{IB} = \underbrace{\left(\sum_o netOI_{o,t-1} \times \Gamma_o(t-1, S_{j,t-1}^{close}) \times S_{j,t}^{15:30} \times Mult_o\right)}_{(\star)} \times \frac{S_{j,t-1}^{close}}{100} \times \frac{1}{\text{ADV}_{j,t-1}^{\text{end}}}. \tag{4}$$

The term $(\star)$ in Equation (4) denotes the total dollar gamma imbalance for a given stock at day t. It is the dollar amount delta-hedgers need to trade in the underlying for each one-dollar move in the underlying stock price S. By multiplying $(\star)$ by the underlying price divided by 100, we obtain the dollar gamma imbalance for a one percent move in S. This facilitates comparison over time and in the cross-section. Finally, we scale by the average dollar volume in the last half hour of a trading day, $\text{ADV}_{t-1}^{\text{end}}$, computed over the last month. Thereby, we express the delta-hedgers' gamma imbalance in a given stock as a fraction of the typical trading taking place in the last half hour, which allows

[7] OptionMetrics calculates the gamma of an option as the absolute change in delta given a \$1.00 change in the underlying.

[8] Since we cannot observe intraday variations of the net open interest, we assume that $\Gamma(t, S)$ only changes due to innovations in the stock price. We hence use the observed stock price at 15:30 on day t to compute the best possible estimate of the amount to be traded by delta-hedgers.

us to obtain a timely proxy for the potential price impact of hedging adjustments.

$\Gamma^{IB}_{j,t}$ denotes the amount of hedging market makers would have to do for a 1%-move in the underlying stock j at time t. We combine this measure with information on how much the underlying has moved before the start of the hedging window. The return from the close of day $t-1$ to the start of the hedging window at 15:30 is denoted as $r^{\text{pre}}_{j,t}$. The percentage hedging pressure is thus the interaction between $\Gamma^{IB}_{j,t}$ and $r^{\text{pre}}_{j,t}$:

$$\Gamma^{HP}_{j,t} = 100 \times \Gamma^{IB}_{j,t} \times r^{\text{pre}}_{j,t}. \tag{5}$$

Γ^{HP} is our main variable of interest. It directly captures the amount of hedging required for delta-hedgers to remain delta-neutral after observing intraday return r^{pre}.

2.3. Measuring Leveraged ETF Rebalancing Pressure

To understand how leveraged ETFs may contribute to intraday price momentum, let L be the fund's leverage factor, say -3 for a bear and $+3$ for a bull fund, and A_t the leveraged ETF's assets under management (AUM). The required notional amount of total return swaps at day t is $S_t = L \times A_t$. Hence, if the return on the index at time $t+1$ is r^{bench}_{t+1}, then $A_{t+1} = A_t \times (1 + L \times r^{bench}_{t+1})$. Therefore, the required notional amount of total return swaps at $t+1$ becomes $S_{t+1} = L \times A_{t+1} = L \times A_t \times (1 + L \times r^{bench}_{t+1})$. However, the actual exposure of the total return swaps at $t+1$ is $E_{t+1} = S_t \times (1 + r^{bench}_{t+1}) = L \times A_t \times (1 + r^{bench}_{t+1})$. The difference between the required and the actual notional is the rebalancing amount, equal to $S_{t+1} - E_{t+1} = L \times (L-1) \times A_t \times r^{bench}_{t+1}$. Given that the required hedging multiple $L \times (L-1)$ is strictly positive for $L \in \mathbb{R} \setminus [0, 1]$, leveraged ETF swap counterparties always have to trade in the same direction as the return of the underlying index, which may induce an end-of-day return momentum effect in the stocks included in the index referenced by the leveraged ETF.[9] Unlike the gamma imbalance effect, hedging demand by leveraged ETFs for a given stock is not necessarily proportional to the return of that specific stock but, rather, to the return of the underlying index. Therefore, leveraged ETFs rebalancing may induce return momentum in a stock, even if its return in the first part of the trading day is zero.

To compute the amount of rebalancing affecting an individual stock, let stock j be

[9] We assume that returns swap providers (or their counterparties) ultimately need to hold a quantity of the underlying stock proportional to the size of the swap contract. If some of the swaps are based on other instruments or on correlated assets, this would introduce noise in our proxy and bias our results toward zero.

included in the underlying of a leveraged ETF i with a weight of $w_{i,j,t}$ on day t. If the swap counterparty starts rebalancing their exposure at 15:30, and $r^{\text{pre}}_{i,t;bench}$ denotes the return on the benchmark ETF up until that moment, the relative amount of rebalancing required in stock j is the sum over all leveraged ETFs in which stock j is included in:

$$\Omega^{LETF}_{j,t} = \frac{\sum_{i=1}^{N_{i,t}} L_i \times (L_i - 1) \times A_{i,t-1} \times w_{i,j,t-1} \times r^{\text{pre}}_{i,t;bench}}{\text{ADV}^{\text{end}}_{j,t-1}} \tag{6}$$

We scale by the average dollar volume in the last half hour ($\text{ADV}^{\text{end}}_{t-1}$) to compare the impact of rebalancing with the average amount of trading at the end of the trading day.

2.4. Summary Statistics

In this section, we provide summary statistics for the gamma hedging pressure and leveraged ETF rebalancing quantity.

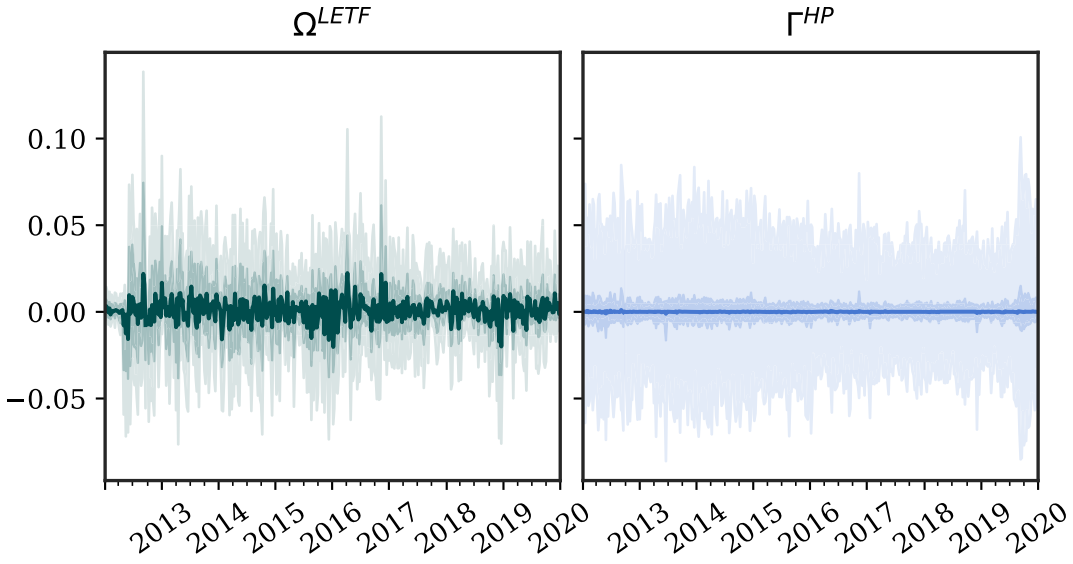


Fig. 2. **Time Series Cross-sectional Distribution of Ω^{LETF} and Γ^{HP}**
This figure shows weekly averages of the cross-sectional distribution of the demand pressure from the rebalancing of leveraged ETFs (Ω^{LETF}) and Gamma imbalance (Γ^{HP}). The dark line represents the cross-sectional median, the dark-colored area the 25th/75th percentile, and the light-colored area the 5th/95th percentile.

Figure 2 shows the cross-sectional distribution of Ω^{LETF} over time and compares it with the distribution of Γ^{HP}. While both are time-varying, the distribution of Γ^{HP} is

wider, with a larger interquartile range, suggesting that on the same day delta-hedgers can have a large negative gamma imbalance on some stocks and, at the same time, a large positive gamma imbalance on another stock. Table A.1 in the Appendix provides more granular summary statistics on Γ^{HP} and Ω^{LETF}. Table A.1 shows that the aggregate pressure arising from leveraged ETFs is on average higher than the one from delta-hedging, but the distribution of absolute Γ^{HP} has a heavy right tail.

Table 1: **The Cross-Section of Demand Pressure from Gamma and Leveraged ETF Rebalancing**
This table reports the average daily number (N) and percentage (Share) of stocks with a combined rebalancing amount in the last 30 minutes of the trading day exceeding a certain threshold (first column), as a percentage of the average dollar volume. The fourth column (% Gamma) reports the proportion of the total demand pressure due to the Gamma imbalance Γ^{HP} relative to the combined rebalancing amount: $|\Gamma^{HP}|/(|\Gamma^{HP}| + |\Omega^{LETF}|)$. Additionally, the table reports the average mean (Mean), standard deviation (Std), and 10%- and 90%-percentile of the share of the combined rebalancing amount conditional on exceeding the specified level. The sample contains stocks that are optionable and included in the benchmark index of at least one leveraged ETF. The sample period is from January 2012 – December 2019.

	N	Share	% Gamma	Mean	Std	10%	90%
1%	601.7	60.23%	49.95%	5.11	8.29	1.30	9.81
2%	354.4	35.48%	55.76%	7.68	10.41	2.86	13.67
5%	182.6	18.28%	63.01%	12.07	13.72	5.47	20.28
10%	69.5	6.95%	71.86%	21.23	19.63	10.83	34.30
15%	34.0	3.41%	76.47%	30.50	24.34	16.43	48.97
25%	12.5	1.25%	80.91%	49.00	31.88	28.92	76.72
50%	3.4	0.34%	84.75%	94.33	44.86	72.69	122.20

Table 1 documents the economic significance of the demand pressure arising from the rebalancing activity of the leveraged ETFs and the delta hedging activity in the options market. The second column reports the average number of stocks in the joint sample for which the combined dollar rebalancing amount due to both channels exceeds a certain threshold of the average dollar volume in the last 30 minutes of the trading day (ADV^{end}), depicted in the first column. We find that, on average, for 69 (12) stocks the combined rebalancing amount exceeds the 10% (25%) threshold. For these stocks, the average rebalancing dollar amount is equal to 21.23% (49.00%) of ADV^{end}. The cross-sectional dispersion is large and for stocks in the 90th percentile (last column) the average rebalancing dollar amount exceeds 34.30% (76.72%) of the total end-of-day trading volume. The fourth column provides information about the relative importance

of each channel. The majority of the rebalancing amount is driven by Γ^{HP}. For the group exceeding a rebalancing threshold of 10% (25%), 71.86% (80.91%) of the average absolute combined rebalancing originates from the hedging activity of option market makers. Moreover, the share of Γ^{HP} in the combined rebalancing amount is monotonically increasing in the threshold level, suggesting that the demand pressure from this channel dominates for large total rebalancing flows.

2.5. Abnormal Order Flow

We define the abnormal order flow in the last 30 minutes of trading day t for stock j as

$$\mathrm{RSVOL}^{\mathrm{end}}_{j,t} = \mathrm{SVOL}^{\mathrm{end}}_{j,t} / \mathrm{ADV}^{\mathrm{end}}_{j,t-1}\,. \tag{7}$$

$\mathrm{SVOL}^{\mathrm{end}}_{j,t}$ is the difference between the trading volumes in up- and down-minutes, defined as

$$\mathrm{SVOL}^{end}_{j,t} = \sum_{m \in 15{:}30 \to Close} \mathrm{VOL}_{j,t,m} \times \mathbb{1}_{r_{j,t,m}>0} \quad - \sum_{m \in 15{:}30 \to Close} \mathrm{VOL}_{j,t,m} \times \mathbb{1}_{r_{j,t,m}<0}, \tag{8}$$

where m denotes the minutes within the hedging window for target day t.[10]

3. Empirical Results

3.1. Main Results

Consider the case in which $\Gamma^{IB}_{j,t}$ for stock j at time t is negative. If delta-hedgers want to maintain delta-neutrality, they have to sell the underlying stock if it has depreciated intraday, leading to additional momentum. Instead, for a positive $\Gamma^{IB}_{j,t}$, the delta-hedgers have to buy the underlying stock, trading against its previous price change, which gives rise to intraday reversals. Following Equation (6), rebalancing of the swap portfolios replicating the exposure of leveraged ETFs will always be in the same direction as the previous movement of the benchmark ETF, potentially leading to momentum for a large number of stocks. Another important difference between the two rebalancing sources is in the cross-section. Since Γ^{HP} can greatly differ across stocks even on the same day t, the effect is inherently stock-specific. For Ω^{LETF} in contrast, all stocks that belong to

[10] We scale by $\mathrm{ADV}^{\mathrm{end}}_{t-1}$ to make the signed volume comparable across stocks and over time.

the same benchmark index are exposed to the same price pressure, proportional to their weighting in the index.

In Table 2, we investigate the impact of the rebalancing activity of delta-hedgers and leveraged ETF swap counterparties on the order-flow and returns of individual stocks in the last half hour before the close. To do so, we employ two types of panel regressions:

$$\text{RSVOL}_{j,t}^{\text{end}} = \beta_0 r_{j,t}^{\text{pre}} + \beta_1 \Gamma_{j,t}^{HP} + \beta_2 \Omega_{j,t}^{LETF} + FE_j + FE_t + \epsilon_{j,t}, \tag{9}$$

$$r_{j,t}^{\text{end}} = \beta_0 r_{j,t}^{\text{pre}} + \beta_1 \Gamma_{j,t}^{HP} + \beta_2 \Omega_{j,t}^{LETF} + FE_j + FE_t + \epsilon_{j,t}. \tag{10}$$

including asset (FE_j) and date fixed effects (FE_t).[11] Standard errors are double-clustered by day and asset, and return observations weighted by the stock's market capitalization in the previous month to rule out that our results are driven by microstructure issues in small stocks.

Table 2: **Impact of Delta-Hedging and Leveraged ETF Rebalancing**
The table reports the results to regressing the returns in the last half hour of a trading day on returns until 15:30 (r^{pre}), on option hedging pressure Γ^{HP} and leveraged ETF rebalancing quantity Ω^{LETF} following Equation (9). T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity-fixed effects in all specifications and value-weight observations. The sample period is 2012 – December 2019.

	(1)	(2)	(3)	(4)	(5)
Dependent	$\text{RSVOL}_t^{\text{end}}$	r_t^{end}	$\text{RSVOL}_t^{\text{end}}$	r_t^{end}	r_t^{end}
Γ_t^{HP}	-9.438**	-9.464***			-10.960***
	(-2.116)	(-4.709)			(-5.521)
Ω_t^{LETF}			109.572***	43.394***	46.738***
			(5.450)	(4.797)	(5.198)
r_t^{pre}	-1.263***	-0.757***	-1.512***	-0.902***	-0.820***
	(-5.111)	(-6.907)	(-5.640)	(-8.139)	(-7.452)
Observations	1,882,332	1,882,332	1,882,332	1,882,332	1,882,332
R^2 (%)	0.141	0.284	0.234	0.325	0.370
Entity FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

[11] In Appendix C, we add control variables $\boldsymbol{X}_{j,t}$ to Equation (9) to rule out that neither informational channels nor risk-based explanations are driving our results.

Gamma Hedging Pressure Column (1) shows that Γ^{HP} negatively impacts $\text{RSVOL}_t^{\text{end}}$ with a coefficient of -9.44 (t-value: -2.12), which establishes a significant link between the gamma-induced trading activity of delta-hedgers and end-of-day order flow. At the same time, this additional directed volume has a direct impact on end-of-day returns (column 2). Specifically, the slope coefficient of Γ^{HP} is negative and highly significant at -9.46 (t-value: -4.71), suggesting that a negative shock to Γ^{HP} (additional buying pressure) amplifies end-of-day returns. We furthermore control for the impact of other momentum-based intraday effects by including the previous return from last day's close to the start of the supposed hedging window at 15:30. While the estimated slope coefficient is negative and significant, its inclusion does not materially change the estimated impact of Γ^{HP} on end-of-day returns.

In Online Appendix G, we mitigate concerns that investors might trade options based on past information which is correlated with future intraday returns. Our results remain quantitatively and qualitatively the same when replacing Γ^{HP} with an estimate based on older option inventories of delta-hedgers following Ni et al. (2020).

Leveraged ETFs If the rebalancing pressure from leveraged ETFs is sufficiently large, we should find that it positively relates to end-of-day order flow, as well as returns. Column (3) shows how a positive shock to Ω^{LETF} clearly leads to additional buying pressure at the close. The slope coefficient is estimated at 109.57 and is highly significant (t-value: 5.45). The ensuing buying pressure also translates to higher end-of-day returns with a coefficient of 43.40 (t-value: 4.80), as shown in column (4).

Joint Rebalancing The two rebalancing channels may affect closing stock returns in different ways. In the example reported in the Introduction, for instance, an investor considering only gamma hedging flows would have predicted positive returns for Amazon on June 23rd, 2016, given the large and negative magnitude of Γ^{HP}. Had she also known about the significant amount of expected selling from leveraged ETFs, she may have revised her prediction.

In column (5) of Table 2 we estimate a model accounting for the two channels jointly, resulting in increases of both coefficients in absolute magnitude to -10.96 for Γ^{HP} and 47.74 for Ω^{LETF}, both of which are significant at the 1%-level. A one standard deviation decrease in Γ^{HP} is associated with a 113% increase in the magnitude of end-of-day returns relative to the stock-level mean return during that period, while a standard deviation increase in Ω^{LETF} corresponds to an increase of 430%.

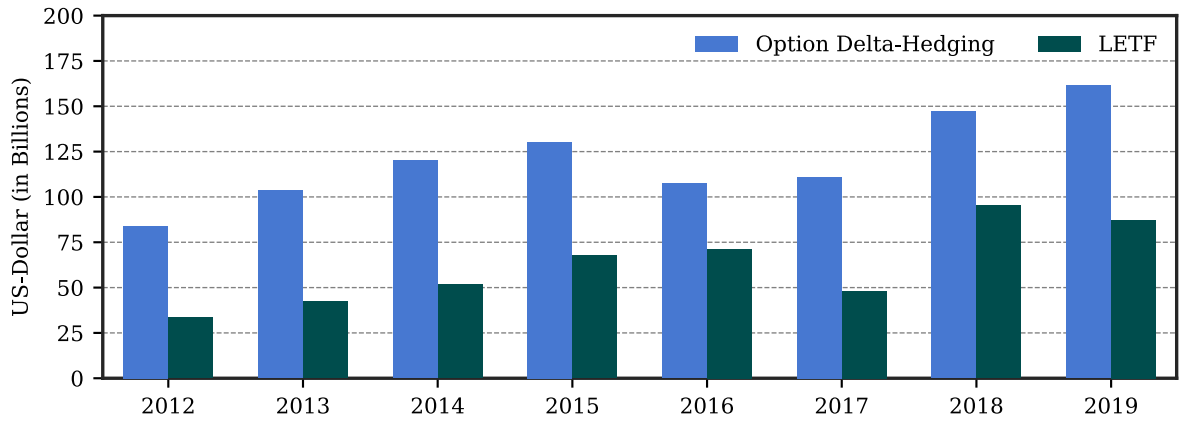


Fig. 3. **Year-by-Year Rebalancing Flows**
The figure shows the estimated overall absolute rebalancing flows from leveraged ETFs and option delta-hedging for each year in the sample in billion USD.

We find that a shock to the buying pressure coming from leveraged ETF rebalancing flows is estimated to impact end-of-day returns of a target stock two to four times as much as a similarly-sized shock from option delta-hedgers. But how do the estimated rebalancing amounts stack up when considered across all stocks in the universe of optionable stocks, which are also included in leveraged ETFs? Figure 3 shows the absolute year-by-year rebalancing amount in billion USD, summed across the entire cross-section of stocks in our sample. For each year, the total amount that would have to be traded in the underlying stocks to assure delta-neutrality exceeds that required by leveraged ETF counterparties. In fact, in most years the total rebalancing flows from delta-hedgers are at least twice as large. We find that the total trading amount for both rebalancing sources has increased over time, from approximately 80 billion USD in 2012 for Γ^{HP} to 160 billion USD in 2019 – a twofold increase. This growth has two reasons: first, the general growth in options trading. Second, the staggered introduction of weekly options on single equities, which, due to their short lifespan and corresponding large gamma values, have a large impact on the overall gamma position of delta-hedgers, potentially prompting larger rebalancing trading. The total amount of rebalancing from leveraged ETFs has increased even more dramatically, starting at 35 billion USD in 2012 and growing to 85 billion USD in 2019, which mimics the rapid growth of the industry.

Information Quality of Rebalancing Proxies Why do gamma hedging flows have a seemingly smaller impact on end-of-day order flow and returns, despite being larger in terms of USD traded? It is well known that noise in the explanatory variable biases

regression estimates towards zero, which potentially explains the seemingly smaller one-to-one impact of gamma rebalancing. In this section, we investigate the role of uncertainty attached to acquiring the information necessary to determine the amount of rebalancing flows from option delta-hedgers. As a counterpoint, consider the estimation of the rebalancing pressure from leveraged ETFs in Equation (6). To quantify Ω^{LETF} we need to know the leverage factor, the assets-under-management, the return on the benchmark ETF before hedging begins, and the weight of the target stock in the benchmark ETF. All of these inputs are readily observable. Furthermore, leveraged ETFs are mandated to provide a daily multiple exposure to the benchmark, forcing their swap counterparties to trade near the market close. Consequently, all necessary information to determine their rebalancing activity is observed with little uncertainty.

In contrast, the information used to quantify the total amount of rebalancing from delta-hedgers is more uncertain for a number of reasons: first, we only observe approximately 50-70% of the transacted volume in equity options. In fact, the fragmented exchange landscape of equity options trading contributes to the opacity of rebalancing flows. Only a fraction of exchanges reports on detailed volume data by type of market participant, while the remainder is invariably unobservable. Second, delta-hedgers are not explicitly required to rebalance at the close, but may do so at different times during the day, or even hold on to some directional risk overnight. Third, the inventory data provides only an end-of-day snapshot of the delta-hedgers portfolio, which disregards intraday changes in the gamma imbalance.

To understand the impact of information uncertainty on the estimation of how Γ^{HP} influences end-of-day returns, we resort to a simulation study, outlined in detail in Appendix B. Figure 4 shows the results for two levels of $\pi \in [10\%, 50\%]$, which denotes the probability that delta-hedgers rebalance at a different time than between 15:30 and 16:00. We also vary the amount of unobserved fluctuation $\sigma(\Delta)$ in the empirically observed gamma imbalance, relative to the fluctuation of the true value, which addresses the influence of intraday shifts in the delta-hedger's portfolio, as well as the unobserved portion of the total option inventory.

Given the evidence in Figure 3 that the cumulative amount of rebalancing is much larger for Γ^{HP} than it is for Ω^{LETF}, we set the true impact of Γ^{HP} on end-of-day returns to the empirical estimate for Ω^{LETF} of -41 as a conservative estimate in the simulation. This true value is shown by the green line. We also highlight our empirical estimate of around -11. The blue line denotes the average estimated coefficient for 1,000 simulations over 2,000 time steps, the blue-shaded area the 95% confidence bounds.

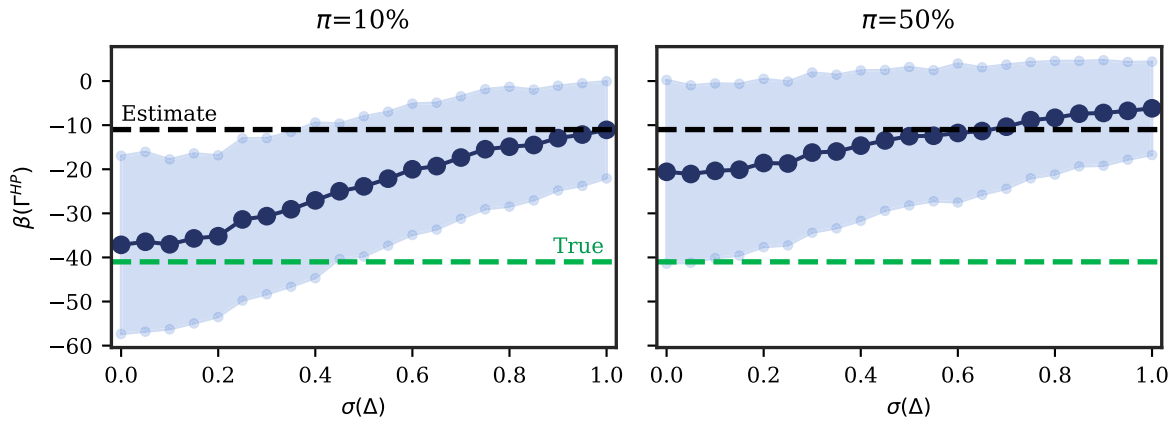


Fig. 4. **Simulating Error Sources in Γ^{HP}**
The figure shows the impact of unobserved fluctuation in the empirical gamma estimate $\sigma^(\Delta)$ and the probability with which option delta-hedgers may decide to rebalance their exposure at a different time than at the close, denoted as π. The setup of the simulation study is outlined in Appendix B. The green line shows the true slope coefficient of Γ^{HP} in the simulation, which we fix at -41, the black line highlights our empirical estimate from Table 2. The blue line denotes the average estimated coefficient for 1,000 simulations over 2,000 time steps, the blue-shaded area the 95% confidence bounds.

For a low level of $\pi = 10\%$ we find that we would require a very high level of $\sigma(\Delta)$, about as high as the fluctuation of its true value. If we increase the probability that delta-hedgers establish delta-neutrality of their positions at their own discretion and may deviate from the supposed hedging window at the close to 50%, we find average coefficients much closer to our empirical estimate of ≈ -11. In fact, an uncertainty level $\sigma(\Delta)$ of around 0.6 generates the empirically-observed impact of Γ^{HP} on end-of-day returns. In the following section, we empirically investigate the level of discretion of delta-hedgers, which can severely impede the estimation of its true impact as shown in the simulation.

3.2. Discretionary Rebalancing Times

Intraday jumps potentially present occasions where delta-hedgers are incentivized to deviate from end-of-day hedging. Consequently, we focus on the price effects of rebalancing activities following large price movements. We detect price jumps for the underlying stocks in the case of Γ^{HP} and the benchmark ETF for Ω^{LETF}. For this, we compare the return in each 30-minute interval of a given trading day with the return distribution of the same interval over the last year. We denote stock intervals where the returns exceed the top or bottom 2.5th percentile as jump events. We then record the cumulative return from the previous day close up to the end of the interval when the jump has occurred

($r_t^{\text{incl. jump}}$) and relate it to the return in the subsequent 30-minute interval (r_t^{next}), and to Γ^{HP} and Ω^{LETF} computed using $r_t^{\text{incl. jump}}$. In Table 3 we separately report results for jump events occurring either before noon or between noon and 15:00.

As evident in column (1), we find large and significantly negative coefficients for jumps before noon in the case of Γ^{HP} (-44.42, t-value: -1.98). This suggests that option delta-hedgers rebalance their exposure shortly after particularly large price movements. We do not find any effect for jumps in the second part of the trading day. In stark contrast to these results, the impact for leveraged ETF rebalancing is insignificant and economically negligible at all times, regardless of whether we consider jumps before or after noon (columns (3) and (4)).

To further investigate how option delta-hedgers exploit the discretionary nature of their re-hedging needs, we focus on the gamma imbalance accumulated only after a jump event. For this, we assume the delta-hedger hedges her accrued exposure right after the jump, and hedges the remainder once more at the close. The remaining exposure consequently amounts to

$$\Gamma_{j,t}^{HP} = 100 \times \Gamma_{j,t-1}^{IB} \times r_{j,t}^{\text{after jump}}, \tag{11}$$

where $r^{\text{after jumps}}$ measures the return of the underlying after the identified jump period until 15:30.

The results reported in columns (5) and (6) in Table 3 provide evidence of end-of-day rebalancing on identified jump-days only for jumps that happen after noon. In fact, the pattern of significance and the magnitude of coefficients inversely match the pattern for the impact of hedging directly after large price movements. Taken together, we conclude that delta-hedgers tend to rapidly rebalance after early jumps, but decide to wait to re-hedge after late jump events to take advantage of favorable liquidity patterns at the close (Lou et al., 2019; Andersen and Bollerslev, 1997).

Although we find substantial evidence that end-of-day rebalancing by delta-hedgers impacts stock prices, the results presented here highlight the significant uncertainty regarding the timing of the rebalancing activity. As the simulation study shows, this uncertainty introduces errors in the empirical measurement of Γ^{HP}, which attenuates its estimated impact on end-of-day returns, driving coefficients towards zero.

Table 3: **Rebalancing Flows on Jump Days**
The table reports the results to regressing intraday jump returns on returns in the subsequent 30-minute interval. For Γ^{HP}, we identify jumps in the underlying stock returns. For Ω^{LETF}, we identify jumps in the benchmark of the leveraged ETF. To detect jumps, we compare each non-overlapping 30-minute return with returns of the same interval over the last year. If the return is higher (lower) than the 97.5% (2.5%) percentile, we regard the return as a jump. Next, we record the return from yesterday's close until the end of the 30-minute interval, where the jump has occurred ($r_t^{\text{incl. jump}}$). We collect also the return of the subsequent 30-minute interval (r_t^{next}) on the stock level. In the case of leveraged ETFs, we select all stocks in leveraged ETFs for which a jump in the benchmark index of the leveraged ETF has occurred. Jumps can take place before noon or after noon, but before 3pm. Equipped with $r_t^{\text{incl. jump}}$, r_t^{next}, and the intraday return of the benchmark index, we reconstruct Γ_t^{HP} and Ω_t^{LETF} for each affected stock j. Subsequently, we run Equation (9). T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity-fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{next}	(2) r_t^{next}	(3) r_t^{next}	(4) r_t^{next}	(5) r_t^{end}	(6) r_t^{end}
Γ_t^{HP}	-44.416**	0.251			-34.546	-77.316***
	(-1.984)	(0.013)			(-1.575)	(-2.668)
Ω_t^{LETF}			-0.220	-0.493		
			(-0.686)	(-0.642)		
$r_t^{\text{incl. jump}}$	-0.952***	-1.288***	-0.711**	-0.266		
	(-4.240)	(-6.311)	(-2.083)	(-1.561)		
$r^{\text{after jump}}$					-0.825***	-1.176***
					(-3.678)	(-3.955)
Observations	339,552	305,694	500,036	293,872	339,552	305,694
R^2 (%)	0.160	0.199	0.057	0.029	0.101	0.112
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]
Jumps Until	Pre-Noon	After-Noon	Pre-Noon	After-Noon	Pre-Noon	After-Noon

3.3. Reversal

If the mechanical rebalancing activities of delta-hedgers and leveraged ETF swap counterparties move prices sufficiently far from fundamental values, we should observe a quick reversal of these effects at the next open as other market participants enter the market and correct these deviations. Consequently, we expect returns on the following market open to relate positively to Γ^{HP} and negatively to Ω^{LETF}. To test this hypothesis,

we run the panel regression

$$r_{j,t}^{\text{night}} = \beta_0 r_{j,t}^{\text{end}} + \beta_1 \Gamma_{j,t}^{HP} + \beta_2 \Omega_{j,t}^{LETF} + FE_j + FE_t + \epsilon_{j,t}, \tag{12}$$

where $r_{j,t}^{\text{night}}$ is the return from the close of day t to 10:00 on day $t+1$.

Table 4: **Next-Day Reversal**
The table reports the results to regressing returns from closure of day t to 10:00 on day $t+1$ on returns until 15:30, r^{pre}, on option hedging pressure Γ^{HP} and leveraged ETF rebalancing quantity Ω^{LETF} following Equation (9). We also consider the return from 15:30 of day t to 10:00 on $t+1$, $r^{\text{end+night}}$, and the 24h return from day t's 15:30 to day $t+1$'s 15:30, $r^{\text{end+next pre}}$, as dependent variables. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{night}	(2) r_t^{night}	(3) r_t^{night}	(4) $r^{\text{end+night}}$	(5) $r^{\text{end+next pre}}$
r_t^{pre}	0.951**	1.268***	1.057***	0.235	-0.626
	(2.444)	(3.126)	(2.737)	(0.585)	(-1.163)
Γ_t^{HP}	25.573***		28.118***	17.132*	13.316
	(3.208)		(3.427)	(1.844)	(1.551)
Ω_t^{LETF}		-70.943***	-79.521***	-32.584	-40.628
		(-3.382)	(-3.634)	(-1.515)	(-1.402)
Observations	1,882,234	1,882,234	1,882,234	1,882,234	1,879,754
R^2 (%)	0.030	0.028	0.040	0.007	0.006
Entity FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

Table 4 highlights the relatively swift reversal of the impact of Γ^{HP} and Ω^{LETF}: the coefficient on Γ^{HP} in column (1) is positive at 25.57 (t-value: 3.28), while that on Ω^{LETF} in column (2) is negative at -70.95 (t-value: -3.38). In column (3), we include both rebalancing flows jointly and show a magnification of the proposed reversal effects for both.

In order to assess if the initial price impact is fully reverted over night, we substitute the overnight return in Equation (12) by the return from 15:30 of day t to 10:00 on $t+1$, $r^{\text{end+night}}$. A full reversal of the effect would render the impact of Γ^{HP} and Ω^{LETF} on $r^{\text{end+night}}$ insignificant. As column (4) in Table 4 shows, this is the case for Ω^{LETF}. However, the coefficient on Γ^{HP} is statistically significant at the 10% level, suggesting

that rebalancing flows from delta-hedgers take more time to be fully reversed. Indeed, we observe a full reversal effect for the impact of gamma rebalancing when considering the return from t's 15:30 to $t+1$'s 15:30, $r^{\text{end+next pre}}$, as shown in column (5).

3.4. Liquidity Provision

How do market participants react when delta-hedgers or leveraged ETF swap counterparties have to trade particularly large quantities of a certain stock? Strategic traders could engage in predatory trading by amplifying the rebalancing activities from option delta-hedgers and leveraged ETF swap counterparties (Brunnermeier and Pedersen, 2005). Predators initially buy (sell) the underlying while rebalancing flows exert significant buying (selling) pressure and unwind their position as the rebalancing concludes. Contrarily, strategic traders may provide liquidity instead, which Admati and Pfleiderer (1991) term a *sunshine equilibrium*. In this scenario, the price impact of rebalancing flows is minimized.

To investigate how strategic traders react to large rebalancing flows, we augment Equation (9) with a dummy interaction for days when the rebalancing flow exceeds a certain percentile:

$$r_{j,t}^{\text{end}} = \beta_0 r_{j,t}^{\text{pre}} + \beta_1 F_{j,t} + \beta_2 F_{j,t} \times \mathrm{L}_{j,t,x} + FE_j + FE_t + \epsilon_{j,t}, \tag{13}$$

where F is one of our two proxies for order flow, and $\mathrm{L}_{j,t,x}$ is a dummy variable that is equal to 1 if the absolute magnitude of the respective rebalancing flow of stock j exceeds the xth cross-sectional percentile. We consider three thresholds for x, i.e, $x \in \{95, 97.5, 99\}$.

Columns (1) to (3) of Table 5 report the results for Γ^{HP}, whereas columns (4) to (6) do so for Ω^{LETF}. The estimated coefficients for the interaction terms are positive, but not statistically significant for Γ^{HP}, implying that we neither find evidence in favor of predatory trading nor in favor of a sunshine equilibrium. On the contrary, the estimated coefficients for Ω^{LETF} rebalancing flows are negative and highly significant for large flows that exceed the 95th percentile of Ω^{LETF}. We find a monotonic increase in β_2 in the threshold used to identify large leveraged ETF rebalancing flows. This pattern is consistent with a sunshine equilibrium in the spirit of Admati and Pfleiderer (1991), in which other market participants accommodate the liquidity needs by large rebalancing flows from leveraged ETFs.

Table D.1 in the Appendix provides supporting evidence using the change in the bid-ask spread of individual stocks. A shock to the absolute value of Γ^{HP} prompts market

Table 5: **Liquidity Provision Around Large Rebalancing Flows**
The table reports the results to regressing returns in the last half hour of a trading day on returns until 15:30 on delta-hedging and leveraged ETF rebalancing flows including dummy variables for large rebalancing flows. $\mathrm{L}_{j,t,x}$ is a dummy variable that is equal to 1 if the size of the respective rebalancing flow of stock j exceeds the xth cross-sectional percentile of the absolute rebalancing flow. We consider three thresholds for x, i.e, $x \in \{95, 97.5, 99\}$. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity-fixed effects in all specifications and value-weight observations.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}	(4) r_t^{end}	(5) r_t^{end}	(6) r_t^{end}
Γ_t^{HP}	-11.588***	-11.294***	-10.333***			
	(-3.407)	(-4.358)	(-4.114)			
$\Gamma^{HP} \times \mathrm{L}_{t,x}$	2.828	3.252	2.696			
	(1.081)	(1.373)	(0.855)			
Ω_t^{LETF}				57.027***	53.729***	49.683***
				(4.803)	(5.230)	(5.672)
$\Omega^{LETF} \times \mathrm{L}_{t,x}$				-22.267**	-26.226***	-30.561***
				(-2.327)	(-2.725)	(-3.111)
r_t^{pre}	-0.754***	-0.753***	-0.755***	-0.909***	-0.908***	-0.908***
	(-6.905)	(-6.902)	(-6.894)	(-8.212)	(-8.183)	(-8.187)
Observations	1,882,332	1,882,332	1,882,332	1,882,332	1,882,332	1,882,332
R^2 (%)	0.284	0.285	0.284	0.334	0.335	0.333
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]
xth perc.	95.0	97.5	99.0	95.0	97.5	99.0

makers in the underlying stock to increase the corresponding spreads in advance. In contrast, we find no such evidence for rebalancing flows from leveraged ETFs. Together with the results regarding large flows in Table 5, we find convincing evidence in favor of a better liquidity provision for the flows of leveraged ETF counterparties, driven by a greater transparency on their timing and of their magnitude. Conversely, market makers in the underlying stocks face the risk of adverse selection from the flows of option market makers due to their opaque nature and discretionary timing.

3.5. Trading Strategies

Next, we quantify the economic significance of these frictions by studying trading strategies that use Γ^{HP} and Ω^{LETF} as timing signals. First, we construct a long-short portfolio based on Γ^{HP}. On each trading day at 15:30, we sort our stock sample according to Γ^{HP} into decile portfolios. Subsequently, we build a long-short strategy, denoted by LmH^{Γ}, by taking a long (short) position in the lowest (highest) decile portfolio. We close our positions at the market close of the same trading day, such that the strategy is active for only 30 minutes each day. In the second strategy, HmL^{Ω}, we use Ω^{LETF} as a timing signal. We construct decile portfolios based on Ω^{LETF} and take a long (short) position in the highest (lowest) decile. Finally, we use both Γ^{HP} and Ω^{LETF} as signals in a combined strategy. For each stock at 15:30, we sort our stock sample according to their aggregate imbalance $\Omega^{LETF} - \Gamma^{HP}$ into decile portfolios and subsequently build a long-short strategy, denoted by $\text{HmL}^{\Gamma,\Omega}$.

Table 6: Γ^{HP} **and** Ω^{LETF} **Based Trading Strategies**
The table reports the economic value of timing the last half-hour market return using using Γ^{HP}, Ω^{LETF} and a joint signal based on both. The Γ^{HP}-strategy, LmH^{Γ}, takes a long (short) position in a stock when the stock's Γ^{HP} is in the lowest (highest) decile. The Ω^{LETF}-strategy, HmL^{Ω}, takes a long (short) position in a stock when the stock's Ω^{LETF} is in the highest (lowest) decile. The combined strategy, $\text{HmL}^{\Gamma,\Omega}$, computes a cross-sectional rank according to Γ^{HP} and Ω^{LETF} and takes a long (short) position in a stock when the stock's aggregated ranking is in the highest (lowest) decile. As a benchmark, $\text{Market}^{\text{end}}$ denotes investing in all stocks from 15:30 to 16:00. Portfolios are value-weighted, including for $\text{Market}^{\text{end}}$. For each strategy, we report the average return (Avg ret), standard deviation (Std dev), Sharpe ratio (Sharpe), Sharpe ratio assuming an effective spread of 25% of the quoted bid-ask spread (Net Sharpe), skewness, kurtosis, and success rate (Success). The returns are annualized and in percentage. Newey and West (1987) robust t-statistics are in parentheses, and significance at the 1%, 5%, or 10% level is denoted by ***, **, or *, respectively. The sample period is January 2012 – December 2019.

	Avg ret	Std dev	Sharpe	Net Sharpe	Skewness	Kurtosis	Success
LmH^{Γ}	5.20***	1.36	3.83	1.76	0.17	3.00	61.07
HmL^{Ω}	5.78***	2.31	2.51	0.05	0.13	3.54	56.29
$\text{HmL}^{\Gamma,\Omega}$	6.63***	1.54	4.29	1.77	0.27	2.80	62.70
$\text{Market}^{\text{end}}$	-1.39	3.26	-0.43	-0.43	-1.05	12.27	51.92

Table 6 reports summary statistics of the resulting strategies using value-weighted portfolios. LmH^{Γ} yields an average total excess return equal to 5.20% per year which are significant at the 1%-level. Although HmL^{Ω} yields a slightly higher average total excess return per year, 5.78%, LmH^{Γ} exhibits a higher Sharpe ratio (3.83 vs. 2.51). Similarly, we observe higher success ratios for LmH^{Γ} compared to HmL^{Ω} (61% vs. 56%).

The combination of the two signals Ω^{LETF} and Γ^{HP} reduces the noise and yields an improvement of the success ratio with an increase in the average return to 6.62% per year, which is significant at the 1%-level. Consistent with this, the combined strategy exhibits the highest (gross) Sharpe ratio of 4.29 and a success rate of 62%. It is notable that the skewness of the three strategies are positive, compared to the negative skewness of the return on the market.[12]

Table 7: **Risk-adjusted Returns of Γ^{HP} and Ω^{LETF} Trading Strategies**
The table reports the estimation results from regressing returns of strategies timing the last half-hour based on Γ^{HP}, Ω^{LETF} and a combination of both on the returns of all stocks from 15:30 to 16:00 ($\text{Market}^{\text{end}}$), the equity market excess return (MKT), size (SMB), book-to-market (HML), profitability (RMW), investment (CMA), momentum (MOM), and an intermediary capital asset pricing factor (IC, proposed by He et al., 2017). The Γ^{HP}-strategy, LmH^{Γ}, takes a long (short) position in a stock when the stock's Γ^{HP} is in the lowest (highest) decile. The Ω^{LETF}-strategy, HmL^{Ω}, takes a long (short) position in a stock when the stock's Ω^{LETF} is in the highest (lowest) decile. The combined strategy, $\text{HmL}^{\Gamma,\Omega}$, takes a long (short) position in a stock when the stock's $\Omega^{LETF} - \Gamma^{HP}$ is in the highest (lowest) decile. Portfolios are value-weighted. Newey and West (1987) robust t-statistics are in parentheses, and significance at the 1%, 5%, or 10% level is denoted by ***, **, or *, respectively. The sample period is January 2012 – December 2019.

	Intercept	$\text{Market}^{\text{end}}$	MKT	SMB	HML	RMW	CMA	MOM	IC	R2	R2 adj
LmH^{Γ}	5.21	0.01								0.03%	-0.02%
	(9.37***)	(0.42)									
	4.84		1.24	1.11	0.46	2.89	0.89	2.13	0.72	0.88%	0.45%
	(7.99***)		(1.00)	(0.88)	(0.27)	(1.58)	(0.31)	(2.10**)	(0.86)		
HmL^{Ω}	5.77	-0.01								0.02%	-0.03%
	(6.07***)	(-0.30)									
	6.22		0.24	0.93	-1.11	-4.98	1.03	-4.06	-1.04	0.72%	0.29%
	(5.87***)		(0.09)	(0.42)	(-0.36)	(-1.27)	(0.25)	(-2.87***)	(-0.61)		
$\text{HmL}^{\Gamma,\Omega}$	6.66	0.02								0.2%	0.1%
	(10.24***)	(1.82*)									
	6.30		3.05	0.19	-0.38	-0.02	4.79	1.30	-0.02	1.10%	0.70%
	(8.70***)		(2.62***)	(0.12)	(-0.19)	(-0.01)	(1.55)	(1.14)	(-0.02)		

Because of the significant turnover ratio required in such a strategy, it is important to investigate the potential impact of transaction costs if an trader were to implement such a strategy. We assume an effective spread of 25% of the quoted bid-ask spread at 15:30 as a measure for transaction costs. As evident from Table 6, we find that the trading strategy based on Γ^{HP} continues to yield high risk-adjusted performance with a net Sharpe ratio of 1.76. On the other hand, the net Sharpe ratio for HmL^{Ω} is reduced to 0.05. The strategy that combines the two signals $\text{HmL}^{\Gamma,\Omega}$ generates the highest net

[12]Jiang, Kelly, and Xiu (2020) use convolutional neural networks on price path images to predict the future return direction. They achieve a success rate of up to 53.6%.

Sharpe ratio, which is equal to 1.77. This confirms that the economic frictions generated by these two channels jointly are indeed economically significant, although Γ^{HP} plays a greater role.

How much of these strategies' end-of-day returns are explained by exposure to well-established risk factors? Table 7 summarizes the results of regressing LmH^{Γ}, HmL^{Ω}, and $\text{HmL}^{\Gamma,\Omega}$ on the end-of-day market return ($\text{Market}^{\text{end}}$) and the full-day Fama-French five-factor model (Fama and French, 2015), augmented by a momentum as well as an intermediary capital asset pricing factor (He et al., 2017).[13] We find that neither the market nor the augmented five-factor model can explain the returns of LmH^{Γ}, HmL^{Ω}, and $\text{HmL}^{\Gamma,\Omega}$. The R^2 is small on average. The only factor that is statistically significant for the joint strategy is MKT, the other factors in the extended five-factor model are not significant. Notwithstanding the statistical importance of MKT, the alphas of these regressions are statistically significant and virtually unchanged and account for almost the entire average gross excess return.

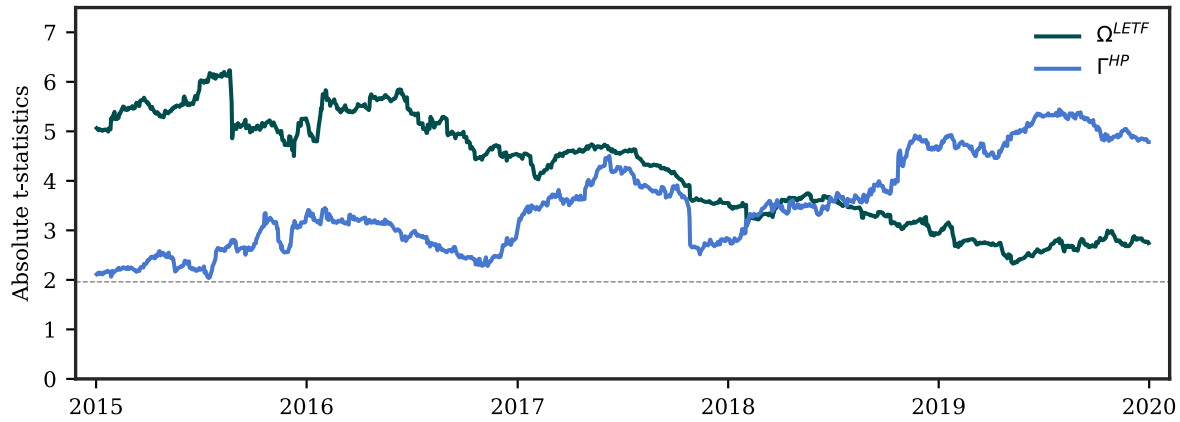


Fig. 5. **Rolling Regressions**
The figure shows the absolute t-statistics of estimated Γ^{HP} and Ω^{LETF} coefficients over time, as measured on a rolling basis for each three year-period starting in January 2012. Specifically, we estimate Equation (9) for the first three years in our sample, shift the starting and end date by one day and repeat this for all days in the sample (January 2012 – December 2019).

[13]Data on the Fama-French five-factor model and the momentum factor is taken from Kenneth French's website, `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`, whereas data on the intermediary capital asset pricing factor is obtained from Zhiguo He's website, `https://voices.uchicago.edu/zhiguohe/data-and-empirical-patterns/intermediary-capital-ratio-and-risk-factor/`.

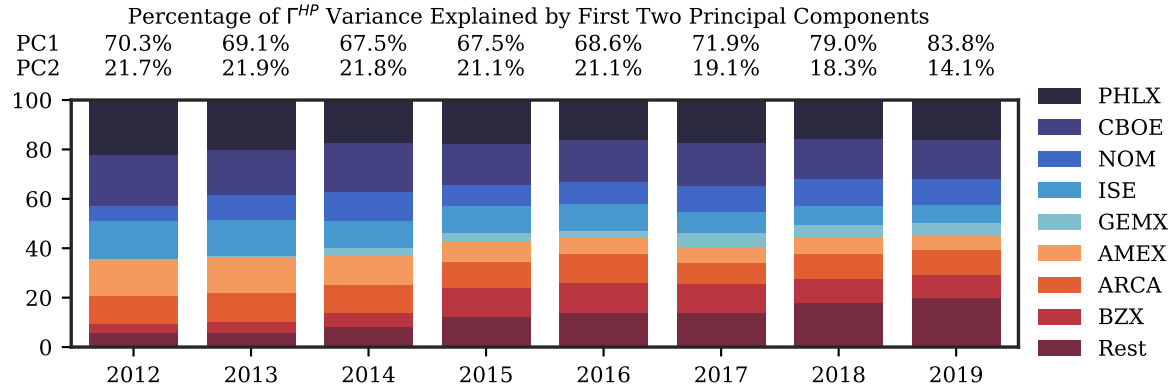


Fig. 6. **Market Share of Option Exchange Volume**
The figure shows the annual share of transacted volume in equity options on each U.S. options exchange. The market share is given for the five option exchanges in our sample (PHLX, CBOE, NOM, ISE, and GEMX), as well as the three largest exchanges among the remainder (AMEX, ARCA, and BZX). *Rest* combines volumes on exchanges BOX, BXO, C2, EDGX, EMLD, MERC, MIAX, and PEARL. The upper part of the figure shows the percentage of Γ^{HP} variance explained by the first two principal components across all option exchanges in our sample on a year-by-year basis. PC1 and PC2 denote the first and second principal component, respectively.

3.6. The Evolution of the Imbalance and its Price Impact

We investigate how the price impact of Γ^{HP} and Ω^{LETF} changed over time by running our baseline regression using a rolling window of three years. Figure 5 displays the absolute t-statistics of the coefficients for Γ^{HP} (blue line) and for Ω^{LETF} (green line). We find that the statistical significance of Γ^{HP} has stayed nearly unchanged until the end of 2016 and has since then further increased. On the other hand, the estimated significance for Ω^{LETF} has decreased over time. Figure E.1 in Appendix E shows that the difference in statistical significance over time is also reflected in the performance of the timing strategies based on Γ^{HP} and Ω^{LETF}.

A probable reason for the difference in the statistical performance over time lies in the high fragmentation of the U.S. options market, which reduces the availability and quality of information about the inventory positions of delta-hedgers.[14] This limits the ability of arbitrageurs to construct a precise measure of the rebalancing demand of delta-hedgers. Figure 6 shows the year-by-year market share for the five exchanges in our sample, as well as the three largest remaining exchanges, measured by volume in 2019. The figure shows a trend towards higher dispersion of option volumes across exchanges. This was

[14]See Andersen, Archakov, Grund, Hautsch, Li, Nasekin, Nolte, Pham, Taylor, and Todorov (2021) for a current account on the history of the U.S. options market.

helped by the opening of new exchanges in recent years, which allows market makers to strategically hide information about their inventory positions. At the start of the sample, there were nine active option exchanges; since then, seven new exchanges have been opened (Andersen et al., 2021). These new exchanges represent about a fifth of the entire volume in equity options in 2019. Moreover, not all option venues provided information useful to approximate the inventory of delta-hedgers. For example, AMEX and ARCA, which correspond to roughly 30% of traded options volume at the start of our sample, started reporting a detailed breakdown of option volumes in March of 2022 (NYSE, 2022). The latter informational friction can be mitigated if option inventories of delta-hedgers are highly correlated across option exchanges. We run a principal component decomposition of the gamma imbalance across the option exchanges for which we observe the data. We find that the first principal component can explain between 67% and 84% of the total variance of the gamma imbalance. The second principal component explains between 14% and 22%, suggesting that information from multiple exchanges is indeed needed to obtain a precise proxy of the aggregate gamma imbalance. We also find that the average 3-year rolling correlations range from 30% to 80% and are decreasing for ISE and CBOE over time (see Figure D.1 in the Online Appendix). These results suggest the presence of an informational friction that may explain why the abnormal excess returns earned by strategies based on Γ^{HP} have been large and persistent.

4. Conclusion

A growing literature focuses on cross-sectional intraday return variation, linking it to investor heterogeneity on the stock level. In this paper, we provide novel insights into how derivative markets add to cross-sectional variation towards the end of the trading day. By drawing upon a unique dataset merging data from several exchanges identifying types of market participants in U.S. stock options and the portfolio composition of U.S. equity-focused leveraged ETFs, we document large price pressure on end-of-day returns when option market makers engage in delta-hedging and leveraged ETF swap counterparties rebalance their positional exposure.

We show that delta-hedging and leveraged ETF rebalancing exert an economically large price pressure on end-of-day returns. Whereas leveraged ETFs contribute to a basket-wide momentum effect, delta-hedging can either have a stabilizing effect in the form of end-of-day reversal, but may also exaggerate intraday momentum. The direction is determined by the previous return of the underlying and the aggregate option inven-

tory of market makers. Moreover, our results reveal that option market makers have discretion on the execution of their hedging strategies, especially after intraday jumps. On the contrary, leveraged ETF swap counterparties are required to establish the target exposure of the fund at the close. This institutional flexibility translates into information asymmetry faced by liquidity providers, constituting a friction for liquidity provision. Consistent with the assumption that the impact of rebalancing is mechanical in nature, we also document that it is transitory, as the effects fully reverse during the next trading day. Such a reversal materializes in a shorter time frame for leveraged ETFs price pressure than for that induced by gamma imbalances.

We finally show that the effect of option market imbalances is persistent over the years, while that of leveraged ETFs is significantly decreasing, further suggesting a key role of information asymmetry as the main friction underlying their dynamics.

Appendix

Appendix A Detailed Summary Statistics on Γ^{HP} and Ω^{LETF}

Table A.1 provides detailed summary statistics on the main variables of interest in the main paper.

Table A.1: **Summary Statistics**
The table reports means, standard deviations, and quantiles for the gamma, leveraged ETF variables and returns used in the regression models. The descriptive statistics are first computed for each day and subsequently averaged across all days. r^{pre}, r^{night} and r^{end} are denoted in basis points. ADV^{end} is given in million USD. $\text{RSVOL}^{\text{end}}$, Γ^{IB}, Γ^{HP}, Ω^{LETF} and their corresponding absolute values are given percentage. The sample period is from January 2012 to December 2019.

	Mean	Std	2.5%	10%	25%	50%	75%	90%	97.5%
r^{pre}	4.75	203.33	-379.68	-187.53	-81.11	4.67	89.38	194.47	394.41
r^{end}	0.02	35.56	-73.4	-38.44	-17.9	-0.53	17.21	38.98	77.24
r^{night}	4.16	167.3	-249.0	-117.38	-50.88	2.97	57.01	125.12	267.0
ADV^{end}	21.72	59.1	0.25	0.78	2.19	6.79	19.94	47.4	136.93
$\text{RSVOL}^{\text{end}}$	0.56	41.86	-86.08	-48.77	-23.39	0.49	24.45	50.03	87.63
Γ^{IB}	0.52	3.0	-3.96	-1.32	-0.26	0.08	0.93	2.91	6.99
$\lvert\Gamma^{IB}\rvert$	1.45	2.67	0.0	0.03	0.14	0.55	1.67	3.76	8.03
Γ^{HP}	0.04	3.27	-7.0	-2.25	-0.48	0.0	0.51	2.37	7.28
$\lvert\Gamma^{HP}\rvert$	1.61	2.86	0.0	0.02	0.1	0.48	1.75	4.51	10.38
Ω^{LETF}	0.15	1.79	-2.97	-1.72	-0.86	0.1	1.11	2.08	3.48
$\lvert\Omega^{LETF}\rvert$	1.65	1.75	0.03	0.16	0.42	1.03	2.37	3.92	6.25
$\Omega^{LETF} - \Gamma^{HP}$	0.11	3.81	-7.99	-3.29	-1.27	0.1	1.52	3.56	8.1
$\lvert\Omega^{LETF} - \Gamma^{HP}\rvert$	2.64	3.14	0.05	0.22	0.61	1.63	3.48	6.25	11.43

Appendix B Errors in Measuring Γ^{HP} – A Simulation Study

We simulate returns from last day's close to 15:30 (r^{pre}) and during the last half hour of a trading day (r^{end}), as well as the true Gamma imbalance of market makers for $N = 1000$ stocks, over a total of $T = 2000$ time steps, mirroring the eight years of daily data in our sample.

We allow the delta-hedger to deviate from hedging at the close. She will hedge during a different time with probability π, i.e. $p \sim \text{Bern}(\pi)$, such that the true Gamma hedging pressure for stock j on day t equals

$$\Gamma^{HP}_{j,t}(*) = r^{\text{pre}}_{j,t} \times \Gamma^{IB}_{j,t}(*) \times p,$$

where $r^{\text{pre}}_{j,t} \sim \mathcal{N}(0, \sigma^{\text{pre}} = 0.08)$. The choice of whether the market maker hedges end-of-day is inherently unobservable to other market participants. At the same time, $\Gamma^{IB}(*)$ may only be approximated with the market maker's inventory measured at the end of the *last* trading day. Intraday changes in the Gamma profile are likewise unobservable. Consequently, the empirically-observed hedging pressure is measured with error:

$$\Gamma^{HP}_{j,t} = r^{\text{pre}}_{j,t} \times (\Gamma^{IB}_{j,t}(*) + \varepsilon^{\Gamma})_{j,t},$$

where $\varepsilon^{\Gamma} \sim \mathcal{N}[0, \sigma^{\Delta} \times (\sigma^{\text{pre}} + \sigma^{\text{post}})\sqrt{2}]$, such that the standard deviation is scaled as a fraction of the variation of the true but unobservable $\Gamma^{IB}(*)$.

Next, imply a direct relationship between end-of-day returns and $\Gamma^{HP}(*)$,

$$r^{\text{end}}_{j,t} = \beta^{\Gamma}(*) \times \Gamma^{HP}_{j,t}(*) + \varepsilon^{\text{post}}_{j,t},$$

with $\varepsilon^{\text{post}} \sim \mathcal{N}(0, \sigma^{\text{post}})$ and $\sigma^{\text{post}} = 0.04$. We set the true slope coefficient $\beta^{\Gamma}(*)$ to -41, mimicking the impact of leveraged ETF rebalancing. Empirically, we can estimate the impact of Gamma rebalancing on end-of-day returns only through the noisy proxy Γ^{HP}:

$$r^{\text{end}}_{j,t} = \beta^{\Gamma} \times \Gamma^{HP}_{j,t} + \varepsilon^{\text{post}}_{j,t},$$

We are consequently interested in how the degree of noise (σ^{Δ}) and probability of discretionary trading at a different time than the close (π) impact the empirical estimate of the slope coefficient, β^{Γ}. Results to this are shown in Figure 4.

Appendix C Controlling for Risk- and Information-based Alternative Explanations

The unveiled empirical relationship may be consistent with alternative hypotheses. In particular, it may arise from intraday oscillations in the level of risk, for volatility-target investors might rebalance their portfolio away from risky assets towards the end of the trading day if volatility changes, generating directional order flow. Since option demand is also a positive function of investors expectations about future volatility, changes in risk may impact the inventory of market makers and, potentially, their Γ imbalance. These two observations may generate a mechanical and negative relationship between gamma imbalance and order-flow, driven by the dynamics of expected volatility.

To rule out the above explanations, specifications in columns (1) to (3) in Table C.1 add controls for proxies of expected volatility, namely the level of the at-the-money implied volatility (IV) on day $t-1$ and forecasts of future variance using an EGARCH model, $\hat{RV}^{end}$ and $\hat{RV}^{pre}$. The EGARCH models the sum of 1-minute squared returns for each stock in the last half hour of a trading day, using historical data up until day $t-1$ (see Moskowitz, Ooi, and Pedersen, 2012).

Results show that neither of these two controls is significant, nor does the inclusion of either materially alter the coefficient and statistical significance of Γ^{HP} and Ω^{LETF}.

Besides risk-based explanations, our findings may arise from information being transmitted from the options market to the market of the underlying. A stream of the literature argues that options are often used because of their implicit leverage in presence of market frictions, for instance when short selling is expensive (Ge et al., 2016). Accordingly, Blau et al. (2014) find that an increase in the put-call ratio negatively predicts future returns at the daily, weekly, and monthly frequency. Roll et al. (2010) and Hu (2014) show that the option-to-stock volume predicts options' underlying returns. We thus extend the baseline regression to control for these explanatory variables. Columns (4) and (5) summarize the results and show that the coefficients on Γ^{HP} and Ω^{LETF} are unchanged both in magnitude and significance, suggesting that our results are not driven by an information channel.

Table C.1: **Controlling for Risk-based and Informational Channels**
The table reports the results to regressing the returns in the last half hour of a trading day on returns until 15:30 (r^{pre}), on options market maker hedging pressure Γ^{HP} and leveraged ETF rebalancing quantity Ω^{LETF} following Equation (9). IV_{t-1} denotes implied volatility at time $t-1$. $\hat{RV}_t^{end}$ ($\hat{RV}_t^{pre}$) denote the square root of predicted realized variance for the time period from 15:30 to 16:00 (from previous day's close to 15:30) on day t. PC_{t-1} is the put-call-ratio and $O/S^{\$}_{t-1}$ denotes the option-to-stock volume in dollar terms. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity-fixed effects in all specifications and value-weight observations. The sample period is 2012 – 2019.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}	(4) r_t^{end}	(5) r_t^{end}	(6) r_t^{end}
Γ_t^{HP}	-10.925***	-10.958***	-10.938***	-10.961***	-10.961***	-10.914***
	(-5.513)	(-5.522)	(-5.507)	(-5.521)	(-5.523)	(-5.503)
Ω_t^{LETF}	46.832***	46.768***	46.790***	46.755***	46.738***	46.835***
	(5.208)	(5.202)	(5.198)	(5.195)	(5.198)	(5.204)
r_t^{pre}	-0.822***	-0.820***	-0.820***	-0.820***	-0.820***	-0.821***
	(-7.460)	(-7.453)	(-7.452)	(-7.453)	(-7.452)	(-7.454)
IV_{t-1}	4.463***					4.040***
	(5.079)					(5.086)
$\hat{RV}_t^{pre}$		19.132				-28.421*
		(1.065)				(-1.754)
$\hat{RV}_t^{end}$			414.582***			336.520***
			(3.156)			(2.777)
PC_{t-1}				0.005		-0.003
				(0.030)		(-0.019)
$O/S^{\$}_{t-1}$					255.550	100.469
					(0.782)	(0.321)
Observations	1,882,332	1,882,236	1,882,236	1,879,828	1,882,327	1,879,753
R^2 (%)	0.388	0.371	0.380	0.370	0.370	0.393
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

Appendix D Liquidity Provision: Evidence from Spreads

This appendix provides additional evidence of impeded liquidity provision for Gamma rebalancing flows, which potentially arises because of their opacity. To do so, we measure the relative change in the bid-ask spread on each stock as

$$\widehat{S}_t = \frac{S_t^{16:00:00} - S_t^{15:30:00}}{\sum_{\tau=09:30:00}^{15:00:00} S_t^{\tau}/12}, \qquad \text{with} \quad S_t = ask_t - bid_t,$$

and regress it on rebalancing flows Γ^{HP} and Ω^{LETF}. Table D.1 shows that market makers active in the stocks underlying options with large rebalancing flows increase the corresponding spreads to alleviate the issue of adverse selection. We find no such evidence for rebalancing flows from leveraged ETFs.

Table D.1: **Spreads and Rebalancing Flows**
The table reports the results to regressing the relative change in the bid-ask spread on Γ^{HP} and Ω^{LETF}. We measure the relative change in the spread as $\widehat{S}_t = \frac{S_t^{16:00:00} - S_t^{15:30:00}}{\sum_{\tau=09:30:00}^{15:00:00} S_t^{\tau}/12}$, where $S_t = ask_t - bid_t$ using consolidated national best bid and offer quotes from the NYSE TAQ database (Holden and Jacobsen, 2014). T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity-fixed effects in all specifications and value-weight observations.

	(1)	(2)	(3)
Dependent	$\widehat{S}_t$	$\widehat{S}_t$	$\widehat{S}_t$
$\lvert\Gamma_t^{HP}\rvert$	6.081**		5.555*
	(2.058)		(1.876)
$\lvert\Omega_t^{LETF}\rvert$		16.353	14.572
		(0.843)	(0.750)
r_t^{pre}	0.026	0.029	0.025
	(0.235)	(0.259)	(0.228)
Observations	1,863,111	1,863,111	1,863,111
R^2 (%)	0.008	0.006	0.013
Entity FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]

Appendix E Performance of Trading Strategies Over Time

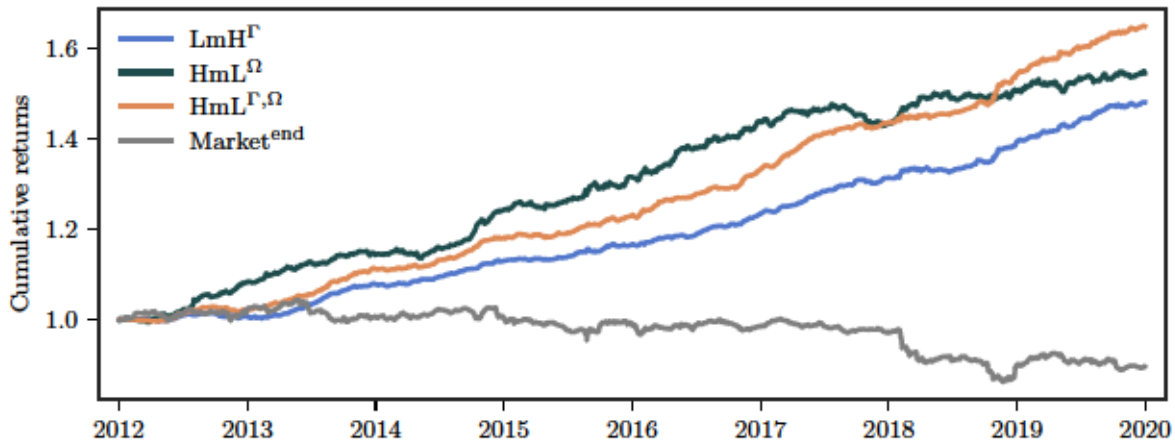


Fig. E.1. **Performance of Trading Strategies Over Time**
The figure shows the cumulative performance of a one-dollar investment into trading strategies based on Γ^{HP} and Ω^{LETF}. The Γ^{HP}-strategy, LmH$^{\Gamma}$, takes a long (short) position in a stock when the stock's Γ^{HP} is in the lowest (highest) decile. The Ω^{LETF}-strategy, HmL$^{\Omega}$, takes a long (short) position in a stock when the stock's Ω^{LETF} is in the highest (lowest) decile. The combined strategy, HmL$^{\Gamma,\Omega}$, takes a long (short) position in a stock when the stock's $\Omega^{LETF} - \Gamma^{HP}$ is in the highest (lowest) decile. As a benchmark, Market^{end} denotes investing in all stocks from 15:30 to 16:00. The sample period is January 2012 – December 2019.

Bibliography

Admati, A. R., Pfleiderer, P., 1991. Sunshine trading and financial market equilibrium. Review of Financial Studies 4, 443–481.

Andersen, T., Archakov, I., Grund, L., Hautsch, N., Li, Y., Nasekin, S., Nolte, I., Pham, M. C., Taylor, S., Todorov, V., 2021. A descriptive study of high-frequency trade and quote option data. Journal of Financial Econometrics 19, 128–177.

Andersen, T. G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4, 115–158.

Baltussen, G., Da, Z., Lammers, S., Martens, M., 2021. Hedging demand and market intraday momentum. Journal of Financial Economics 142, 377–403.

Barbon, A., Buraschi, A., 2020. Gamma fragility. University of St.Gallen, School of Finance Research Paper.

Ben-David, I., Franzoni, F., Moussawi, R., 2018. Do etfs increase volatility? Journal of Finance 73, 2471–2535.

Benveniste, L. M., Marcus, A. J., Wilhelm, W. J., 1992. What's special about the specialist? Journal of Financial Economics 32, 61–86.

Blau, B. M., Nguyen, N., Whitby, R. J., 2014. The information content of option ratios. Journal of Banking & Finance 43, 179 – 187.

Bogousslavsky, V., 2020. The Cross-Section of Intraday and Overnight Returns.

Bogousslavsky, V., Muravyev, D., 2020. Who trades at the close? implications for price discovery, liquidity.

Brøgger, S. B., 2021. The market impact of predictable flows: Evidence from leveraged vix products. Journal of Banking & Finance 133, 106280.

Brunnermeier, M. K., Pedersen, L. H., 2005. Predatory trading. Journal of Finance 60, 1825–1863.

Chen, H., Joslin, S., Ni, S. X., 2019. Demand for Crash Insurance, Intermediary Constraints, and Risk Premia in Financial Markets. Review of Financial Studies 32, 228–265.

Cheng, M., Madhavan, A., 2010. The dynamics of leveraged and inverse exchange-traded funds. Journal of Investment Management 7, 43–62.

Chordia, T., Kurov, A., Muravyev, D., Subrahmanyam, A., 2021. Index option trading activity and market returns. Management Science 67, 1758–1778.

Cremers, M., Weinbaum, D., 2010. Deviations from put-call parity and stock return predictability. Journal of Financial and Quantitative Analysis 45, 335–367.

Easley, D., O'Hara, M., Srinivas, P. S., 1998. Option volume and stock prices: Evidence on where informed traders trade. Journal of Finance 53, 431–465.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Forster, M. M., George, T. J., 1992. Anonymity in securities markets. Journal of Financial Intermediation 2, 168–206.

Fournier, M., Jacobs, K., 2020. A Tractable Framework for Option Pricing with Dynamic Market Maker Inventory and Wealth. Journal of Financial and Quantitative Analysis 55, 1117–1162.

Frey, R., 1998. Perfect option hedging for a large trader. Finance and Stochastics 2, 115–141.

Frey, R., Stremme, A., 1997. Market volatility and feedback effects from dynamic hedging. Mathematical Finance 7, 351–374.

Gao, L., Han, Y., Zhengzi Li, S., Zhou, G., 2018. Market intraday momentum. Journal of Financial Economics 129, 394–414.

Ge, L., Lin, T.-C., Pearson, N. D., 2016. Why does the option to stock volume ratio predict stock returns? Journal of Financial Economics 120, 601–622.

Gemmill, G., 1996. Transparency and liquidity: A study of block trades on the london stock exchange under different publication rules. Journal of Finance 51, 1765–1790.

Golez, B., Jackwerth, J. C., 2012. Pinning in the s&p 500 futures. Journal of Financial Economics 106, 566 – 585.

Gârleanu, N., Pedersen, L. H., Poteshman, A. M., 2009. Demand-Based Option Pricing. Review of Financial Studies 22, 4259–4299.

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

Heston, S. L., Korajczyk, R. A., Sadka, R., 2010. Intraday patterns in the cross-section of stock returns. Journal of Finance 65, 1369–1407.

Holden, C. W., Jacobsen, S., 2014. Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. Journal of Finance 69, 1747–1785.

Hu, J., 2014. Does option trading convey stock price informationα. Journal of Financial Economics 111, 625–645.

Jacobs, K., Mai, A. T., 2020. The role of intermediaries in derivatives markets: Evidence from vix options.

Jiang, J., Kelly, B. T., Xiu, D., 2020. (re-) imag (in) ing price trends. Chicago Booth Research Paper .

Johnson, T., Liang, M., Liu, Y., 2016. What Drives Index Options Exposures?*. Review of Finance 22, 561–593.

Johnson, T. L., So, E. C., 2012. The option to stock volume ratio and future returns. Journal of Financial Economics 106, 262–286.

Komarov, O., 2017. Intra-day momentum.

Lou, D., Polk, C., Skouras, S., 2019. A tug of war: Overnight versus intraday expected returns. Journal of Financial Economics 134, 192–213.

Madhavan, A., 1995. Consolidation, fragmentation, and the disclosure of trading information. Review of Financial Studies 8, 579–603.

Madhavan, A., 2000. Market microstructure: A survey. Journal of financial markets 3, 205–258.

Moskowitz, T. J., Ooi, Y. H., Pedersen, L. H., 2012. Time series momentum. Journal of Financial Economics 104, 228 – 250, special Issue on Investor Sentiment.

Ni, S. X., Pan, J., Poteshman, A. M., 2008. Volatility information trading in the option market. Journal of Finance 63, 1059–1091.

Ni, S. X., Pearson, N. D., Poteshman, A. M., 2005. Stock price clustering on option expiration dates. Journal of Financial Economics 78, 49–87.

Ni, S. X., Pearson, N. D., Poteshman, A. M., White, J., 2020. Does Option Trading Have a Pervasive Impact on Underlying Stock Prices? Review of Financial Studies 34, 1952–1986.

NYSE, 2022. Nyse options open/close volume summary client specification.

Pan, J., Poteshman, A. M., 2006. The Information in Option Volume for Future Stock Prices. Review of Financial Studies 19, 871–908.

Platen, E., Schweizer, M., 1998. On feedback effects from hedging derivatives. Mathematical Finance 8, 67–84.

Porter, D. C., Weaver, D. G., 1998. Post-trade transparency on nasdaq's national market system. Journal of Financial Economics 50, 231–252.

Roll, R., Schwartz, E., Subrahmanyam, A., 2010. O/s: The relative trading activity in options and stock. Journal of Financial Economics 96, 1 – 17.

Shim, J. J., Todorov, K., 2021. Etfs, illiquid assets, and fire sales.

Shum, P., Hejazi, W., Haryanto, E., Rodier, A., 2016. Intraday share price volatility and leveraged etf rebalancing. Review of Finance 20, 2379–2409.

Sircar, K. R., Papanicolaou, G., 1998. General black-scholes models accounting for increased market volatility from hedging strategies. Applied Mathematical Finance 5, 45–82.

Tuzun, T., 2014. Are leveraged and inverse etfs the new portfolio insurers? In: *Paris December 2014 Finance Meeting EUROFIDAI-AFFI Paper*.

Wilmott, P., Schönbucher, P. J., 2000. The feedback effect of hedging in illiquid markets. SIAM Journal on Applied Mathematics 61, 232–272.

Wu, Y., Jegadeesh, N., 2020. Closing auctions: Information content and timeliness of price reaction.

Online Appendix

Online Appendix A Cleaning of High-Frequency Data

To analyze intraday momentum effects in individual stocks, we rely on the NYSE Trades and Quote (TAQ) database for January 1996 through July 2019. Since the NYSE provides the raw tape of all trades performed on the included exchanges, multiple cleaning steps are required. Additionally, we merge the TAQ database with CRSP to use the PERMNO as a unique identifier per common share of any company.

A.1 Cleaning Procedure

We retain only trades with trade correction indicators "00" an "01", which refer to correctly recorded trades, and those that have been subsequently altered, but reflect the actual trade price at the time. We further keep only trades with trade sale corrections "Z", "B", "U", "T", "L", "G", "W", "K", and "J", as well as an accompanying "I" reflecting odd lot trades. To rely on trade prices during regular trading hours only, we discard all observations before 9:30 and after 16:01. We explicitly include the minute of 4pm, as many closing trades (denoted by sale conditions "6" or "M") fall within the first few seconds afterwards.

If multiple trades occur at exactly the same point in time, we take the median price as the "correct price". To make sure that prices are consistent by Ticker, we employ a bounce-back filter following Andersen, Bondarenko, and Gonzalez-Perez (2015), which effectively checks whether any trade's price deviates by more than 15 times the median absolute deviation of the day. If this is the case, we will kick this observation if we observe a reversion to the previous price within the next five minutes, or 10 trades, whichever encompasses more trades. We also drop price observations which deviate by more than two times $|\log(p_t/\hat{p})|$, where $\hat{p}$ is the median for the day. We choose trade-based filters to check for the internal consistency, instead of relying on the Quote database also provided by the NYSE, as some observations are falsely recorded in both. Afterwards, we span a minute-by-minute grid between 9:30 and 16:01 and map trades to these trading minutes, taking the volume-weighted average price within each minute to limit the impact of microstructure noise and single trades.

A.2 Merge with CRSP

The TAQ database provides intraday trade prices, but lacks information about distributions, mergers, and delistings. We obtain this information from CRSP, as the low-

frequency database in financial economics.

We use the PERMNO provided by CRSP as an identifier that is unique over time. In a first step, we merge the historical CUSIP by CRSP (NCUSIP) with the CUSIPs in the master file taken from TAQ. In some occasions, we cannot merge available tickers in the trade file this way, and in a second step merge the two databases by the root ticker and ticker suffix, which indicates different share classes. We keep only stocks with share codes 10 and 11, denoting common shares, as well as exchange codes 1, 2, 3, 31, 32 and 32, representing the NYSE, AMEX and NASDAQ, respectively. Using this procedure, we can merge most stocks for which we have intraday data available and cover most of the CRSP sample.

Online Appendix B Summary Statistics of the Underlying

Table B.1: **Industry Distribution of Underlying Stocks**
The table reports time-series averages of industry distributions of the Fama-French 12-industry classification. For comparability to the CRSP universe, the distribution of the full CRSP sample is included.

FF-12 Industry	Our sample	CRSP sample
Consumer durables	0.44%	0.19%
Consumer nondurables	4.57%	2.71%
Manufacturing	10.23%	4.78%
Energy	5.64%	3.4%
Chemicals	3.22%	1.26%
Business Equipment	15.68%	8.62%
Telecom	2.38%	1.95%
Utilities	3.73%	1.9%
Wholesale	11.75%	4.91%
Healthcare	10.64%	5.51%
Finance	10.31%	44.15%
Other	21.41%	20.62%

Table B.2: **Summary Statistics of Underlying Stocks**
The table reports summary statistics on the stock-day sample for the underlying stocks. Panel A reports the time-series summary statistics and Panel B reports the time-series average of cross-sectional distributions. Percent coverage of the stock universe (EW) is the number of stocks in the sample, divided by the total number of CRSP stocks. Percent coverage of the stock universe (VW) is the total market capitalization of sample stocks divided by the total CRSP market capitalization. Percent coverage of stocks traded at NYSE or AMEX is the number of stocks in the sample trading at NYSE or AMEX, divided by the total number of stocks. The firm size percentiles are computed using the full CRSP sample. Number of LETF is the number of LETF a stock is included in. The sample period is January 2012 – December 2019.

	Mean	Std	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: Time-Series Distribution							
Number of stocks in the sample each month	992.53	227.92	877.2	973.0	1043.0	1101.25	1140.0
Stock coverage of stock universe (EW)	13.86	3.17	12.43	13.65	14.57	15.26	15.92
Stock coverage of stock universe (VW)	48.79	10.17	48.47	49.76	50.92	51.85	52.72
Stock traded at NYSE or AMEX	38.74	8.07	35.78	37.91	40.19	42.88	43.42
Panel B: Time-Series Average of Cross-Sectional Distributions							
Firm size in million	15050	44338	454	1079	3152	11182	31883
Firm size CSRP percentile	77	17	51	66	81	92	96
Number of LETF	9	3	6	7	9	11	14

Online Appendix C Summary Statistics of Leveraged ETFs

Table C.1: **Summary Statistics of Leveraged ETFs**
The table reports time-series summary statistics on the underlying leveraged ETFs in our sample. Number of LETF is the number of leveraged ETFs. Number of benchmark indices is the number of unique benchmark indices underlying all leveraged ETFs. Aggregated AUM denotes the assets under management across all leveraged ETFs, in million. Aggregated leverage-adjusted AUM is the assets under management adjusted for the rebalancing leverage of each leveraged ETF. Percentage of inverse LETF (EW) is the number of inverse ETFs divided by the total number of leveraged ETFs. Percentage of inverse LETF (VW) is the assets under management weighted proportion of inverse ETFs to the total leveraged ETF sample. Average absolute leverage factor (EW) is the average absolute leverage factor. Average absolute leverage factor (VW) is the assets under management weighted absolute leverage factor. The sample period is January 2012 – December 2019.

	Mean	Std	10%	25%	50%	75%	90%
Number of LETF	71.67	11.05	66.0	72.0	74.0	76.0	79.0
Number of benchmark indices	23.72	3.7	21.0	23.0	25.0	25.0	27.0
Agg. AUM	17.71	4.69	12.03	15.29	16.92	21.33	24.43
Agg. leverage-adjusted AUM	95.33	28.55	64.23	73.01	86.97	121.12	131.71
Perc. of inverse LETF (EW)	45.29	3.38	41.67	43.24	45.07	45.95	51.28
Perc. of inverse LETF (VW)	33.14	15.66	16.31	22.75	28.48	41.77	62.64
Avg. absolute leverage factor (EW)	2.35	0.13	2.21	2.22	2.32	2.45	2.5
Avg. absolute leverage factor (VW)	2.43	0.14	2.29	2.35	2.44	2.53	2.6

Online Appendix D Correlation of Exchange Specific Γ^{HP} and Total Γ^{HP}

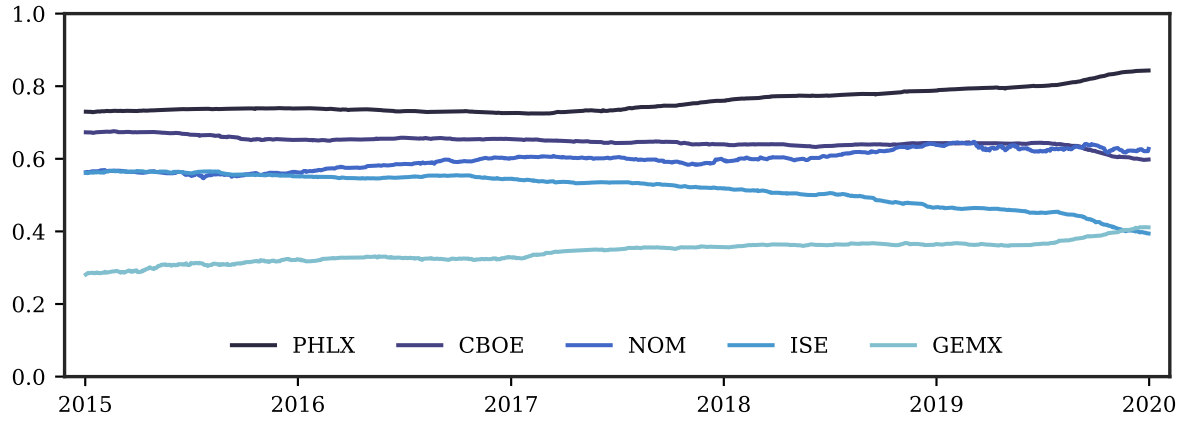


Fig. D.1. **Average Three-year Rolling Γ^{HP} Correlations**
The figure shows average three-year rolling correlations between Γ^{HP} based on the delta-hedgers option inventory of one single exchange (either PHLX, CBOE, NOM, ISE, or GEMX) and Γ^{HP} based on the pooled delta-hedgers option inventory across all exchanges in our sample.

Online Appendix E Different ADV Measures

We investigate whether the choice of the time period over which we calculate the average dollar volume (ADV) effects our results. We calculate ADV over the previous week (W) or previous quarter (W) and using the last half hour of a trading day (I) or the entire daily trading volume (D). As Table E.1 shows, the choice of denominator for Γ^{HP} and Ω^{LETF} has little impact on our results.

Table E.1: **Effect of Different ADV Measures**
The table reports the results to regressing returns in the last half hour of a trading day on returns until 15:30, r^{pre}, gamma hedging pressure Γ^{HP}, and leveraged ETF rebalancing flows Ω^{LETF} following Equation (9). We exchange the standard ADV measure as the average volume in the last half trading hour over the last month by similar measures using weekly (W) and quarterly horizons (Q), as well as measures using daily volume in their constructed (denoted by "-D"). T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}	(4) r_t^{end}	(5) r_t^{end}	(6) r_t^{end}
Γ_t^{HP}	-10.006***	-55.821***	-10.960***	-59.616***	-10.976***	-60.342***
	(-5.518)	(-7.094)	(-5.521)	(-6.949)	(-5.244)	(-6.763)
Ω_t^{LETF}	41.180***	177.953***	46.738***	190.533***	48.616***	201.118***
	(4.658)	(6.015)	(5.198)	(6.094)	(5.084)	(6.099)
r_t^{pre}	-0.817***	-0.803***	-0.820***	-0.804***	-0.824***	-0.809***
	(-7.430)	(-7.176)	(-7.452)	(-7.148)	(-7.501)	(-7.227)
Observations	1,882,332	1,882,332	1,882,332	1,882,332	1,882,332	1,882,332
R^2 (%)	0.363	0.380	0.370	0.382	0.376	0.391
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]
ADV	W-I	W-D	M-I	M-D	Q-I	Q-D

Online Appendix F Different Sets of Delta-Hedgers

In the main part of the paper, we assume that delta-hedgers are composed of market makers and broker/dealers. We exchange the set of delta-hedgers to include only market makers (Table F.1) or to include market makers, broker/dealers, and firm proprietary traders (Table F.2). The choice of the set of likely delta-hedgers has little impact on our results.

Table F.1: **Market Makers as the only Delta-Hedgers**
The table reports the results to regressing returns in the last half hour of a trading day on returns until 15:30, r^{pre} and gamma hedging pressure Γ^{HP}, following Equation (9). We assume that market makers are delta-hedgers. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}	(4) r_t^{end}	(5) r_t^{end}	(6) r_t^{end}
Γ_t^{HP}	-10.354***	-10.292***	-10.345***	-10.312***	-10.355***	-10.355***
	(-3.094)	(-3.077)	(-3.093)	(-3.077)	(-3.094)	(-3.094)
r_t^{pre}	-0.739***	-0.741***	-0.739***	-0.739***	-0.739***	-0.739***
	(-7.034)	(-7.050)	(-7.036)	(-7.039)	(-7.035)	(-7.034)
IV_{t-1}		4.517***				
		(5.029)				
$\hat{RV}_t^{pre}$			20.913			
			(1.159)			
$\hat{RV}_t^{end}$				418.390***		
				(3.143)		
PC_{t-1}					-0.014	
					(-0.092)	
$O/S_{t-1}^{\$}$						244.126
						(0.770)
Observations	1,882,332	1,882,332	1,882,236	1,882,236	1,879,827	1,882,327
R^2 (%)	0.267	0.286	0.269	0.277	0.267	0.268
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

Table F.2: **Market Makers, Broker/Dealer, and Firm Proprietary Traders as Delta-Hedgers**
The table reports the results to regressing returns in the last half hour of a trading day on returns until 15:30, r^{pre} and gamma hedging pressure Γ^{HP}, following Equation (9). We assume that market makers, broker/dealer, and firm proprietary traders are delta-hedgers. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}	(4) r_t^{end}	(5) r_t^{end}	(6) r_t^{end}
Γ_t^{HP}	-11.014***	-10.977***	-11.005***	-10.976***	-11.014***	-11.012***
	(-6.965)	(-6.945)	(-6.968)	(-6.947)	(-6.965)	(-6.963)
r_t^{pre}	-0.706***	-0.708***	-0.706***	-0.707***	-0.706***	-0.706***
	(-6.817)	(-6.834)	(-6.818)	(-6.821)	(-6.817)	(-6.817)
IV_{t-1}		4.516***				
		(5.076)				
$\hat{RV}_t^{pre}$			18.816			
			(1.034)			
$\hat{RV}_t^{end}$				410.953***		
				(3.101)		
PC_{t-1}					-0.010	
					(-0.063)	
$O/S_{t-1}^{\$}$						280.615
						(0.832)
Observations	1,882,332	1,882,332	1,882,236	1,882,236	1,879,823	1,882,327
R^2 (%)	0.296	0.314	0.297	0.306	0.296	0.296
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

Online Appendix G Γ^{HP} based on Older Positions

Endogeneity concerns regarding Equation (9) might be raised given that traders with public and/or private information on the return of stock j in the last 30 minutes, r_j^{end}, choose the option market to exploit this information. To mitigate any concerns, we follow Ni et al. (2020). Precisely, we split Γ^{HP} in two parts: gamma hedging pressure originating from older positions held by delta-hedgers at time $t-\tau$, and gamma hedging pressure stemming from new positions between $t-\tau$ and t. Ni et al. (2020) argue that option positions which existed at $t-\tau$ cannot be established due to private and/or public information after the close at $t-\tau$. In case of short-lived private and/or public information, positions at $t-\tau$ are of no use for predicting the return in the last 30 minutes on t, r^{pre}.

To decompose Γ^{HP}, we first define the gamma imbalance for stock j at time t which is based on option positions of delta-hedgers at time $t-\tau$ as

$$\Gamma^{IB}_{j,t;\tau} = \left(\sum_{o=1}^{N^t_{t-\tau}} netOI_{o,t-1-\tau} \times \Gamma_o(t-1, S^{close}_{j,t-1}) \times S^{15:30}_{j,t} \times Mult_o \right) \times \frac{S^{close}_{j,t-1}}{100} \times \frac{1}{\text{ADV}^{\text{end}}_{j,t-1}}, \tag{G.1}$$

where $N^t_{t-\tau}$ denotes the number of options contracts on stock j that are available at time $t-\tau$ and expire after t. The difference between Equation (G.1) and Equation (4) in the main paper is that Equation (G.1) uses the net open interest of likely delta-hedgers at time $t-\tau$ and sums over options expiring after t. However, both definitions use the gamma of option contracts at the previous trading day, $t-1$. Hence, Equation (G.1) denotes the gamma imbalance at time t of old positions.

Next, we define the gamma hedging pressure due to old positions as

$$\Gamma^{HP}_{j,t;\tau} = 100 \times \Gamma^{IB}_{j,t;\tau} \times r^{\text{pre}}_{j,t}. \tag{G.2}$$

We decompose Γ^{HP} at time t into the part that is due to old positions existing at time $t-\tau$, $\Gamma^{HP}_{j,t;\tau}$, and the part due to new positions, $\Gamma^{HP}_{j,t;\text{new}}$, as follows

$$\Gamma^{HP}_{j,t} = \Gamma^{HP}_{j,t;\tau} - \Gamma^{HP}_{j,t;\text{new}}, \tag{G.3}$$

where $\Gamma^{HP}_{j,t;\text{new}} = \Gamma^{HP}_{j,t} - \Gamma^{HP}_{j,t;\tau}$.

Finally, we run the following specification where we control for fixed effects and other control variables, $\boldsymbol{X}_{j,t}$,

$$r^{\text{end}}_{j,t} = \beta_0 r^{\text{pre}}_{j,t} + \beta_1 \Gamma^{HP}_{j,t;\tau} + \beta_2 \Gamma^{HP}_{j,t;\text{new}} + \boldsymbol{\gamma}' \boldsymbol{X}_{j,t} + FE_j + FE_t + \epsilon_{j,t}. \tag{G.4}$$

We hypothesize that β_1 is negative and statistically significant. Table G.1 and Table G.2 show results with τ set to five and ten business days, respectively. Both tables confirm our hypothesis.

Table G.1: **Option Hedging Pressure based on 5-Business Day Old Positions**
The table summarizes the results of regressions of returns in the last half hour of a trading day on gamma hedging pressure based on five business day old positions $\Gamma^{HP}_{t;\tau}$ and hedging pressure based on new positions, $\Gamma^{HP}_{t;\text{new}}$, after controlling for returns until 15:30 (r^{pre}), as in specification Equation (G.4). IV_{t-1} denotes implied volatility at time $t-1$. $\hat{RV}^{end}_t$ denote the square root of predicted realized variance for the time period from 15:30 to 16:00. PC_{t-1} is the put-call-ratio and $O/S^{\$}_{t-1}$ denotes the option-to-stock volume in dollar terms. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r^{end}_t	(2) r^{end}_t	(3) r^{end}_t	(4) r^{end}_t	(5) r^{end}_t	(6) r^{end}_t
$\Gamma^{HP}_{t;\tau}$	-16.375***	-16.324***	-16.370***	-16.366***	-16.376***	-16.372***
	(-6.954)	(-6.942)	(-6.954)	(-6.947)	(-6.954)	(-6.955)
$\Gamma^{HP}_{t;\text{new}}$	-1.107***	-1.117***	-1.110***	-1.124***	-1.107***	-1.104***
	(-2.601)	(-2.638)	(-2.611)	(-2.664)	(-2.600)	(-2.590)
r^{pre}_t	-0.810***	-0.813***	-0.811***	-0.811***	-0.810***	-0.810***
	(-6.765)	(-6.776)	(-6.767)	(-6.767)	(-6.766)	(-6.765)
IV_{t-1}		5.133***				
		(4.964)				
$\hat{RV}^{pre}_t$			22.224			
			(1.058)			
$\hat{RV}^{end}_t$				435.201***		
				(2.795)		
PC_{t-1}					0.064	
					(0.299)	
$O/S^{\$}_{t-1}$						283.201
						(0.899)
Observations	1,015,379	1,015,379	1,015,333	1,015,333	1,014,433	1,015,375
R^2 (%)	0.341	0.363	0.342	0.351	0.341	0.341
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

Table G.2: **Option Hedging Pressure based on 10-Business Day Old Positions**
The table summarizes the results of regressions of returns in the last half hour of a trading day on gamma hedging pressure based on ten business day old positions $\Gamma^{HP}_{t;\tau}$ and hedging pressure based on new positions, $\Gamma^{HP}_{t;\text{new}}$, after controlling for returns until 15:30 (r^{pre}), as in specification Equation (G.4). IV_{t-1} denotes implied volatility at time $t-1$. $\hat{RV}^{end}_t$ denote the square root of predicted realized variance for the time period from 15:30 to 16:00. PC_{t-1} is the put-call-ratio and $O/S^{\$}_{t-1}$ denotes the option-to-stock volume in dollar terms. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

Dependent	(1) r^{end}_t	(2) r^{end}_t	(3) r^{end}_t	(4) r^{end}_t	(5) r^{end}_t	(6) r^{end}_t
$\Gamma^{HP}_{t;\tau}$	-17.973***	-17.912***	-17.972***	-17.959***	-17.973***	-17.973***
	(-6.714)	(-6.694)	(-6.715)	(-6.710)	(-6.713)	(-6.715)
$\Gamma^{HP}_{t;\text{new}}$	-1.234***	-1.244***	-1.237***	-1.251***	-1.234***	-1.231***
	(-2.777)	(-2.814)	(-2.787)	(-2.839)	(-2.776)	(-2.765)
r^{pre}_t	-0.821***	-0.823***	-0.821***	-0.821***	-0.821***	-0.821***
	(-6.837)	(-6.848)	(-6.839)	(-6.838)	(-6.838)	(-6.837)
IV_{t-1}		5.138***				
		(4.969)				
$\hat{RV}^{pre}_t$			22.482			
			(1.069)			
$\hat{RV}^{end}_t$				435.447***		
				(2.791)		
PC_{t-1}					0.064	
					(0.302)	
$O/S^{\$}_{t-1}$						287.581
						(0.910)
Observations	1,015,379	1,015,379	1,015,333	1,015,333	1,014,433	1,015,375
R^2 (%)	0.332	0.354	0.334	0.343	0.332	0.333
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]

Online Appendix H Scaled Returns

Instead of using simple returns and value-weighted observations, Moskowitz et al. (2012) propose the use of scaled returns, which expresses returns in terms of units of expected risk:

$$\hat{r} = \frac{r}{\sigma_r}, \tag{H.1}$$

where σ_r is calculated using an exponentially-weighted moving average from the realized variance of using 5-minute squared returns from last close to 15:30. The half life is chosen to equal 60 days.

As shown in Table H.1, scaling returns does not alter our results.

Table H.1: **Using Scaled Returns**
The table reports the results to regressing returns in the last half hour of a trading day on returns until 15:30, r^{pre} and gamma hedging pressure Γ^{HP}, following Equation (9). The regression setup follows that of Table 2 but uses scaled returns. T-statistics in parentheses are derived from standard errors clustered by date and entity. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects and weight returns by the stock's market capitalization. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}
Ω_t^{LETF}		165.089***	173.968***
		(4.936)	(5.233)
Γ_t^{HP}	-29.157***		-34.042***
	(-3.501)		(-4.167)
r_t^{pre}	-0.060***	-0.068***	-0.063***
	(-7.644)	(-8.695)	(-7.948)
Observations	1,882,236	1,882,236	1,882,236
R^2 (%)	0.487	0.536	0.563
Entity FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]

Online Appendix I Mechanical Flows

Index versus Stock Specific Effects

While Γ^{HP} rebalancing activity depends on the return of the underlying stock, the rebalancing amount for leveraged ETFs depends on the return of the benchmark index of the leveraged ETF. We exploit this unique structure of the rebalancing flows from leveraged ETFs, by conditioning on those stocks for which the return until hedging begins is near zero (below 10 basis points in absolute terms). Given that the impact of rebalancing flows is mechanical, we still expect an effect of Ω^{LETF} in roughly the same magnitude as for the full sample. This approach allows us to investigate the impact of leveraged ETF rebalancing without being conflated by other intraday return phenomena (Heston et al., 2010; Lou et al., 2019; Bogousslavsky, 2020), or option rebalancing in the form of Γ^{HP}. The results in Table I.1 confirm our prior.

Table I.1: **Near Zero Returns of the Underlying Stock**
The table reports the results to regressing the returns in the last half hour of a trading day on the Leveraged ETF rebalancing quantity Ω^{LETF}. We have included only stock-day observations for which the absolute return from the previous day's close until 15:30 of stock j is below 10 basis points in absolute terms. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

	(1)
Dependent	r_t^{end}
Ω_t^{LETF}	69.400***
	(4.961)
r_t^{pre}	-0.899
	(-0.434)
Observations	120,913
R^2 (%)	0.119
Entity FE	Yes
Time FE	Yes
SEs	[t;j]

What is the role of news on fundamentals?

To test that delta-hedging effects are unrelated to fundamental news on the underlying stock, we perform sample splits on earnings announcements and material news. Information on earnings announcement days is obtained from Compustat and I/B/E/S. Whenever the announcement date for the same stock differs between Compustat and I/B/E/S, we follow Dellavigna and Pollet (2009) and use the earlier date. Compustat and I/B/E/S are matched to our CRSP data via the matching algorithms provided by WRDS. We use the Dow Jones version of Ravenpack News Analytics and its sentiment scores to identify days with significant news for each underlying. We restrict our news sample to articles that are most relevant for a particular stock, i.e. a relevance score of 100. Furthermore, we only include news that is highly positive (sentiment score above 0.75) or highly negative (sentiment score below 0.25).

We identify the two weeks centered around earnings announcements and the releases of material news as indicated by RavenPack for each stock in our sample. The estimated coefficients for stocks with and without material news releases or earnings announcements in Table I.2 barely differ, confirming the mechanical nature of Γ^{HP}-flows.

Table I.2: **Impact of Fundamental Information**
The table reports the results to regressing the returns in the last half hour of a trading day on returns until 15:30, r^{pre}, and gamma hedging pressure Γ^{HP}. The regression results are reported for several subsamples where we focus on or exclude days with fundamental information, either earnings announcements (EA) or fundamental news releases identified by RavenPack. Specification (1) uses the subsample for which Compustat and/or I/B/E/S do not report earnings announcements (EA). Specification (2) excludes day-asset observations for which Compustat and/or I/B/E/S report earnings announcements. In specification (3) we exclude asset-day observations for which RavenPack documents either at least one negative (score $\geq$ 25) or one positive (score $\leq$ 75) news appearance, whereas specification (4) includes only these observations. Whenever earnings announcements or fundamental news are released on day t for asset j, we exclude also a window of 5 trading days around t for asset j. t-statistics are in parentheses below. T-statistics are in parentheses below and are computed using time-and-entity-clustered standard errors. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects in all specifications and value-weight observations. The sample period is January 2012 – December 2019.

	(1)	(2)	(3)	(4)
Dependent	r_t^{end}	r_t^{end}	r_t^{end}	r_t^{end}
Γ_t^{HP}	-9.865***	-8.117*	-10.551***	-8.671***
	(-5.557)	(-1.879)	(-6.635)	(-2.780)
r_t^{pre}	-0.749***	-0.744***	-0.662***	-0.850***
	(-6.907)	(-4.296)	(-6.824)	(-5.168)
Observations	1,607,439	274,893	1,567,086	315,246
R^2 (%)	0.263	0.335	0.219	0.390
Entity FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]
Subsample	Excluding EA	Only EA	Excluding News	Only News

Online Appendix J Effects Across Industries

Investor attention for stocks in respective sectors changes over time. To compare how hedging pressure from the options market impacts end-of-day returns, we sort stocks into their respective industry following the classification on Kenneth French's website.

Table J.1 displays results. $\Gamma^{IB}-$ and $\Omega^{LETF}-$effects are present and statistically significant in all industries.

Table J.1: **Effects in Different Industries**
The table reports the results to regressing returns in the last half hour of a trading day on returns until 15:30, r^{pre} and gamma hedging pressure Γ^{HP}, following Equation (9). The sample is split by the industry classification based on SIC codes, following Kenneth French's website, `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. T-statistics in parentheses are derived from standard errors clustered by date and entity. ***, **, * denotes significance at the 1%, 5%, 10% level. We include entity fixed effects and weight returns by the stock's market capitalization. The sample period is January 2012 – December 2019.

Dependent	(1) r_t^{end}	(2) r_t^{end}	(3) r_t^{end}	(4) r_t^{end}	(5) r_t^{end}
Γ_t^{HP}	-8.777***	-6.822***	-6.919*	-20.417***	-16.849***
	(-5.212)	(-3.145)	(-1.814)	(-4.111)	(-5.734)
Ω_t^{LETF}	31.260***	21.224***	47.491***	95.533***	38.054***
	(6.586)	(3.642)	(8.304)	(6.488)	(4.318)
r_t^{pre}	-0.680***	-0.839***	-0.978***	-0.959***	-0.654***
	(-5.585)	(-4.495)	(-4.630)	(-4.689)	(-3.491)
Observations	346,342	427,144	366,477	201,183	541,186
R^2 (%)	0.269	0.295	0.430	0.748	0.308
Entity FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
SEs	[t;j]	[t;j]	[t;j]	[t;j]	[t;j]
Industry	Consumer	Manuf.+Energy	Business	Health	Other

Bibliography

Andersen, T. G., Bondarenko, O., Gonzalez-Perez, M. T., 2015. Exploring return dynamics via corridor implied volatility. Review of Financial Studies 28, 2902–2945.

Bogousslavsky, V., 2020. The Cross-Section of Intraday and Overnight Returns.

Dellavigna, S., Pollet, J. M., 2009. Investor Inattention and Friday Earnings Announcements. Journal of Finance 64, 709–749.

Heston, S. L., Korajczyk, R. A., Sadka, R., 2010. Intraday patterns in the cross-section of stock returns. Journal of Finance 65, 1369–1407.

Lou, D., Polk, C., Skouras, S., 2019. A tug of war: Overnight versus intraday expected returns. Journal of Financial Economics 134, 192–213.

Moskowitz, T. J., Ooi, Y. H., Pedersen, L. H., 2012. Time series momentum. Journal of Financial Economics 104, 228 – 250, special Issue on Investor Sentiment.

Ni, S. X., Pearson, N. D., Poteshman, A. M., White, J., 2020. Does Option Trading Have a Pervasive Impact on Underlying Stock Prices? Review of Financial Studies 34, 1952–1986.

Chapter 2

Credit Variance Risk Premiums

*Manuel Ammann, Mathis Mörke**

ABSTRACT

This paper studies variance risk premiums in the credit market using a novel data set of swaptions quotes on the CDX North America Investment Grade and High Yield indices. We find that returns of credit variance swaps are negative and economically large, irrespective of the credit rating class. Shorting credit variance swaps yields annualized Sharpe ratios well above their counterparts in other asset classes. The returns remain highly statistically significant when accounting for transaction costs and cannot be explained by established risk-factors and structural model variables. By means of corridor variance swaps, we also dissect the overall variance risk premium into receiver and payer variance risk premiums. We show that credit variance risk premiums are mainly driven by the payer corridor, which is associated with worsening macro-economic conditions.

Keywords: option-implied moments, option-implied skewness, downside risk

JEL classification: C14, G11, G12, G13, G14.

*We thank Torben Andersen, Marc Arnold, Nils Friewald, Joachim Grammig, Jens Jackwerth, Axel Kind, Angelo Ranaldo, Stefan Ruenzi, Christian Schlag, Stace Sirmans, Erik Theissen, Viktor Todorov, Michael Weber, and Florian Weigert, as well as the participants of the 7th annual meeting of the Paris Financial Management Conference 2019, the SoFiE Financial Econometrics Summer School 2019, of the doctoral workshop at the 26th annual meeting of the German Finance Association 2019, the Brown Bag Seminar of the University of St.Gallen, the 2019 joint seminar session of the University of St.Gallen and the University of Konstanz, and the 2018 Topics in Finance Seminar in Davos for insightful and constructive comments.

1. Introduction

The analysis of variance risk and its associated market price has gained a lot of interest over the last one and a half decades. Carr and Wu (2009) document in their seminal work strongly negative variance risk premiums in equity markets. Similar findings have been documented in most asset classes. However, variance risk is much less understood in credit markets than in other asset classes. This is surprising given that credit derivatives constitute a non-negligible part of the global over-the-counter derivatives market. Moreover, credit volatility products have sparked interest by investment banks and the asset management industry during the last years.

In case of equity markets, negative variance risk premiums have been reconciled with exposure to rare disasters (see Dew-Becker, Giglio, Le, and Rodriguez, 2017). Previous literature has shown that credit default swaps (CDS) feature similar characteristics (Kelly, Manzo, and Palhares, 2019; Wang, Zhou, and Zhou, 2013) and can be explained by similar underlying principles (Gabaix, 2012; Gourio, 2013; Bai, Goldstein, and Yang, 2020). From this perspective, the existence and magnitude of a risk premium related to variability in CDS spreads is ambiguous. Moreover, it remains a priori unclear if variance risk premia differ for various credit rating classes. Prior literature has found that a large market price of risk for jumps is needed to explain investment grade CDS spreads, whereas high yield CDS spreads are more sensitive to diffusive risks (Huang and Huang, 2012; Huang, Nozawa, and Shi, 2019; Bai et al., 2020), hence pointing towards notable differences between the two rating classes.

To the best of our knowledge, we are the first to study variance risk premiums in credit markets, drawing upon a new data set of options written on two of the most prominent corporate credit indices in the world, the CDX North America Investment Grade and High Yield 5 Year indices. We use the notion of synthetic variance swaps to assess the sign, significance and magnitude of credit variance risk premiums (CVP, hereafter) over the period from 2012 to 2021. In a variance swap, the ex-post realized variance is exchanged for a fixed variance swap rate, set at swap inception. The difference between the swap rate and the ex-post realized variance constitutes the variance risk premium. We use a model-free estimation approach to explicitly avoid any stand on microfoundations. Moreover, we follow latest developments for calculating realized variance to avoid biases due to discretization errors and jumps (see Neuberger, 2012; Bondarenko, 2014).

We find large and significant variance risk premiums in credit markets. Moreover, CVP is equally high across credit rating classes. An investment strategy capturing CVP

by means of shorting 1-month variance swaps yields average monthly excess returns of 0.58% (0.73%) for investment grade (high yield) credit. Our result are also quantitatively the same across rating classes in terms of risk-adjusted performance measures. The associated annualized Sharpe ratios are well above 0.8.

After having established our main results, we aim at uncovering more details on CVP: What drives the richness in CVP? How does it compare to other trading strategies capturing volatility risk in credit and other asset classes. Can CVP be explained by standard asset pricing risk-factors or structural model variables? Are there commonalities to equity and fixed-income variance risk premiums?

In order to shed light on the sources of CVP, we use corridor variance risk premiums following Andersen and Bondarenko (2010) and Kaeck (2018). In corridor variance swaps, price moves of the underlying contribute only to realized variance if they happen to be within a specific price corridor. We decompose CVP into payer and receiver variance risk. Payer (receiver) variance risk premiums are associated with price moves of the CDS spread above (below) the CDS spread observed at swap inception. Therefore, payer variance risk premiums are associated with bad economic states, receiver variance risk premiums with good economic states of the world. We find that both corridor variance risk premiums are significantly different from zero and sizeable. However, payer variance risk premiums are larger in absolute magnitude compared to receiver variance risk premiums.

To address the second question, we first contrast CVP with another popular trading strategy exposed to variance risk, namely, investing in at-the-money straddles. Additionally, we compare CVP to shorting equity and fixed-income variance swaps. None of the strategies generate similar pure return or risk-adjusted performances. As our data set spans the onset of the Covid-19 pandemic, we are able to compare CVP with risk premiums in other asset classes across different market regimes. Precisely, we find that, though shorting credit variance swaps have generated negative returns during the first half of 2020, its performance quickly regained its relative differential to variance strategies in other asset classes in the second half of 2020, as evident before 2020.

Subsequently, we show that CVP cannot be explained by structural model variables. Moreover, we find that CVP, and also corridor variance risk premiums remain largely unaltered when we consider standard asset pricing factors, such as the Fama and French (2018) six-factor equity model, augmented by the intermediary capital risk factor of He, Kelly, and Manela (2017), or the four-factor model of Bai, Bali, and Wen (2019) for corporate bonds.

Next, we analyse the comovement of CVP with equity, fixed-income, and variance-of-

variance risk premiums. We show that CVP loads significantly and positively on equity variance risk premiums. However, the latter are not sufficient to fully explain CVP.

Finally, our analysis yields credit implied volatility indices as by-products. We study the informational content of credit implied variance and its term structure. Credit implied variance possesses strong predictive power for financial stress, measured by the St. Louis Fed Stress Index.

Our paper is related to various strands of the literature. First, our results extend findings by Kita and Tortorice (2021) to the index level whose instruments are the most liquid and most traded credit derivatives in the world. Investors pay high risk premiums to be insured against credit volatility risk. In contrast to Kita and Tortorice (2021), our analysis does not need to rely on a structural model, as it is completely model-free and employs credit swaptions. Although we confirm the author's findings that credit variance risk premiums are larger than equity variance risk premiums, we do not find support that this relation is inversely related to the leverage ratio.

Second, we add to the research on documenting variance risk premiums in other asset classes. Carr and Wu (2009) has been the first to document negative and economically large variance risk premiums, focusing on equity markets. Ammann and Buesser (2013) adopt the approach and study foreign exchange variance risk premiums. Trolle and Schwartz (2010) and Prokopczuk, Symeonidis, and Wese Simen (2017) analyze commodity markets, and Trolle and Schwartz (2014) and Choi, Mueller, and Vedolin (2017) focus on fixed income. In line with other asset classes, we document strongly negative credit variance risk premiums. However, credit variance risk premiums dwarf their analogue in all other asset classes. Rather, our findings suggest that credit variance risk premiums serve as catastrophic risk premiums. Credit index tranches are conceptually similar, but not equal to credit index swaptions. Previous research has shown that these instruments also contain catastrophic or systemic risk premiums. Longstaff and Rajan (2008) estimate that 8% of the credit index spread is attributable to an systemic event in which the majority of firms go bankrupt. Seo and Wachter (2018) show that a model with time-varying probability of economic disaster can explain credit index tranches reasonably well.

Moreover, our findings complement research on general risk premia and trading strategies in credit markets. Longstaff, Pan, Pedersen, and Singleton (2011) dissect sovereign credit default spreads into their risk premium and a default risk component. Both components are strongly related to global macroeconomic factors. More recently, Kelly et al. (2019) show that the risk premium in single-name CDS is mainly driven by the compensation for facing volatility risks. Contrary to our findings, the authors show that Sharpe

ratios on single-name CDS returns are depending on the underlying firm's leverage ratio. CDS of low leverage firms are higher than for high leverage firms. Additionally, the findings in this paper also relate to Jarrow, Li, Ye, and Hu (2019) who explore mispricings in the term structure of CDS spreads by means of a reduced-form credit risk model. Their trading strategy involving market-neutral portfolios of misvalued single firm CDS contracts yields on average Sharpe ratios of around one, close to the Sharpe ratios we document for shorting credit variance swaps.

Our paper also adds to the extant literature analysing the determinants of credit spreads. The theoretical foundation for empirically investigating the drivers of credit spreads is the structural model of Merton (1974), in which a firm's default probability is depending on the firm's leverage, equity volatility, and the level of the risk-free rate. Ericsson, Jacobs, and Oviedo (2009) confirm the predictions of Merton's model and show that equity volatility, firm leverage, and changes in interest rate levels suffice to explain the time and cross-sectional variation in single-name CDS spreads. Pires, Pereira, and Martins (2015) provide further evidence that high-risk firms are more sensitive to changes in the above explanatory variables than low-risk firms by means of quantile regressions. Contrary to the previous literature, credit variance risk premiums are not explained by structural model variables.

Given the theoretical link between credit spreads and equity volatility in Merton (1974), a large body of research has concentrated on the connection of CDS spreads to equity volatility. Cremers, Driessen, Maenhout, and Weinbaum (2008) incorporate implied alongside realized volatility in their regression model. They document an improved estimation fit when including implied volatilities from firms' individual options in explaining CDS levels. Cao, Yu, and Zhong (2010) refine the use of option implied information. As the authors argue, CDS exhibit similar features to out-of-the-money put options as both offer effective protection against downside risks. They find that firms' put option-implied volatility is superior to historical volatility. On the one hand, the authors relate their finding to the ability of implied volatility to predict future realized volatility. On the other hand, they find that volatility risk premium in option prices co-moves with CDS spreads. Another strand of research has looked at lead-lag-relationships between equity and credit markets. Norden and Weber (2009) document that equity returns are leading CDS spreads, whereas the opposite does not hold true. Moreover, CDS are more sensitive to the stock market than to bond markets. Our results confirm these findings as credit variance risk premiums co-move strongly with equity variance risk premiums. This co-movement is stronger than for fixed-income variance risk premiums.

Moreover, our analysis produces credit implied volatility indices as by-products. Equity implied volatilities have gained a prominent role in academia and practice. The VIX, computed from options on the S&P 500, is known as the fear gauge due to its tendency to increase in times of negative equity market returns (Whaley, 2009). It exhibits predictive power for future economic activity and financial stability, as documented by Bekaert and Hoerova (2014). Attention has been also paid to the term structure of implied volatilities. Aït-Sahalia, Karaman, and Mancini (2020) show that in case of equities it is upward sloping. Its counterpart in Treasury bond markets is, however, downward sloping and its slope exhibits predictive power for economic growth and economic stress, as reported by Choi et al. (2017). Similarly to equity markets, the term structure of credit implied volatilities is downward sloping. Moreover, we find that it exhibits predictive power for financial stress and systemic risk.

The remainder of the paper is structured as follows. Section 2 details our methodological framework. In section 3, we present our data set. Section 4 describes and discusses our main empirical findings, whereas section 5 concludes.

2. Methodology

2.1. Credit Default Index Swaptions

A credit default swap (CDS) is a contract giving its holder the right to sell a bond for its face value in case of default or otherwise specified credit event by the bond issuer. Consequently, a CDS index (CDIS) tracks the cost of buying protection for each company in a portfolio. A buyer of protection on a CDIS receives insurance against losses stemming from defaults of any of the index's single-name constituents during the lifetime of the contract. In exchange, the protection buyer pays a fixed coupon and a cash upfront payment to the protection seller at trade inception.

Generally, CDIS can be either spread or price quoted. If CDIS are price quoted, they are effectively regarded as a credit risky floating rate note. Investment grade CDIS are usually quoted as an index spread, that is, the spread which the index should be priced if it were a simple CDS contract (O'Kane, 2008). In practice, it is standard to use a reduced-from model, the ISDA standard model[1], to convert back and forth between spread and price quotation. In the remainder, we assume that CDIS are spread-quoted.

[1]We refer to the homepage of the ISDA standard model, `https://www.cdsmodel.com/`, for documentation on it.

In the following, we use the notation of Rutkowski (2012). Let $T_0 < T_1 < \cdots < T_m$ denote the tenor structure of a forward-start CDIS, where $T_0 = T$ is the inception date, T_m is the maturity date, and T_j is the jth fee payment date for $j = 1, \ldots, m$. We further assume that the recovery rate $\delta \in [0, 1]$ is predetermined, constant and the same for all reference entities in the CDIS. More generally, we assume that firms are identical. We denote by B the process for the money-market account which is assumed to be strictly positive. Let n be the initial number of reference entities and τ_i the time at which entity i defaults, then $J_t = n - \sum_{i=1}^{n} \mathbb{1}_{\{t \geq \tau_i\}} = \sum_{i=1}^{n} \mathbb{1}_{\{\tau_i > t\}}$ denotes the number of entities which have survived up to time t.

We will later consider options written on a forward CDIS with expiry $T = T_0$. Their pricing will be based on a change of numéraire argument. For the remainder, we assume that the event $\{\tau_n \leq T\}$ has zero probability under $\mathbb{Q}$. This ensures that the candidate numéraire is strictly positive and does not vanish with non-zero probability. Economically, the technicality rules out the default of all reference entities before swaption maturity. Morini and Brigo (2011) refer to the default of all entities as the "Armageddon" event. Incorporating the event into option prices requires knowledge of the risk-neutral conditional distribution of τ_n, which can be estimated via index tranches. The authors show in a case study surrounding the unfolding of the financial crisis in 2007 that neglecting the "Armageddon" event only materially effects longer term option prices when the default correlation is above 90%.[2]

From the perspective of the CDIS seller, the value of the forward CDIS with spread κ at time $t \in [0, T_0]$ is

$$S_t^n(\kappa) = \mathbb{E}_t^{\mathbb{Q}}(P_t^n) - \kappa \mathbb{E}_t^{\mathbb{Q}}(A_t^n),$$

where

$$\begin{aligned} P_t^n &= (1-\delta) B_t \sum_{i=1}^{m} B_{\tau_i}^{-1} \mathbb{1}_{\{T_0 < \tau_i \leq T_m\}} \\ A_t^n &= B_t \sum_{j=1}^{m} \alpha_j B_{T_j}^{-1} J_{T_j}, \end{aligned}$$

with $\alpha_j = T_j - T_{j-1}$ for every $j = 2, \ldots, m$ and $\mathbb{E}_t^{\mathbb{Q}}$ denotes the risk-neutral expectation given the information available up to time t. P_t^n is the discounted payoff of the protection

[2]Precisely, the authors use nine-month options, whereas the focus here is on one-month options for which the "Armageddon" effect should be less severe.

leg, whereas A_t^n is the discounted payoff of the fee leg per one unit of the spread.[3] The latter is also known as risky annuity or defaultable price value of the basis point. As A_t^n is not directly observable in the market, it is market convention to a) compute the implied risk-neutral probabilities for different maturities using a flat single-name CDS curve with a constant spread equalling κ_t^n and b) assume that all firms are identical. Therefore, we can approximate A_t^n by $J_t PV_t(\kappa_t^n)$, where $PV_t(\kappa_t^n)$ is the risky annuity calibrated by a flat CDS curve with spread equalling the actually quoted CDIS spread in the market.

An option on a CDIS, also referred to as a swaption, is an agreement to sell or buy protection on the underlying index with a certain maturity at a pre-agreed spread. Swaptions are European options. The option holder of a payer index option has the right, but no obligation, to buy protection in the index at the strike spread. On the contrary, a receiver option entitles its holder the right to sell protection in the index at the strike spread.

Assume that the CDIS was initiated at time 0, with constant spread κ_0^n. κ_T^n denotes the prevailing market spread at time T. $\mathcal{P}_T(t,\kappa)$ represents the time-t value of a European payer swaption with maturity $T = T_0$ and strike spread κ. At expiration, the swaption has a payoff of

$$\left(PV_T(\kappa_T^n) J_T(\kappa_T^n - \kappa_0^n) - nPV_T(\kappa)(\kappa - \kappa_0^n) + L_T\right)^+, \tag{1}$$

where $L_t = (1-\delta)\sum_{i=1}^n \mathbb{1}_{\{\tau_i \le t\}}$ denotes the loss process of the CDIS. We would like to point out several noteworthy aspects with respect to Equation (1). First, unlike forward CDIS, swaptions provide protection from losses occurring between option inception and option maturity T. Secondly, $PV_T(\kappa)$ denotes the risky annuity at time T and hence is random in the interest rates. On the contrary, $PV_T(\kappa_T)$ is random in interest rates and the index spread κ_T. We follow market convention and approximate future interest rates by current forward rates for the same date. Consequently, randomness in $PV_T(\kappa)$ vanishes and $PV_T(\kappa_U)$ remains to be random with respect to κ_T only. We can rewrite the swaption payoff in Equation (1) to

$$\left(S_T^n(\kappa) + L_T\right)^+ = \left(S_T^a(\kappa)\right)^+, \tag{2}$$

where $S_T^a(\kappa) = S_T^n(\kappa) + L_T$ denotes the loss-adjusted forward CDIS and we have used the approximation $PV_T(\kappa)n \approx PV_T(\kappa_T)J_T \approx A_T^n$ (see Section 8.5.3 Rutkowski, 2012). The

[3]We suppress the dependence of A_t^n on the inception date of the forward-start CDIS, $T_0 = T$, for ease of notation.

price of the loss-adjusted forward CDIS at time $t \in [0, T]$ is given by

$$S_t^a(\kappa) = \mathbb{E}_t^{\mathbb{Q}}(P_t^a) - \kappa \mathbb{E}_t^{\mathbb{Q}}(A_t^n),$$

where $P_t^a = P_t^n + B_t B_T^{-1} L_T$. Consequently, the loss-adjusted fair forward CDIS spread at any time $t \in [0, T]$ is the random variable $\kappa_{t,T}^a$ which solves $S_t^a(\kappa_{t,T}^a) = 0$. For every $t \in [0, T]$,

$$\kappa_{t,T}^a = \frac{\mathbb{E}_t^{\mathbb{Q}}(P_t^a)}{\mathbb{E}_t^{\mathbb{Q}}(A_t^n)},$$

and the price of the forward CDIS can be expressed as, for every $t \in [0, T]$

$$S_t^a(\kappa_t^a) = A_t^a(\kappa_{t,T}^a - \kappa). \tag{3}$$

By inspection of Equations (2) and (3), we see that through a suitable change of numéraire (see Section 8.5.3.3 Rutkowski, 2012), the price of a payer option is given as

$$\mathcal{P}_T(t, \kappa) = \mathrm{A}_\mathrm{t}^\mathrm{a}\, \mathbb{E}_t^{\mathbb{A}} \left[(\kappa_{t,T}^a - \kappa)^+ \right], \tag{4}$$

where $\mathbb{A}$ denotes the risky annuity measure associated with $\mathrm{A}_\mathrm{t}^\mathrm{a}$ as numéraire. The corresponding receiver swaption, denoted by $\mathcal{R}_T(t, \kappa)$, has a time-t price of

$$\mathcal{R}_T(t, \kappa) = \mathrm{A}_\mathrm{t}^\mathrm{a}\, \mathbb{E}_t^{\mathbb{A}} \left[(\kappa - \kappa_{t,T}^a)^+ \right]. \tag{5}$$

Equations (4) and (5) illustrate that a payer swaption can be associated with a call option on the CDIS spread, whereas a receiver swaption resembles the notion of a put option.

2.2. Variance Swap Contracts

This subsection introduces variance swaps in credit markets. In a variance swap, the long side receives the difference between the realized variance of the CDIS spread over the life of the contract and a pre-determined rate K

$$(RV_{t,T} - K)\, A_T^a, \tag{6}$$

where $RV_{t,T}$ denotes the realized variance of $\kappa^a_{t,T}$ over a partition $\Pi = \{t = t_0 < \ldots\ t_n = T\}$ of the interval from t to T and is defined as

$$RV_{t,T} = 2 \times \left[\left(\frac{\kappa^a_{t_1,T}}{\kappa^a_{t_0,T}} - 1 - \log \frac{\kappa^a_{t_1,T}}{\kappa^a_{t_0,T}} \right) + \ldots + \left(\frac{\kappa^a_{t_n,T}}{\kappa^a_{t_{n-1},T}} - 1 - \log \frac{\kappa^a_{t_n,T}}{\kappa^a_{t_{n-1},T}} \right) \right]. \quad (7)$$

The definition in Equation (7) is based on Neuberger (2012) and Bondarenko (2014) and has been adopted frequently in the recent literature (see Choi et al., 2017; Kaeck, 2018). The popularity of realized variance as in Equation (7) stems from two facts. First, its risk-neutral expectation can be calculated in a model-free way if $\kappa^a_{t,T}$ follows a martingale, and hence, the latter is not restricted to be a continuous process. Second, results are robust to discretization biases. The inclusion of the risky annuity in Equation (6) is adapted from Mele and Obayashi (2015).

The fixed variance swap rate is determined as the value for K such that the variance swap has zero value at t, it solves

$$\mathbb{E}^{\mathbb{Q}}_t \left[e^{-\int_t^T r(s)ds} \left(RV_{t,T} - K \right) A^a_T \right] = 0,$$

and is given by

$$K \equiv IV_{t,T} = \mathbb{E}^{\mathbb{A}}_t \left[RV_{t,T} \right] = \frac{2}{A^a_t} \int_0^\infty \frac{\mathcal{M}_T(t,\kappa)}{\kappa^2} d\kappa, \quad (8)$$

where

$$\mathcal{M}_T(t,\kappa) = \begin{cases} \mathcal{R}_T(t,\kappa) & \text{if } \kappa \leq \kappa^a_t \\ \mathcal{P}_T(t,\kappa) & \text{if } \kappa > \kappa^a_t \end{cases} \quad (9)$$

To shed further light on the sources of credit variance risk premiums, we follow the ideas of Andersen and Bondarenko (2010) and Kaeck (2018) and decompose (implied and realized) return variation into price ranges by means of corridor variances. Generally, corridor variance is constructed as

$$IV^{B_u,B_d}_{t,T} = \frac{2}{A^a_t} \int_{B_d}^{B_u} \frac{\mathcal{M}_T(t,\kappa)}{\kappa^2} d\kappa, \quad (10)$$

where $\mathcal{M}_T(t,\kappa)$ as in Equation (9), and B_d (B_u) denotes the lower (upper) bound of the price corridor. The realized counterpart is given as

$$RV^{B_u,B_d}_{t,T} = \sum_{i=1}^{n} g(\kappa^a_{t_i,T}) - g(\kappa^a_{t_{i-1},T}) - g'(\kappa^a_{t_{i-1},T}) \cdot \left(\kappa^a_{t_i,T} - \kappa^a_{t_{i-1},T} \right), \quad (11)$$

where

$$g(x) = \begin{cases} 2 \times \left(-\log B_u - \frac{x}{B_u} + 1 \right) & \text{if } x > B_u \\ -2 \times \log x & \text{if } x \in [B_d, B_u] \\ 2 \times \left(-\log B_d - \frac{x}{B_d} + 1 \right) & \text{if } x < B_d \end{cases}$$

and $g'(x)$ denotes the derivative of g with respect to x. Model-invariance and robustness to discretization are retained by Equation (11) (see Bondarenko, 2014). We focus on the two most intuitive price ranges, namely upside and downside variances. Upside (downside) variance is the return variation when the current forward CDIS spread is above (below) the starting forward CDIS spread. More precisely, in the former case we have $B_d = 0$ and $B_u = \kappa^a_{t_0,T}$, whereas in the latter $B_d = \kappa^a_{t_0,T}$ and $B_u = \infty$. Since Equation (10) breaks into either payer or receiver swaptions, we term the semi-variances payer and receiver variances.

Equations (8) and (10) require a continuum of option prices. However, only a finite number of strikes is available. In line with Carr and Wu (2009), we obtain and sort all out-of-the money swaptions for each date t and maturity T. Subsequently, we construct a grid of 2000 implied volatility points within a strike range of ± 8 standard deviations from the current forward price. Thereby the standard deviation is approximated by the average implied volatility. Third, we linearly interpolate across moneyness. For strikes higher (lower) than the highest (lowest) listed strike price, we use the implied volatility at the highest (lowest) available strike. Finally, we convert the implied volatilities back into option prices using Black (1976).

3. Data

Credit swaptions data are obtained from Markit. Markit is considered the leading financial services and pricing data company in the area of credit default swaps. We focus on two of the most prominent credit indices, the CDX North America Investment Grade Index (CDX IG, hereafter) and the CDX North America High Yield Index (CDX HY). Markit is owner and administrator of both indices. CDX IG comprises 125 of the most liquid North American entities with investment grade credit rating, whereas CDX HY includes 100 of the most liquid North American entities with high yield credit ratings. As CDX HY is quoted on price, contrarily to CDX IG, we use the ISDA CDS standard model to compute equivalent spread quotes from price quotes.

Table 1: **CDX North America Options Data Set**
This table summarizes information on options written on the CDX North America Investment Grade and High Yield indices. The options data is grouped into buckets depending on the time to maturity. For each bucket, we report the number of days with observed quotes, the average and median time to maturity, moneyness level, and number of strikes per maturity. Options are written on the on-the-run indices.

		Maturity Bucket						
		Days ≤ 30	$30 <$ Days ≤ 60	$60 <$ Days ≤ 90	$90 <$ Days ≤ 120	$120 <$ Days ≤ 150	$150 <$ Days ≤ 180	Days > 180
Panel A: Investment Grade								
# Active Days		2095	2098	2173	2084	1732	1257	725
Time To Maturity	Mean	17.24	45.56	75.00	105.60	135.49	164.23	203.83
	Median	16	47	76	105	135	163	198
Moneyness	Mean	0.86 - 1.51	0.81 - 1.73	0.78 - 1.83	0.76 - 1.88	0.75 - 1.92	0.75 - 1.87	0.75 - 1.72
	Median	0.88 - 1.34	0.83 - 1.62	0.80 - 1.72	0.78 - 1.78	0.76 - 1.84	0.76 - 1.79	0.76 - 1.70
Number of Strikes	Mean	12.30	16.85	18.67	19.05	19.04	17.42	14.22
	Median	11	15	17	17	17	17	14
Panel B: High Yield								
# Active Days		2007	2042	2108	2039	1630	1149	548
Time To Maturity	Mean	17.17	45.39	74.91	105.53	135.20	163.61	202.26
	Median	16	44	75	105	134	162	197
Moneyness	Mean	0.90 - 1.36	0.86 - 1.51	0.84 - 1.56	0.82 - 1.57	0.82 - 1.57	0.82 - 1.53	0.81 - 1.43
	Median	0.91 - 1.31	0.87 - 1.48	0.85 - 1.52	0.83 - 1.54	0.82 - 1.54	0.83 - 1.50	0.81 - 1.43
Number of Strikes	Mean	12.37	16.94	18.28	18.32	18.12	16.40	14.22
	Median	12	16	18	18	18	16	14

In 2017, the average daily trading volume in CDX IG (CDX HY) options are 5.3 billion USD (1.6 billion USD), whereas about 60% of the trading is in receiver options for both indices (Toublan, 2017).[4] Turning to market participants and their trading volume share, Toublan (2017) estimate that 29% of the trading volume is due to hedge funds, followed by banks with 22% in 2016. Other notable market participants are asset managers (20%) and broker/dealer (18%). Moreover, whereas banks' share has shrunken over time (2013: 33%), asset managers and hedge funds seem to become more dominant (combined share 2013: 39% vs. 2016: 49%).

As credit swaptions are traded over-the-counter, Markit receives quotes on payer and receiver options from associated buy-side and sell-side accounts. After assuring data quality, Markit provides composite payer and receiver quotes per option maturity and strike. The composite quotes are formed by averaging the latest quotes over the tightest time window possible on each day ensuring a sufficient amount of quotes. Collin-Dufresne et al. (2020a) estimate at least 10 daily trades per day at the beginning of our sample for swaptions on CDX IG. Towards the end, the number has more than doubled and reached a maximum of 88 during the onset of the Covid-19 pandemic.

The data are available at a daily frequency from March 2012 to September 2021. The maturity date is the third Wednesday of each month. Data span the on-the-run series and the first off-the-run series of the five-year contracts, being the most liquid (see Collin-Dufresne, Junge, and Trolle, 2020b).[5] As there is considerably less trading in options on off-the-run indices, we restrict our sample to on-the-run series. Data for the underlying reference indices are also provided by Markit. Furthermore, we keep only days where each of the four nearest maturities carry at least three strikes. We furthermore verify option implied forward spreads with their theoretical value using the ISDA CDS standard model. Our filters leave us with 437'686 (401'900) receiver and payer swaptions for CDX IG (CDX HY).

Table 1 provides summary statistics for our data set by reporting the number of days with observed quotes, the average and median of the time to maturity, moneyness levels,

[4]Whereas trading volumes in single-name CDS has steadily fallen, trading volume in CDX IG and CDX HY have remained stable over our sample period (Toublan, 2017; Collin-Dufresne, Junge, and Trolle, 2020a).

[5]Every sixth months, precisely on CDS IMM dates in March and September, CDIS are rolled, i.e. a new index is issued. The new issue is known under an incremented series. The new series of the index is termed on-the-run index until a new series is created at the next roll date. It then becomes the first off-the-run index. The constituents of a specific series remain constant over the life-time of the CDIS. However, new issues might have different constituents compared to the previous issue. Moreover, the coupon is usually kept constant across all issues of an index. It amounts to 100 (500) basis points for investment grade (high yield).

and number of strikes. Statistics are reported for maturity buckets, each spanning 30 days. As Table 1 shows, the market for options with time to maturity between 30 and 120 days is most liquid. CDX IG swaptions seem to be more traded than CDX HY swaptions, gauging by the number of active days. However, the difference is relatively small. Moreover, the moneyness range is generally skewed towards higher strike prices, e.g., the average moneyness range reads $0.81 - 1.73$ for CDX IG swaptions with time to maturity between 30 and 60 days. Finally, across all maturity buckets, we observe on average on any given day at least 11 different strikes. Though this seems to be little as compared to equity markets, it is noteworthy, that first studies on variance risk premiums on single stocks report number of strikes well below our numbers (see Carr and Wu, 2009).

Our analysis on theoretically motivated determinants on credit variance risk premiums follows Collin-Dufresne, Goldstein, and Martin (2001). First, we construct a composite leverage ratio for each credit index. Therefore, we take stock prices and shares outstanding for each constituent of the CDX IG and HY indices from the Center for Research in Security Prices. Quarterly balance sheet and accounting data is obtained from Compustat. We treat market leverage as book debt (the sum of Compustat items Long-Term Debt - Total and Debt in Current Liabilities) over the sum of book debt and the market value of equity. We assume that accounting variables become publicly available after one quarter. The market value of equity is calculated as the product of the number of common shares outstanding and share price. For estimating the jump component, we resort to implied volatilities from S&P 500 index options, taken from OptionMetrics.

Finally, we collect data on S&P 500, VIX index and the exchange traded fund TLT options from OptionMetrics. The options are European-style in case of S&P 500 and VIX and written on the spot indices. Options on TLT are American-style and we use the method of Barone-Adesi and Whaley (1987) to convert them to European options. For every day in our sample, we use put-call parity for the at-the-money option pair for which put and call prices are closest and back out implied forward prices.

4. Empirical Results

4.1. Preliminary Data Analysis

We start by analysing annualized implied volatility of CDX IG and CDX HY (CIV IG and CIV HY, hereafter), defined as $\sqrt{(T-t)IV_{t,T}}$. We proceed as follows to construct

these measures for constant maturities of 45, 75, and 105 days.[6] At each day in our sample, we straddle the time to maturity of interest and compute implied variances for the neighbouring maturities. Subsequently, we interpolate linearly in implied variance to obtain a value at a fixed time to maturity. We have chosen 45 days as the shortest time to maturity instead of 30 days out of two reasons. First, we thereby avoid extrapolation as we observe for some days in our sample swaptions with maturities strictly above 30 days. Secondly, we disregard options with less than 7 days to maturity in the calculation of CIV. We do not consider time to maturities beyond 105 days since they are less actively traded.

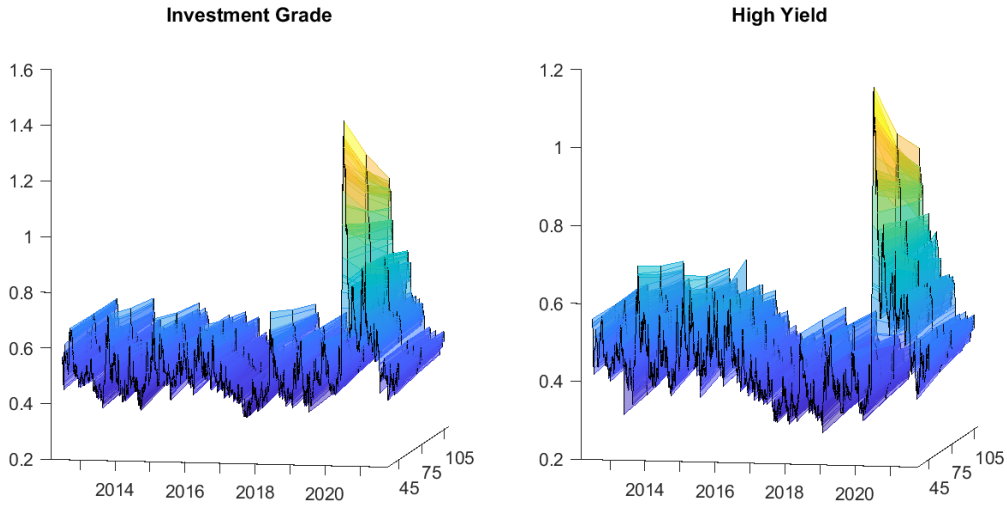


Fig. 1. Credit Implied Volatility
The figure shows the time series of credit implied volatility for three different constant maturities (45, 75, and 105 days) over the sample period from March 2012 to September 2021. Implied volatility is the annualized square-root of implied variance; that is $\sqrt{(T-t)IV_{t,T}}$, where $IV_{t,T}$ is defined in Equation (8).

Figure 1 displays credit implied volatility for the three time to maturities and both credit ratings. First, credit implied volatility exhibits strong time-variation. The lowest levels are attained around mid-2017 after a period of slow, but gradually decline. At this point in time, CIV IG (CIV HY) for 45 days is around 35% (30%). Maximums in CIV levels are reached at at the onset of the Covid-19 pandemic in 2020, where credit implied volatility is above 100%. Second, the term structure of implied volatility is most of the

[6]Chen, Doshi, and Seo (2020) use the same data set to construct a VIX type corporate bond implied volatility index by means of synthetic defaultable floating rate notes. As the VIX index, their corporate bond VIX constitutes the expected 1-month risk-neutral volatility.

time upward sloping. This is best illustrated between mid-2016 and mid-2017. Third, we observe several noteworthy spikes besides 2020. They appear to be around the US taper tantrum in 2013 and the "implosion" of inverse equity volatility products in early 2018. During both periods, credit implied volatility levels across all maturities and credit ratings nearly double within a few days. Furthermore, during periods of high stress, the term structure inverts. Fourth and finally, the steepness of the term structure is a function of market calmness: the more quiet the market, the steeper the term structure, as seen during the first half of 2017.

Table 2: **Summary Statistics on Credit Implied Volatility**
This table reports the summary statistics of credit implied volatilities for three maturities (45, 75, and 105 days). Columns (2)-(7) report the mean, standard deviation, skewness, kurtosis, minimum, and maximum. Levels and first differences of log values of the three implied volatility time series are considered. The sample period runs from April 2012 to September 2021.

Maturity	Mean	Std	Skewness	Kurtosis	Min	Max
Panel A: Investment Grade						
45	0.51	0.12	3.04	17.30	0.34	1.41
75	0.52	0.11	2.69	13.91	0.36	1.24
105	0.52	0.10	2.44	11.69	0.38	1.10
Panel B: High Yield						
45	0.44	0.11	2.09	10.55	0.26	1.15
75	0.45	0.10	1.68	7.79	0.27	0.99
105	0.45	0.09	1.50	6.48	0.30	0.92

Table 2 presents summary statistics for credit implied volatility. The unconditional means of implied volatility reflect the on-average upward sloping term structure, corroborating findings in equity markets (Aït-Sahalia et al., 2020), but being in contrast to Treasury bond implied volatility (Choi et al., 2017). However, the steepness of the term structure of unconditional implied volatilities is gentle, as the unconditional difference between 45 and 105 days average implied volatility amounts to 1%. Furthermore, the minimum level of implied volatility is strictly increasing in time to maturity, whereas the opposite holds true for the maximum. This points towards the inversion of the term structure during times of market distress. The smaller standard deviation for longer maturities adds to that point. Moreover, we observe that all CIV indices exhibit positive skewness and kurtosis, albeit monotonically declining in time to maturity.

As Figure 1 documents, CIV IG attains higher levels of implied volatility than CIV

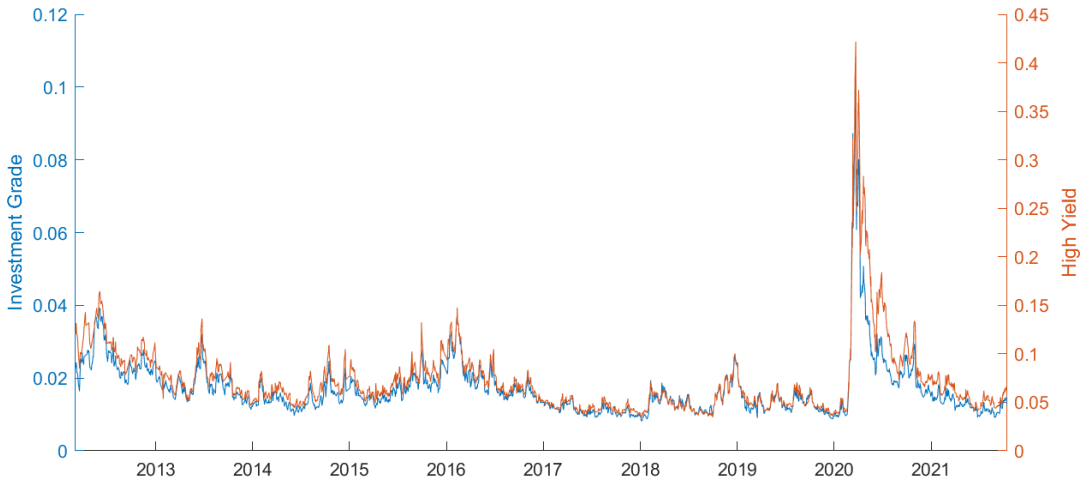


Fig. 2. Credit 45-day Price-based Implied Volatility
The figure shows the time series of credit 45-day price-based implied volatility for investment grade (left-hand side) and high yield (right-hand side) over the sample period from March 2012 to September 2021. Price-based implied volatility is derived from spread implied volatility (see Equation (8)) following Duembgen et al. (2015).

HY during the onset of Covid-19 and also at other occasions during our sample. This may seem counterintuitive. CIV IG and HY are based on spread volatilities. Spread volatility, being the market standard to quote credit volatilities (Duembgen et al., 2015), does not inevitably increase with the riskiness of the underlying CDS. Price volatility, however, shows a positive relation between volatility and riskiness of the underlying CDS. Appendix B empirically illustrates this behavior in case of realizied volatilities. Figure 2 displays 45-day price-based implied volatility which is calculated from spread-based volatility according to Duembgen et al. (2015). As the figure shows, the time-series dynamics of price-based CIV IG and HY are highly correlated, but price-based CIV HY is higher than CIV IG. The former reaches lows (highs) of approximately 5% (40%), whereas the latter records 1% (10%) as its low (high).

4.2. Credit Variance Risk Premiums

After having established characteristics of credit implied volatility, we turn to credit variance risk premiums. We refer to the difference between the (ex post) realized and the implied variance in the variance swap, multiplied by the risky annuity, as the credit variance risk premium (CVP). Our definition resembles the payoff of a long position in

a variance swap. Precisely, it is the profit and loss for each \$1 notional investment in the variance swap contract. Our chosen approach is similar to the existing literature on variance risk premiums, though it differs by the inclusion of the risky annuity in the contract design.[7]

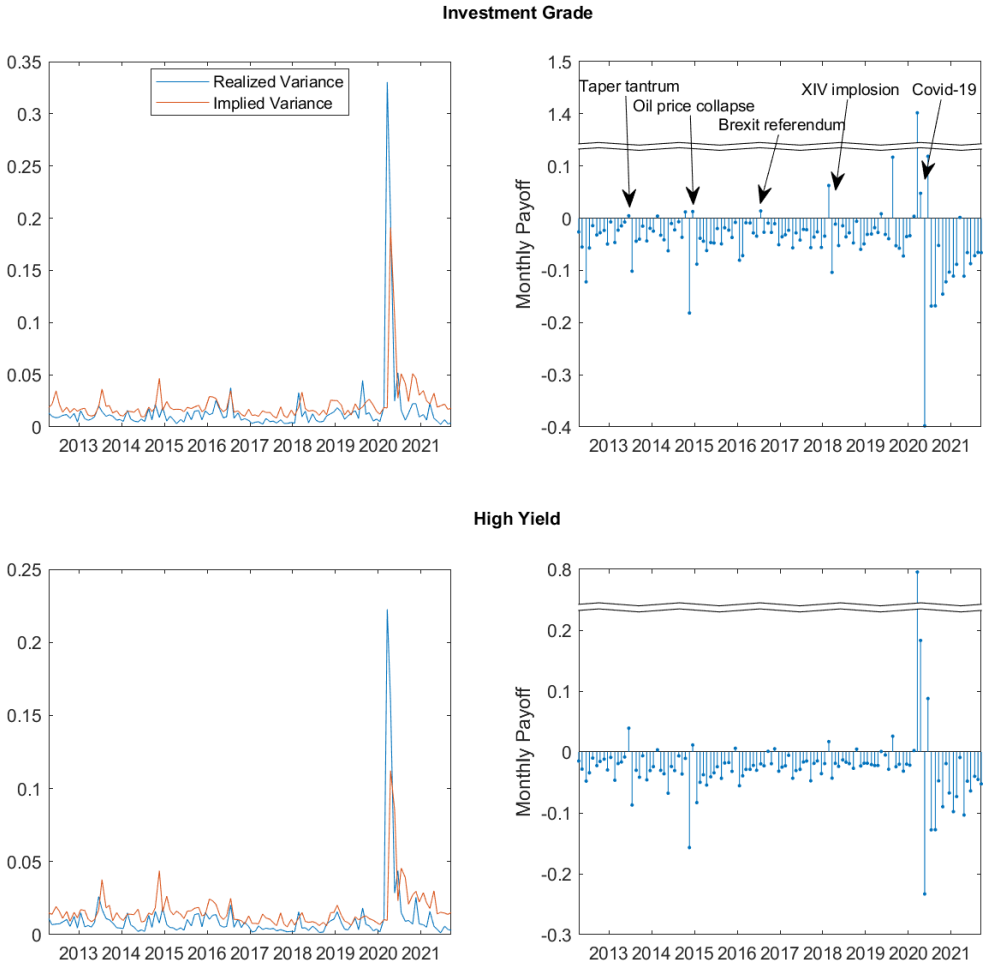


Fig. 3. Credit Variance Risk Premiums
The left graph depicts the monthly realized variances $RV_{t,T}$ and the implied variances $IV_{t,T}$ constructed from CDX North America Investment Grade index options. The right graph shows the corresponding variance swap payoffs. Each month the payoff is computed such that T corresponds to the expiry date of the Investment Grade index options. The sample period is from March 2012 to September 2021.

[7]Closest to our setting, Trolle and Schwartz (2014) include an annuity factor in case of variance and skewness swaps in the interest rate swap market.

We measure CVP by means of realized excess returns on synthetic variance swap contracts. Since variance swaps are zero net present value investments, we assume the existence of an initial capital amount, X, required to enter the swap. Hence, the excess return on variance swaps from t to T is given by

$$\frac{(RV_{t,T} - IV_{t,T}) \times A_T^a}{X}. \tag{12}$$

We adjust X in order to generate a 10% unconditional volatility of the realized excess returns (2.887% per month), following Duarte, Longstaff, and Yu (2007) and Trolle and Schwartz (2014). Each month the return is computed such that T corresponds to the expiry date of CDX index options in the next calendar month and t is the trading day succeeding the expiry date of the current month. Hence, our implementation choice exploits the existence of traded credit swaptions and circumvents the computation of synthetic swaption prices. The floating leg in the variance swap relies on the existence on forward CDIS spreads over the time span of the variance swap contract. We obtain daily forward CDIS spreads via the ISDA model and subsequently calculate realized variance.

Table 3 summarizes the statistics for CVP. Panel A shows results for CVP based on CDX IG, whereas panel B reports results for CVP based on CDX HY. Shorting credit variance swaps yields a monthly average return of 0.58% (0.73%), which is highly significantly different from zero (t-statistic = 2.23 (= 2.56)). Our results corroborate recent findings of Kita and Tortorice (2021) that credit investors are willing to pay a large premium for hedging against volatility risk, but on the index instead of firm level. However, we do not find support on the index level that this is especially the case for low-levered firms as high yield CVP is larger in absolute magnitude than investment grade CVP.

Shorting credit variance swaps shows high negative skewness and high excess kurtosis. Investment strategies with these properties are often called "picking up pennies in front of a steamroller", as they generate positive returns most of the time, but occasionally experience tremendous losses. Figure 3 provides visual evidence. It presents the time series of the realized and implied variance. The realized variance is often well below its implied counterpart. Hence, long positions in variance swaps yield on average negative returns. Notably, during our sample, there are several occurrences for CDX IG where realized variance exceeds implied variance, leading to positive variance swap payoffs. Most positive payoffs can be traced back to periods of general distress, the severest being the onset of Covid-19. Notably, the magnitude of the spikes in late 2014, coinciding

Table 3: **Summary Statistics on Credit Variance Risk Premiums**
This table reports summary statistics of non-overlapping monthly excess returns on one-month variance and semi-variance swaps. On each month, the first trade date after swaption expiry is selected and returns on variance and semi-variance swaps are computed from swaptions with one month to expiry. We assume an initial amount of capital required to entire into a swap contract. We adjust this amount to achieve a 10% annualized excess return volatility. We also include statistics on excess returns of investments in at-the-money straddles and credit default swaps. Returns are displayed in percent. Columns (2)-(10) report the mean, initial amount of capital required per $1 notional of the strategy (Capital), standard deviation, maximum, skewness, kurtosis, Sortino ratio (So. R), Stutzer ratio (St. R) and Sharpe ratio (SR). Standard errors are computed using Newey and West (1987) accounting for autocorrelation and hetereoscedasticity with optimal bandwidth. Sharpe ratios are corrected for return-autocorrelation using the approach of Lo (2002). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 to September 2021.

Strategy	Mean	Capital	Std	Max	Skewness	Kurtosis	So. R	St. R	SR
Panel A: Investment Grade									
CVP	-0.58**	5.07	2.89	27.67	8.17	82.28	-0.59	-0.17	-0.83
	(-2.23)								
CVP Payer	-0.43*	5.06	2.89	28.23	8.56	87.11	-0.49	-0.13	-0.77
	(-1.88)								
CVP Receiver	-0.40*	1.84	2.89	22.95	4.57	42.16	-0.28	-0.13	-0.75
	(-1.93)								
ATM straddle	-0.12	-	2.89	29.30	9.40	96.35	-0.29	-0.04	-0.16
	(-0.45)								
CDS	-0.17	0.20	2.89	21.40	2.94	31.35	-0.09	-0.06	-0.35
	(-0.79)								
Panel B: High Yield									
CVP	-0.73**	3.04	2.89	26.17	7.11	68.05	-0.66	-0.21	-0.86
	(-2.56)								
CVP Payer	-0.69***	2.86	2.89	28.19	8.86	89.53	-0.91	-0.19	-0.92
	(-2.58)								
CVP Receiver	-0.22	1.08	2.89	20.63	3.31	31.83	-0.13	-0.07	-0.41
	(-1.14)								
ATM straddle	-0.05	-	2.89	28.98	9.03	91.29	-0.11	-0.02	-0.06
	(-0.19)								
CDS	-0.50*	0.84	2.89	19.68	2.52	23.21	-0.27	-0.16	-0.83
	(-1.94)								
Panel C: Other Asset Classes									
SPX	-0.09	0.19	2.89	27.89	8.75	85.88	-0.13	-0.03	-0.16
	(-0.38)								
VIX	-0.24	3.05	2.89	21.03	4.57	31.90	-0.27	-0.08	-0.51
	(-1.02)								
TLT	-0.20	0.09	2.89	25.28	5.78	62.71	-0.13	-0.06	-0.44
	(-0.96)								

with the sell-off in crude oil prices, are very specific to credit markets and have not been witnessed in equity markets.[8] Moreover, positive payoffs for investment grade credit variance swaps are generally accompanied by positive payoffs in high yield credit variance swaps, underlying a high degree of commonality between both credit rating classes.

To gauge the magnitude of CVP, we put it into perspective of variance risk premiums in other asset classes. Carr and Wu (2009) and Choi et al. (2017) study equity and fixed-income variance risk premiums. Both studies document large and statistically significant premiums. Panel C in Table 3 reports results for variance risk premiums in several asset classes over our sample period. Previous studies' findings cannot be confirmed for our sample period. The average return of S&P 500 variance swaps amounts to 0.09%, whereas variance-of-variance and fixed-income variance swaps yield roughly -0.24% and -0.20%, respectively. None of the three are statistically significant. Additionally, CVP remains superior in comparison to interest rate variance swap premiums. Trolle and Schwartz (2014) analyze interest rate variance swaps by means of EUR and USD dominated swaptions. They document mean monthly returns to shorting variance between -0.40% and -0.66%.

Another popular strategy for gaining exposure to volatility are at-the-money straddles.[9] Coval and Shumway (2001) document attractive Sharpe ratios above one for trading straddles on S&P 500 index futures. We synthetically construct at-the-money straddles with the same time to maturity as the variance swap. As panels A and B in Table 3 show, shorting investment grade (high yield) at-the-money straddles yield average monthly returns of 0.12% (0.05%), however, being not statistically significant from zero.

Finally, we put CVP into perspective of direct exposure to credit risk. We implement a long credit risk strategy by entering into a forward-starting CDX index, paying the fixed leg and receiving the floating leg. As shown in Table 3, the monthly average return amounts to -0.17% in case of CDX IG, being not statistically significantly different from zero. Contrarily, long credit risk in CDX HY yields average monthly returns of -0.50%, being statistically significant (t-statistic $= -1.94$). Consequently, exposure to direct credit risk is less attractive than compensation for bearing variance risk by entering into variance swaps.

So far we have concentrated on average monthly returns. Table 3 also reports standard

[8] In unreported results, we have visually compared the time series of returns to credit variance swaps with variance swaps on the S&P 500 index.

[9] Selling (at-the-money) straddles is an alternative option strategy for volatility exposure. Choi et al. (2017) argue that this exposure is, however, non-linear. Furthermore, selling straddles leaves exposure to other factors.

risk-adjusted performance measures. Returns to shorting investment grade (high yield) credit variance swaps yield a Sharpe ratio of 0.83 (0.86).[10] As Sharpe ratios rest on the assumption of normality and returns to short credit variance swaps exhibit high positive skewness and kurtosis, we also consider the Sortino and Stutzer Ratio. Both measures confirm the superiority of credit variance swaps.

To shed light on the richness of CVP, we adapt the framework of corridor variance swaps and study their investment returns (Andersen and Bondarenko, 2010; Kaeck, 2018). Precisely, we focus on payer and receiver corridors. The payer (receiver) corridor is determined by the prevailing forward CDIS spread at variance swap initiation as the lower (upper) bound. It is associated with rising (falling) CDIS spreads. Table 3 includes summary statistics for corridor variance swaps. Overall, shorting investment grade payer or receiver or high yield payer corridor variance swaps earn highly negative and statistically significant returns. Findings are more pronounced for returns of payer variance swaps. The latter yield an average monthly return of 0.43% in case of CDX IG, whereas it is 0.69% for CDX HY. The average return for shorting receiver investment grade variance swaps is 0.40% and 0.22% for CDX HY. The return differences are also reflected in risk-adjusted performance measures, as receiver variance swaps appear as more successful investment strategies compared to payer variance swaps. The difference in corridor risk premiums is explained by the investor's search for economic downside protection (Duembgen et al., 2015). In equity markets, demand for downside protection is sought by the purchase of puts. In credit markets, its analogue is manifested in high demand for payer swaptions, causing a positive skew in the volatility surface and leading to the dearness of payer variance risk premiums.

We challenge our findings with three robustness checks presented in the appendix. First, as option trading strategies show non-Gaussian behaviour, we mitigate any concerns on statistical significance by employing a studentized bootstrap approach in Appendix C. Second, the results presented so far have focused on 1-month investment returns. Appendix D varies the investment horizon of variance swap contracts. CVP remain highly attractive considering investment periods of up to 4 months. Third and finally, Appendix E analyzes the impact of bid and ask spreads on the profitability of credit variance swaps. We find that shorting variance remains highly attractive after conservative transaction costs.

[10]We adjust Sharpe ratios for return-autocorrelation when aggregating them to annualized numbers following Lo (2002).

Table 4: **Credit Variance Risk Premiums across Market Regimes**
This table reports summary statistics of non-overlapping monthly excess returns on one-month variance and semi-variance swaps. On each month, the first trade date after swaption expiry is selected and returns on variance and semi-variance swaps are computed from swaptions with one month to expiry. We assume an initial amount of capital required to entire into a swap contract. We adjust this amount to achieve a 10% annualized excess return volatility. We also include statistics on excess returns of investments in at-the-money straddles and credit default swaps. Returns are displayed in percent. We consider three market regimes: before and after the onset of Covid-19 and the onset of Covid-19. The latter being defined as time from February to June 2020. For each sub-period, we report the mean, maximum return and Sharpe ratio. Standard errors are computed using Newey and West (1987) accounting for autocorrelation and hetereoscedasticity with optimal bandwidth. Sharpe ratios are corrected for return-autocorrelation using the approach of Lo (2002). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively.

	Pre Covid-19			Covid-19			Post Covid-19		
Strategy	Mean	Max	Sharpe Ratio	Mean	Max	Sharpe Ratio	Mean	Max	Sharpe Ratio
Panel A: Investment Grade									
CVP	-0.65***	2.31	-4.50	4.63	27.67	1.19	-1.87***	0.03	-6.07
	(-10.78)			(1.16)			(-7.75)		
CVP Payer	-0.50***	2.49	-4.75	3.29	28.23	0.80	-1.29***	0.75	-8.89
	(-8.13)			(0.78)			(-8.24)		
CVP Receiver	-0.43***	3.78	-1.36	3.71	22.95	0.97	-1.62***	0.43	-3.28
	(-3.79)			(1.53)			(-4.37)		
ATM straddle	-0.29***	1.94	-1.50	5.54	29.30	1.44	-0.91***	0.16	-13.62
	(-3.37)			(1.48)			(-11.86)		
CDS	-0.23	4.62	-0.51	0.96	21.40	0.27	-0.17	1.34	-2.25
	(-1.35)			(0.28)			(-1.53)		
Panel B: High Yield									
CVP	-0.82***	1.29	-2.97	5.50	26.17	1.51	-2.22***	-0.30	-6.97
	(-10.87)			(1.46)			(-7.80)		
CVP Payer	-0.77***	2.05	-3.96	4.41	28.19	1.14	-1.90***	0.29	-10.10
	(-10.24)			(1.05)			(-10.78)		
CVP Receiver	-0.28**	2.83	-0.66	3.81	20.63	1.01	-1.22**	2.14	-2.61
	(-2.37)			(1.63)			(-2.23)		
ATM straddle	-0.26**	2.30	-0.77	5.44	28.98	1.43	-0.55***	0.42	-4.21
	(-2.42)			(1.47)			(-3.84)		
CDS	-0.44**	4.98	-1.03	0.61	19.68	0.18	-1.21**	0.85	-1.68
	(-2.52)			(0.15)			(-2.03)		
Panel C: Other Asset Classes									
SPX	-0.23***	3.18	-1.24	3.92	27.89	1.00	-1.29***	0.27	-4.06
	(-3.73)			(0.99)			(-3.67)		
VIX	-0.19	21.03	-0.46	1.23	11.77	0.66	-2.18***	-0.88	-7.67
	(-0.82)			(0.82)			(-7.31)		
TLT	-0.24***	1.53	-1.14	1.63	25.28	0.40	-1.03***	-0.20	-6.66
	(-3.71)			(0.42)			(-4.69)		

4.3. Credit Variance Risk Premiums across Different Market Regimes

The return distribution of selling credit volatility is highly negatively skewed, meaning that the strategy can face substantial losses. The latter is especially the case when

volatility suddenly rises. A prime example is the onset of the Covid-19 pandemic in early 2020. To investigate the performance of CVP across different market regimes, we divide our sample in three non-overlapping periods: the time before Covid-19, the onset of Covid-19 spanning five months (from February to June 2020), and the remainder.[11] Average and maximum returns, as well as Sharpe ratios, for each period are reported in Table 4, providing some interesting findings. First, CVP, either investment grade or high yield, is significantly different from zero before the onset of Covid-19. A similar finding can be observed for selling at-the-money straddles or equity and fixed-income volatility by means of variance swaps. During the onset of the Covid-19 pandemic, all considered short variance strategies experience extreme losses. However, variance strategies are able to quickly recover after abrupt losses. Precisely, they exhibit the ability to generate strong upside returns after times of turmoil as average returns for the post Covid-19 period are multiples of the pre Covid-19 period. Second and most importantly, Table 4 provides interesting results regarding the relative performance of shorting credit variance swaps compared to other asset classes. Credit variance risk premiums are dearer in absolute magnitude before the onset of Covid-19 compared to equity and fixed-income variance risk premiums, especially in risk-adjusted terms. On the contrary, shorting credit volatility performs slightly worse than shorting equity or fixed-income volatility during the onset of Covid-19. Strikingly, CVP returns to its relative richness after the onset. This finding extends Kita and Tortorice's (2021) conjecture to the index level that investors regard credit variance risk premiums as a better proxy for catastrophic risk premiums than equity or fixed-income variance risk premiums.

4.4. Structural Model Variables as Drivers

The literature analysing the determinants of credit spreads is substantial (Collin-Dufresne et al., 2001; Campbell and Taksler, 2003; Cremers et al., 2008; Pires et al., 2015). The theoretical motivation for most studies lies in the structural model of Merton (1974). In his model, a firm's default probability is depending on the firm's leverage, equity volatility, and the level of the risk-free rate. Ericsson et al. (2009) confirm the predictions of Merton's model and show that equity volatility, firm leverage, and changes in interest rate levels suffice to explain the time and cross-sectional variation in single-name CDS spreads. We consider the set of macroeconomic and financial variables suggested by Collin-Dufresne et al. (2001), nesting the variables employed by Ericsson et al. (2009),

[11] Our findings are robust to varying the time period of the onset of Covid-19.

Table 5: **Structural Model Variables as Drivers of the Variance Risk Premiums** This table reports the estimation results from regressing the excess returns on variance (CVP) and semi-variance swaps (CVP Payer, CVP Receiver) on structural model variables. We also consider excess returns to ATM straddels and direct investments in credit default swaps. ΔL is the change in the aggregated leverage ratio, Δ RF the change in the 10-year Treasury rate, $(\Delta RF)^2$ the squared change in the 10-year Treasury rate, ΔS the change in the slope of the yield curve, ΔVIX the change in market volatility measured by the VIX index, ΔJ the change in a jump component, and RM the return on the S&P 500 index. Column (9) reports the adjusted R^2 (R^2_{adj}) of the regression. t-statistics are given in parentheses and adjusted for heteroskedasticity and autocorrelation using Newey and West (1987). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 until December 2020.

Strategy	Intercept	ΔL	ΔVIX	ΔRF	$(\Delta RF)^2$	ΔS	ΔJ	RM	R^2_{adj}
Panel A: Investment Grade									
CVP	-0.89***	49.06**	0.05	0.94	0.45	-0.76	-5.73	0.13**	0.19
	(-6.20)	(2.34)	(1.37)	(1.15)	(0.26)	(-0.65)	(-1.51)	(2.01)	
CVP Payer	-0.68***	44.06***	0.03	0.88	1.76	-0.54	-4.77*	0.03	0.48
	(-6.30)	(2.78)	(1.16)	(1.42)	(1.33)	(-0.61)	(-1.66)	(0.53)	
CVP Receiver	-0.55**	-7.62	0.08	0.03	-3.12	-0.70	1.90	0.15	-0.00
	(-2.10)	(-0.20)	(1.37)	(0.02)	(-0.99)	(-0.33)	(0.28)	(1.37)	
Panel B: High Yield									
CVP	-0.88***	-0.24	0.01	0.56	2.60	-0.65	-2.15	-0.01	-0.03
	(-4.84)	(-0.02)	(0.34)	(0.53)	(1.12)	(-0.44)	(-0.42)	(-0.11)	
CVP Payer	-0.93***	9.58	0.02	1.22	4.19**	-1.11	-3.76	-0.06	0.21
	(-6.17)	(0.90)	(0.48)	(1.39)	(2.17)	(-0.91)	(-0.87)	(-0.85)	
CVP Receiver	-0.24	-14.05	0.03	-0.23	-3.61	-0.45	4.11	0.11	0.05
	(-0.92)	(-0.75)	(0.46)	(-0.15)	(-1.07)	(-0.21)	(0.54)	(0.94)	

as potential determinants of CVP. More precisely, we use changes in the leverage ratio (ΔL_t), in the 10-year Treasury rate (ΔRF), squared changes in the 10-year Treasury rate, $(\Delta RF)^2$, changes in the slope of the yield curve (ΔS), changes in the VIX index (ΔVIX), returns on the S&P 500 index (RM), and changes in a jump component ΔJ. We then estimate the following regression model

$$r_t = \alpha + \beta' \Delta F_t + \epsilon_t, \tag{13}$$

where r_t is the return of credit variance swaps (in month t) and

$$\Delta F_t = \Big[\Delta L_t \quad \Delta RF_t \quad (\Delta RF_t)^2 \quad \Delta S_t \quad \Delta VIX_t \quad RM_t \quad \Delta J_t\Big].$$

Table 5 provides results. Changes in the leverage ratio are negatively and statistically

significantly related to returns of investment grade credit variance swaps, driven by the payer variance swaps. Economically speaking, an increase in riskiness in the constituents of the CDX IG index leads to an increase in the index spread, and hence, to more index variation in the payer corridor. This in turn causes higher payer variance swap payoffs, diminishing returns to shorting payer swaps. However, the intercept is close to the previously documented average monthly returns. Moreover, the intercept is highly statistically significant. We find the same results when considering high yield variance risk premiums. Concluding, structural model variables are not sufficient to explain returns to credit variance swaps.

4.5. Risk-adjusted Returns

Next, we tackle the question of how well CVP can be attributed to established equity and corporate bond asset pricing risk factors. Precisely, we estimate the following regression model

$$r_t = \alpha + \beta' X_t + \epsilon_t, \tag{14}$$

where r_t is the return of credit variance swaps (in month t) and X_t the set of risk-factors. We use the six-factor model of Fama and French (2018), augmented by the intermediary risk factor of He et al. (2017) as equity risk-factors. For corporate bond risk-factors, we use the four factors proposed in Bai et al. (2019).[12]

Table 6 provides results for the equity risk-factors. First, we set X_t to include only the equity market excess return. For both credit classes, the slope coefficient of the market portfolio is negative, but only significantly different from zero for high yield CVP. Shedding light on this, we repeat the regression with corridor variance risk premiums. Payer (receiver) variance risk premiums exhibit strong, negative (positive) exposure to the market portfolio, rendering the overall effect on CVP insignificant. Hence, going long payer variance swaps provides insurance against negative returns of the equity market portfolio. Second, we include the entire set of risk factors into X_t. The loadings on the investment portfolio are negative. Besides the market portfolio, this is the only risk-factor which is statistically significantly related to CVP.

Table 7 reports results for corporate bond risk factors. The corporate bond market

[12]Data on the six-factor model of Fama and French (2018) and the risk-free rate r^f are taken from Kenneth French's website Data on the intermediary capital risk factor is taken from Zhiguo He's website. We thank Turan G. Bali for providing us with data on the corporate bond factors.

Table 6: **Equity Risk-adjusted Variance Risk Premiums**
This table reports the estimation results from regressing the excess returns on variance (CVP) and semi-variance swaps (CVP Payer, CVP Receiver) on the equity market excess return (MRKT), size (SMB), book-to-market (HML), profitability (RMW), investment (CMA), momentum (MOM) and an intermediary asset pricing (IR) factor. Column (10) reports the adjusted R^2 (R^2_{adj}) of the regression. t-statistics are given in parentheses and adjusted for heteroskedasticity and autocorrelation using Newey and West (1987). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 until September 2021 (October 2018 for the multi-factor specification).

Strategy	Alpha	MRKT	SMB	HML	RMW	CMA	MOM	IR	R^2_{adj}
Panel A: Investment Grade									
CVP	-0.69***	-0.02							0.03
	(-9.79)	(-1.07)							
	-0.60***	-0.02	0.01	0.05	-0.03	-0.14**	-0.01	-0.32	0.02
	(-8.86)	(-0.92)	(0.22)	(1.15)	(-0.66)	(-2.27)	(-0.22)	(-0.16)	
CVP Payer	-0.44***	-0.12***							0.44
	(-6.87)	(-9.15)							
	-0.35***	-0.14***	0.04	0.07*	-0.03	-0.14**	0.00	0.44	0.35
	(-5.54)	(-5.61)	(1.17)	(1.83)	(-0.64)	(-2.49)	(0.16)	(0.25)	
CVP Receiver	-0.62***	0.05**							0.06
	(-6.37)	(2.41)							
	-0.63***	0.13***	-0.05	-0.06	-0.07	0.02	0.01	0.72	0.16
	(-5.86)	(3.13)	(-0.93)	(-0.85)	(-0.88)	(0.23)	(0.20)	(0.23)	
Panel B: High Yield									
CVP	-0.78***	-0.04**							0.05
	(-10.17)	(-2.27)							
	-0.77***	-0.04	0.03	0.09	-0.03	-0.14*	0.03	-0.34	0.02
	(-9.28)	(-1.39)	(0.72)	(1.64)	(-0.42)	(-1.73)	(0.89)	(-0.15)	
CVP Payer	-0.73***	-0.12***							0.31
	(-8.83)	(-6.74)							
	-0.69***	-0.12***	0.01	0.05	-0.02	-0.11	0.04	0.88	0.20
	(-8.27)	(-3.85)	(0.16)	(0.96)	(-0.30)	(-1.35)	(1.27)	(0.38)	
CVP Receiver	-0.50***	0.12***							0.20
	(-4.66)	(5.34)							
	-0.46***	0.18***	0.00	0.03	-0.08	-0.00	-0.02	-3.21	0.17
	(-3.79)	(3.75)	(0.05)	(0.40)	(-0.80)	(-0.03)	(-0.47)	(-0.93)	

portfolio is negatively associated with payer variance risk premiums, whereas it positively co-moves with receiver variance risk premiums. In times of market turmoil and falling bond prices due to rising default concerns, payer variance swaps generate positive returns for investors. Whereas in times of vanishing default risks and positive bond returns, credit spreads move lower and cause more variation in the receiver swap, increasing its return. Both effects mean that the explanatory power of corporate bond yields is insignificant for

Table 7: **Corporate Bond Risk-adjusted Variance Risk Premiums**
This table reports the estimation results from regressing the excess returns on variance (CVP) and semi-variance swaps (CVP Payer, CVP Receiver) on the corporate bond market excess return (MRKT), dwonside risk (DRF), credit risk (CRF), and liquidity risk (LRF) factor. Column (7) reports the adjusted R^2 (R^2_{adj}) of the regression. t-statistics are given in parentheses and adjusted for heteroskedasticity and autocorrelation using Newey and West (1987). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 until December 2019.

Strategy	Alpha	MRKT	DRF	CRF	LRF	R^2_{adj}
Panel A: Investment Grade						
CVP	-0.64***	-0.03				-0.01
	(-9.39)	(-0.29)				
	-0.66***	-0.14*	0.03	-0.03	0.19	-0.01
	(-8.67)	(-1.67)	(0.53)	(-0.34)	(1.44)	
CVP Payer	-0.46***	-0.11				0.02
	(-6.89)	(-1.11)				
	-0.47***	-0.22**	0.01	-0.09	0.20	0.04
	(-6.20)	(-2.39)	(0.24)	(-0.99)	(1.55)	
CVP Receiver	-0.51***	0.22*				0.04
	(-5.27)	(1.70)				
	-0.54***	0.22*	0.05	0.16*	-0.05	0.05
	(-5.57)	(1.82)	(0.46)	(1.67)	(-0.24)	
Panel B: High Yield						
CVP	-0.78***	-0.14				0.02
	(-9.33)	(-1.47)				
	-0.78***	-0.21	0.06	-0.02	-0.00	-0.00
	(-8.71)	(-1.62)	(0.78)	(-0.30)	(-0.02)	
CVP Payer	-0.70***	-0.23**				0.07
	(-7.83)	(-2.10)				
	-0.70***	-0.34**	0.05	-0.08	0.09	0.06
	(-7.39)	(-2.19)	(0.62)	(-0.95)	(0.65)	
CVP Receiver	-0.35***	0.21				0.03
	(-3.32)	(1.54)				
	-0.35***	0.31**	0.02	0.14	-0.25	0.04
	(-3.09)	(2.11)	(0.24)	(1.40)	(-1.54)	

CVP as a whole. In all specifications, the estimated alphas are statistically significantly different from zero and close to the sample averages of CVP. Hence, corporate bond risk factors cannot explain returns to shorting credit volatility.

Table 8: **Comovement with Equity Market Variance Risk Premiums**
The table presents the results from regressing excess returns on credit variance (CVP) and semi-variance swap on contemporaneous excess returns on S&P 500, VIX and TLT variance swaps. Column (6) reports the adjusted R^2 (R^2_{adj}) of the regression. We assume an initial amount of capital required to entire into a swap contract. We adjust this amount to achieve a 10% annualized excess return volatility. Returns are displayed in percent. t-statistics are given in parentheses and adjusted for heteroskedasticity and autocorrelation using Newey and West (1987). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 until December 2020.

Strategy	Alpha	S&P 500	VIX	TLT	R^2_{adj}
Panel A: Investment Grade					
CVP	-0.40***	0.95***			0.90
	(-5.16)	(56.92)			
	-0.37**		0.51		0.25
	(-2.10)		(1.42)		
	-0.32***			0.84***	0.70
	(-4.74)			(31.70)	
CVP Payer	-0.24***	0.98***			0.96
	(-3.46)	(46.72)			
	-0.24*		0.56		0.31
	(-1.80)		(1.36)		
	-0.57***			0.95***	0.90
	(-6.96)			(33.55)	
CVP Receiver	-0.49***	0.67***			0.45
	(-2.73)	(5.20)			
	-0.34*		0.02		-0.01
	(-1.69)		(0.15)		
	-0.89***			-0.62***	0.38
	(-3.89)			(-3.71)	
Panel B: High Yield					
CVP	-0.52***	0.89***			0.79
	(-6.22)	(99.60)			
	-0.50**		0.45		0.19
	(-2.56)		(1.25)		
	-0.46***			0.73***	0.53
	(-2.90)			(21.58)	
CVP Payer	-0.47***	0.94***			0.89
	(-8.77)	(121.99)			
	-0.47		0.53		0.27
	(-1.47)		(1.29)		
	-0.79***			0.95***	0.91
	(-8.56)			(34.56)	
CVP Receiver	-0.30*	0.65***			0.42
	(-1.80)	(5.70)			
	-0.15		-0.01		-0.01
	(-0.88)		(-0.09)		
	-0.63***			-0.55***	0.30
	(-2.91)			(-4.01)	

4.6. Comovement with Equity Variance and Variance-of-variance Risk Premiums

Next, we study the comovement of CVP with equity variance, variance-of-variance and Treasury bond variance risk premiums. We obtain options on the S&P 500, the VIX and the exchange traded fund TLT from OptionMetrics and implement the same methodology to estimate excess returns on synthetic variance swaps.[13] Thereby, we ensure that holding periods of variance swaps are perfectly aligned to credit variance swaps.

We estimate the following regression

$$CVP_{t,T} = \alpha + \beta VRP^{x}_{t,T} + \epsilon_t. \tag{15}$$

$VRP^{x}_{t,T}$ denotes the variance risk premium in asset class x, $x \in \{\text{S\&P 500}, \text{VIX}, \text{TLT}\}$.[14]

Table 8 provides results. CVP loads significantly (at the one-percent level) and positively on S&P 500 and TLT variance swaps. Hence, excess returns to credit variance swaps are positively related to their counterparts in equity and fixed-income markets. This is probably not surprising given that in times of economic distress, volatilities in all asset classes rise. However, the alpha remains relatively large and significant. It ranges between -0.32% and -0.40% (-0.46% and -0.52%) for investment grade (high yield) CVP. In terms of adjusted R^2s, equity variance risk premiums yield the best fit by achieving 90% (79%).

We repeat the analysis for payer and receiver variance risk premiums. To compare like for like, we construct corridor variance risk premiums for the S&P 500, VIX and TLT. Payer (receiver) variance risk premiums are associated with bad (good) economic states of the world. The same holds true for downward (upward) equity risk premiums and upward (downard) variance-of-variance and fixed-income variance risk premiums. In estimating Equation (15), we regress credit corridor variance risk premiums on the respective counterparts for S&P 500, VIX and TLT. As shown in Table 8, payer and receiver variance risk premiums significantly comove with equity and fixed-income variance risk premiums. The link seems to be stronger for equity variance risk premiums. This echos findings in Norden and Weber (2009) who show that the CDS market is more sensitive to the equity than the bond market. Moreover, payer variance swaps appear to be better explained

[13] TLT tracks the performance of the Barclays Capital U.S. 20+ Year Treasury Bond Index, and hence, is exposed to long-term U.S. Treasury bonds.

[14] Results in univariate regressions are qualitatively and quantitatively similar to the multivariate setting which we omit for brevity.

by the corresponding equity and fixed-income corridor swaps as gauged by the adjusted R^2. A potential explanation for this finding is that payer variance swaps and equity left tail risk premiums are driven by the same economic principle, investor's high demand for macro-economic downside protection. However, as for total CVP, alphas remain significantly different from zero and economically large. Finally, credit variance risk premiums are not significantly related to corridor variance-of-variance risk premiums, irrespectively of the chosen credit rating and corridor.

4.7. Predictive Power of Credit Implied Volatility

In the realm of equity markets, the VIX is known as a fear gauge, due to its tendency to rise in times of negative equity market returns. Moreover, the VIX proves to exhibit predictive power for future economic activity and financial stability (see Bekaert and Hoerova, 2014). In the following, inspired by Choi et al. (2017), we analyze whether credit implied variance and the slope of the credit implied variance term structure contain predictive power for financial stress, measured by the St. Louis Fed Stress Index (STLFSI, hereafter).[15]

[15]The STLFSI uses 18 weekly data series (including interest rates, yield spreads, and other indicators such as the VIX and the MOVE) and principal component analysis to measure stress in the financial markets. A reading above 0 indicates above-average financial market stress, whereas values below zero indicate below-average financial market stress. In unreported results, we have verified that we obtain similar results when using systemic risk indices such as the SRISK index of Brownlees and Engle (2017), obtained via `https://vlab.stern.nyu.edu/srisk`.

Table 9: **Predictive Regressions St. Louis Fed Stress Index**
This table reports the estimation results from weekly regressions of the St. Louis Fed Stress Index (STLFSI) onto CIV squared and the slope of CIV, $\text{slope}^{\text{CIV}}$, i.e.

$$\text{STLFSI}_{t+n} = \alpha_n + \beta_n^{\text{CIV}}\text{CIV}^2 + \beta_n^{\text{slope CIV}}\text{slope}^{\text{CIV}} + \epsilon_{t+n}, n \in \{0, 4, 8, 12\}.$$

CIV denotes credit implied volatility with 45 days to maturity. $\text{slope}^{\text{CIV}}$ is measured by the difference between implied variance with 105 and 45 days to maturity. All variables are standardized, that means de-meaned and divided by their respective standard deviation. We employ two approaches for statistical inference. The first four columns denote the intercept and slope estimates, as well as, the adjusted R^2 (R^2_{adj}) using the standard least squares approach. t-statistics are given in parentheses and adjusted for heteroskedasticity and autocorrelation using Newey and West (1987). $\text{CIV}^2_{\text{IVX}}$ and $\text{slope}^{\text{CIV}}_{\text{IVX}}$ stand for the slope coefficients estimated by the IVX method recently propsed by Kostakis et al. (2015). Wald statistics under the null hypothesis that the individual slope coefficient is equal to zero are given in parentheses. *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 until September 2021.

α	CIV^2	$\text{slope}^{\text{CIV}}$	R^2_{adj}	$\text{CIV}^2_{\text{IVX}}$	$\text{slope}^{\text{CIV}}_{\text{IVX}}$
Panel A: Contemp.					
0.00	0.75***		0.56	0.77***	
(0.00)	(5.11)			(614.68)	
-0.00		-0.74***	0.55		-0.75***
(-0.00)		(-9.75)			(563.43)
0.00	0.43***	-0.40***	0.62	0.43***	-0.42***
(0.00)	(4.10)	(-5.28)		(82.03)	(76.46)
Panel B: 4 weeks					
-0.00	0.37***		0.14	0.40***	
(-0.04)	(4.66)			(81.46)	
-0.00		-0.40***	0.16		-0.41***
(-0.04)		(-5.38)			(86.69)
-0.00	0.15	-0.28**	0.17	0.17**	-0.28***
(-0.04)	(1.58)	(-2.24)		(5.82)	(14.51)
Panel C: 8 weeks					
-0.01	0.15***		0.02	0.18***	
(-0.06)	(3.40)			(14.12)	
-0.01		-0.15***	0.02		-0.16***
(-0.06)		(-3.00)			(11.59)
-0.01	0.08	-0.09	0.02	0.12	-0.07
(-0.06)	(0.90)	(-0.86)		(2.44)	(0.76)

We focus on investment grade implied credit volatility over the next 45 days and denote it CIV for brevity. Furthermore, we measure the slope of the term structure of implied variances, $\text{slope}^{\text{CIV}}$, by the difference between 105-day and 45-day implied variances. To test the predictive power of CIV and $\text{slope}^{\text{CIV}}$, we examine the following

regression model on a weekly basis for different horizons, n, ranging from 0 to 8 weeks:

$$\text{STLFSI}_{t+n} = \alpha_n + \beta_n^{\text{CIV}}\text{CIV}^2 + \beta_n^{\text{slope CIV}}\text{slope}^{\text{CIV}} + \epsilon_{t+n}, \quad n \in \{0, 4, 8\}.$$

We report results in Table 9 and concentrate first on univariate regressions. We conjecture a positive (negative) relationship between STLFSI and credit implied variances (its term structure). In times of stress, credit implied variance rises and the term structure of credit implied variances inverts. Our results confirm this hypothesis. CIV^2 is positively and significantly related to STLFSI, regardless of the horizon considered. Any one standard deviation change in CIV^2 yields a 0.15 to 0.75 standard deviation increase in STLFSI. A negative and significant effect is also estimated for the slope of the credit implied variance term structure. Moreover, CIV^2 and its slope pose strong predictive power, as suggested by adjusted R^2s. At contemporaneous and short horizons, the adjusted R^2s is above 14%. However, the power diminishes with the horizon considered, as the adjusted R^2 drops sharply to 2% for the longest horizon considered. Our findings generalize to the multivariate setting.

It is well known that volatility and variance show a high degree of persistence. To address potential econometric concerns, we use the recently proposed instrumental variable (IVX) method of Kostakis et al. (2015). It allows for statistical inference in the presence of persistent regressors (unit root, local-to-unit root, near stationary, stationary). Precisely, we compute their IVX-slope coefficient and Wald statistic under the null hypothesis that the slope coefficients equal zero.[16] Results, tabulated in the two right most columns in Table 9, confirm previous inferences and the predictive power of CIV^2.

5. Conclusion

Using a novel data set of swaptions written on the CDX North America Investment Grade and High Yield indices, this paper studies synthetic credit variance risk premiums using variance swaps. We find that investors command a monthly premium of 0.58% (0.73%) for bearing investment grade (high yield) variance risks, well above variance risk premiums in other asset classes. Shorting credit variance swaps is also superior to other asset classes in risk-adjusted performance measures. Our results are quantitatively and qualitatively the same for investment grade and high yield credit and suggest that credit variance risk premiums serve as catastrophic risk premiums. We dissect credit variance

[16]We thank Alexandros Kostakis for providing code for the implementation of the IVX method.

risk premiums into payer and receiver variance risk premiums by means of corridor variance swaps. The former are associated with improving macro-economic conditions, the latter with a deterioration. We find that both corridor variance risk premiums are strongly negative and significant. However, payer variance risk premiums are larger in absolute magnitude than receiver variance risk premium. Neither well-known asset pricing, nor structural credit factors are able to explain credit variance risk premiums. Our results remain robust if we vary the investment horizon of credit variance swaps and account for transaction costs. Finally, we show that the level of credit implied volatility is a highly useful predictor of financial stress.

Appendix

Appendix A Synthesizing Variance Swap Contracts

First, we assume that $\kappa^{\mathrm{a}}_{\mathrm{t,T}}$ follows a continuous process. This allow us to write

$$\frac{\mathrm{d}\,\kappa^{\mathrm{a}}_{\mathrm{t,T}}}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} = \sigma_t dW^{\mathbb{A}}_t,$$

where $dW^{\mathbb{A}}_t$ is a standard Brownian motion under the risky annuity measure $\mathbb{A}$, and σ_t denotes the instantaneous volatility.

We follow Bakshi and Madan (2000), Carr and Madan (2001), and Bakshi, Kapadia, and Madan (2003) and Bondarenko (2014). For any twice-differentiable function g and any fixed positive x, we have

$$\begin{aligned} g(\kappa^a_{T,T}) =& g(x) + g'(x)\left(\kappa^a_{T,T} - x\right) + \int_0^x g''(\kappa)\left(K - \kappa^a_{T,T}\right)^+ d\kappa \\ &+ \int_x^\infty g''(\kappa)\left(\kappa^a_{T,T} - \kappa\right)^+ d\kappa. \end{aligned}$$

Specifically for $x = \kappa^a_{t,T}$, we obtain

$$\begin{aligned} g(\kappa^a_{T,T}) =& g(\kappa^a_{t,T}) + g'(\kappa^a_{t,T})\left(\kappa^a_{T,T} - \kappa^a_{t,T}\right) + \int_0^{\kappa^a_{t,T}} g''(\kappa)\left(K - \kappa^a_{T,T}\right)^+ d\kappa \\ &+ \int_{\kappa^a_{t,T}}^\infty g''(\kappa)\left(\kappa^a_{T,T} - \kappa\right)^+ d\kappa. \end{aligned}$$

Taking as $g(\kappa) = \log\kappa$ yields

$$\begin{aligned} \log\kappa^a_{T,T} =& \log\kappa^a_{t,T} + \frac{1}{\kappa^a_{t,T}}\left(\kappa^a_{T,T} - \kappa^a_{t,T}\right) - \int_0^{\kappa^a_{t,T}} \frac{1}{\kappa^2}\left(K - \kappa^a_{T,T}\right)^+ d\kappa \\ &- \int_{\kappa^a_{t,T}}^\infty \frac{1}{\kappa^2}\left(\kappa^a_{T,T} - \kappa\right)^+ d\kappa. \end{aligned} \tag{A.1}$$

Ito's lemma implies that

$$\mathrm{d}(\log\kappa^{\mathrm{a}}_{\mathrm{t,T}}) = \frac{\mathrm{d}\,\kappa^{\mathrm{a}}_{\mathrm{t,T}}}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} - \frac{1}{2}\sigma^2_t dt.$$

Therefore, the quadratic variation of $\kappa^{\mathrm{a}}_{\mathrm{t,T}}$ can be expressed as

$$\begin{aligned}\int_t^T \sigma_s^2 ds =& -2\log\frac{\kappa^2_{T,T}}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} + \int_t^T \frac{\mathrm{d}\kappa^a_{s,T}}{\kappa^a_{s,T}} \\ =& 2\left(\frac{\kappa^2_{T,T}-\kappa^{\mathrm{a}}_{\mathrm{t,T}}}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} - \log\frac{\kappa^2_{T,T}}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}}\right) + 2\int_t^T \left(\frac{1}{\kappa^a_{s,T}} - \frac{1}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}}\right)\mathrm{d}\kappa^a_{s,T} \\ =& 2\left(\int_0^{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} \frac{\left(\kappa-\kappa^a_{T,T}\right)^+}{\kappa^2}\mathrm{d}\kappa + \int_{\kappa^{\mathrm{a}}_{\mathrm{t,T}}}^{\infty} \frac{\left(\kappa^a_{T,T}-\kappa\right)^+}{\kappa^2}\mathrm{d}\kappa\right) \\ &+ 2\int_t^T \left(\frac{1}{\kappa^a_{s,T}} - \frac{1}{\kappa^{\mathrm{a}}_{\mathrm{t,T}}}\right)\mathrm{d}\kappa^a_{s,T}. \end{aligned} \tag{A.2}$$

Next, we take expectations under the risky annuity measure $\mathbb{A}$ of Equation (A.2), multiply by A^a_t and use that $\kappa^a_{t,T}$ is a martingale under $\mathbb{A}$. This yields

$$\begin{aligned} A^a_t\mathbb{E}^{\mathbb{A}}_t\left[\int_t^T \sigma_s^2 ds\right] =& -2A^a_t\mathbb{E}^{\mathbb{A}}_t\left[\log\kappa^a_{T,T} - \log\kappa^{\mathrm{a}}_{\mathrm{t,T}}\right] \\ =& 2A^a_t\mathbb{E}^{\mathbb{A}}\left[\int_0^{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} \frac{\left(\kappa-\kappa^a_{T,T}\right)^+}{\kappa^2}\mathrm{d}\kappa + \int_{\kappa^{\mathrm{a}}_{\mathrm{t,T}}}^{\infty} \frac{\left(\kappa^a_{T,T}-\kappa\right)^+}{\kappa^2}\mathrm{d}\kappa\right] \\ =& 2\left(\int_0^{\kappa^{\mathrm{a}}_{\mathrm{t,T}}} \frac{\mathcal{R}_T(t,\kappa)}{\kappa^2}\mathrm{d}\kappa + \int_{\kappa^{\mathrm{a}}_{\mathrm{t,T}}}^{\infty} \frac{\mathcal{P}_T(t,\kappa)}{\kappa^2}\mathrm{d}\kappa\right). \end{aligned}$$

Finally, we relax the assumption on continuity of $\kappa^{\mathrm{a}}_{\mathrm{t,T}}$ and follow Bondarenko (2014). Equation (7) can be rewritten as

$$\begin{aligned} RV_{t,T} =& -2\log\frac{\kappa^a_{t_n,T}}{\kappa^a_{t_0,T}} + 2\sum_{i=1}^n \left(\frac{\kappa^a_{t_i,T}-\kappa^a_{t_{i-1},T}}{\kappa^a_{t_{i-1},T}}\right) \\ =& 2\left(\frac{\kappa^a_{t_n,T}-\kappa^a_{t_0,T}}{\kappa^a_{t_0,T}} - \log\frac{\kappa^a_{t_n,T}}{\kappa^a_{t_0,T}}\right) \\ &+2\sum_{i=1}^n \left(\frac{1}{\kappa^a_{t_{i-1},T}} - \frac{1}{\kappa^a_{t_0,T}}\right)\left(\kappa^a_{t_i,T}-\kappa^a_{t_{i-1},T}\right) \\ =& 2\left(\int_0^{\kappa^a_{t,T}} \frac{1}{\kappa^2}\left(K-\kappa^a_{T,T}\right)^+ d\kappa + \int_{\kappa^a_{t,T}}^{\infty} \frac{1}{\kappa^2}\left(\kappa^a_{T,T}-\kappa\right)^+ d\kappa\right) \\ &+ 2\sum_{i=1}^n \left(\frac{1}{\kappa^a_{t_{i-1},T}} - \frac{1}{\kappa^a_{t_0,T}}\right)\left(\kappa^a_{t_i,T}-\kappa^a_{t_{i-1},T}\right), \end{aligned}$$

where the last equation follows from Equation (A.1).

Appendix B Spread vs. Price Realized Volatility

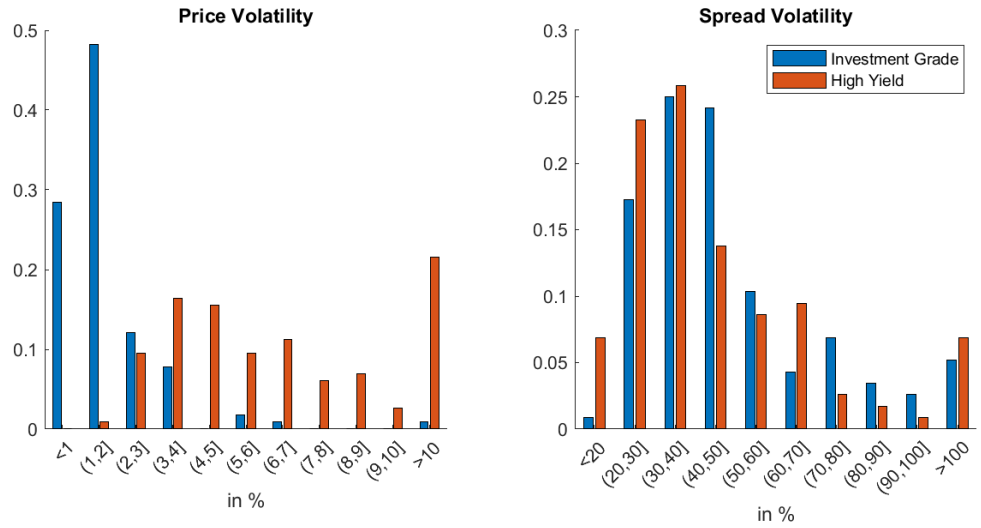


Fig. B.1. Histograms of Spread and Price Volatility
The figure shows the histograms of monthly spread and price volatility for investment grade and high yield credit over the sample period from March 2012 to September 2021.

Appendix C Statistical Significance

Option trading strategies show non-Gaussian behaviour, evident in Table 3, threatening statements on statistical significance. Mitigating any concerns, we employ a studentized bootstrap to obtain 95% confidence intervals on the mean and four versions of alpha for the trading strategies previously presented. The first three versions of alpha refer to Equation (14). The fourth alpha refers to the intercept in Equation (13). We use a sample of 10,000 bootstrapped repetitions to report confidence intervals in Table C.1.

Table C.1 confirms our previous findings for average returns. Average returns for overall and payer credit variance swaps strategies are significantly different from zero, since none of the intervals contain the zero itself. The same applies to the alphas for equity and corporate bond market risk factors and the intercept in the structural model regression.

Table C.1: **Bootstrapped Confidence Intervals**
This table reports 95% bootstrapped confidence intervals for means and alphas for the excess returns on the three trading strategies presented in Panel A and B of Table 3. We consider four versions of alpha: α_{MRKT}, α_{FF6+1} and α_{BOND} refer to the three alphas considered in Table 6, α_{STRUCT} refers to the intercept in Table 5. LB (UB) denotes the lower (upper) bound of the confidence interval. The empirical distribution of returns is obtained from 10,000 studentized bootstrap repititions of our sample.

	Mean		α_{MRKT}		α_{FF6+1}		α_{BOND}		α_{STRUCT}	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
Panel A: Investment Grade										
CVP	-1.58	-0.22	-0.87	-0.53	-0.70	-0.46	-0.83	-0.52	-1.62	-0.82
CVP Payer	-1.54	-0.07	-0.61	-0.25	-0.48	-0.26	-0.63	-0.32	-0.88	-0.55
CVP Receiver	-1.16	0.01	-0.82	-0.46	-0.73	-0.40	-0.71	-0.37	-3.19	0.49
Panel B: High Yield										
CVP	-1.66	-0.35	-1.00	-0.66	-0.90	-0.58	-0.99	-0.61	-1.67	-0.52
CVP Payer	-1.88	-0.33	-0.91	-0.52	-0.78	-0.46	-0.90	-0.52	-1.11	-0.34
CVP Receiver	-0.94	0.24	-0.63	-0.22	-0.67	-0.21	-0.56	-0.13	-2.48	0.20

Appendix D Term Structure of Credit Variance Risk Premiums

We focus on investment holding periods of 2 to 4 months. Each trading day after the swaption expiry, we calculate returns of variance swaps with 2, 3, and 4 months to expiry. We do not include options with more than 4 months time to maturity, as Table 1 has shown that these are less liquid. As for excess returns on 1-month variance swaps, we adjust the excess returns to achieve a 10% annualized volatility. Table D.1 documents average returns, corresponding t-statistics, and Sharpe and Stutzer ratios. The upward sloping implied volatility term structure indicates a downward sloping term structure of variance risk premiums. Table D.1 confirms the intuition. The return in a long 2-month investment grade (high yield) variance swap amounts to -0.97% (-1.19%), whereas a long 4-month variance swap yields -1.51% (-1.76%). Average returns to shorting credit variance swaps remain statistically significant from zero, regardless of the swap tenor and credit quality. The same finding is observed for investment grade and high yield payer swaps. However, the risk-adjusted performance by means of the Sharpe ratio is in absolute terms monotonically decreasing in time to maturity. Sharpe ratios declining in time to maturity have been documented for a range of markets by van Binsbergen and Koijen (2017) and Choi et al. (2017) for bond variance risk premiums.

Table D.1: **Credit Variance Risk Premium (Monthly Term Structure)**
This table reports summary statistics of excess returns on variance and semi-variance swaps with varying holding periods. On each month, the first trade date after swaption expiry is selected and excess returns on variance and semi-variance swaps are computed from swaptions with two, three and four months to expiry. Swap contracts are held until maturity. We assume an initial amount of capital required to entire into a swap contract. We adjust this amount to achieve a 10% annualized excess return volatility. Returns are displayed in percent. Standard errors are computed using Newey and West (1987) accounting for autocorrelation and hetereoscedasticity with optimal bandwidth. Sharpe ratios are corrected for return-autocorrelation using the approach of Lo (2002). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 to September 2021.

	2 months			3 months			4 months		
Strategy	Mean	Sharpe Ratio	Stutzer Ratio	Mean	Sharpe Ratio	Stutzer Ratio	Mean	Sharpe Ratio	Stutzer Ratio
Panel A: Investment Grade									
CVP	-0.97**	-0.49	-0.20	-1.32**	-0.36	-0.22	-1.51*	-0.30	-0.23
	(-2.11)			(-1.99)			(-1.81)		
CVP Payer	-0.68	-0.38	-0.15	-0.96	-0.28	-0.17	-1.09	-0.23	-0.17
	(-1.56)			(-1.52)			(-1.34)		
CVP Receiver	-1.37***	-0.78	-0.30	-1.56***	-0.56	-0.28	-1.97***	-0.52	-0.32
	(-3.59)			(-3.07)			(-3.07)		
Panel B: High Yield									
CVP	-1.19**	-0.53	-0.24	-1.55**	-0.40	-0.26	-1.76**	-0.36	-0.27
	(-2.38)			(-2.24)			(-2.10)		
CVP Payer	-1.10**	-0.52	-0.22	-1.46**	-0.40	-0.24	-1.64*	-0.33	-0.24
	(-2.30)			(-2.18)			(-1.94)		
CVP Receiver	-0.47	-0.33	-0.11	-0.52	-0.22	-0.10	-0.76	-0.26	-0.14
	(-1.34)			(-1.13)			(-1.54)		

Appendix E Transaction Costs

In the following we investigate the impact of transaction costs onto the profitability of shorting credit variance swaps. Consequently, we reconstruct synthetic variance swap rates using effective bid-ask spreads of 25% instead of mid option prices.

Markit provides us with bid and ask option prices for the second half of our sample. To obtain bid prices for previous series, we proceed as follows. We treat the bid-ask spread relative to the mid price as a function of time to maturity and moneyness level. Subsequently, we use k-nearest-neighbours to find the bid option price for swaptions with missing bid-ask spread. Our procedure implicitly assumes that liquidity patterns in the swaptions remain constant over our entire sample period. Trolle and Schwartz (2014) rest on the same assumption while contacting major investment banks to obtain indicative bid-ask spreads at the sample end. Subsequently, the authors form three different strike buckets and apply to each a representative bid-ask spread. Since Trolle and Schwartz (2014) concentrate on a single maturity, we would have to introduce a second dimension to such an approach. We believe that, instead of constructing such a two-dimensional grid, it is more intuitive to use k-nearest-neighbours.

Table E.1 reports the average returns for investment horizons spanning 1 to 4 months. Shorting credit variance swaps remain a successful strategy even in light of transaction costs.

Table E.1: **Transaction Cost Adjusted Credit Variance Risk Premiums**
This table documents summary statistics of excess returns on variance (CVP) and semi-variance swaps (CVP Payer, CVP Receiver). On each month, the first trade date after swaption expiry is selected and returns on variance and semi-variance swaps are computed from swaptions with one, two, three and four months to expiry. Swap contracts are held until maturity. Instead of relying on mid prices, we use effective bid-ask spreads of 25% to account for transaction costs in computing the variance swap rates. We assume an initial amount of capital required to entire into a swap contract. We adjust this amount to achieve a 10% annualized excess return volatility. Returns are displayed in percent. t-statistics are adjusted for heteroskedasticity and autocorrelation using Newey and West (1987). All Sharpe ratios account for return-autocorrelation according to Lo (2002). *, **, and *** denote significance at the 10%, 5% and 1% levels, respectively. The sample period runs from April 2012 until September 2021.

	1 month		2 months		3 months		4 months	
Strategy	Mean	Sharpe Ratio	Mean	Sharpe Ratio	Mean	Sharpe Ratio	Mean	Sharpe Ratio
Panel A: Investment Grade								
CVP	-0.48*	-0.66	-0.81*	-0.40	-1.10*	-0.30	-1.24	-0.25
	(-1.80)		(-1.75)		(-1.65)		(-1.48)	
CVP Payer	-0.37	-0.65	-0.59	-0.33	-0.83	-0.24	-0.94	-0.19
	(-1.59)		(-1.34)		(-1.31)		(-1.15)	
CVP Receiver	-0.29	-0.52	-1.02***	-0.59	-1.13**	-0.41	-1.41**	-0.37
	(-1.34)		(-2.67)		(-2.24)		(-2.22)	
Panel B: High Yield								
CVP	-0.61**	-0.71	-1.01**	-0.45	-1.32*	-0.34	-1.49*	-0.30
	(-2.12)		(-2.02)		(-1.90)		(-1.77)	
CVP Payer	-0.61**	-0.80	-0.98**	-0.47	-1.31*	-0.35	-1.46*	-0.29
	(-2.26)		(-2.05)		(-1.95)		(-1.73)	
CVP Receiver	-0.10	-0.19	-0.19	-0.13	-0.13	-0.06	-0.41	-0.14
	(-0.52)		(-0.53)		(-0.29)		(-0.83)	

Bibliography

Ammann, M., Buesser, R., 2013. Variance risk premiums in foreign exchange markets. Journal of Empirical Finance 23, 16–32.

Andersen, T. G., Bondarenko, O., 2010. Dissecting the market pricing of return volatility. Northwestern University and University of Illinois at Chicago.

Aït-Sahalia, Y., Karaman, M., Mancini, L., 2020. The term structure of equity and variance risk premia. Journal of Econometrics 219, 204 – 230.

Bai, J., Bali, T. G., Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. Journal of Financial Economics 131, 619–642.

Bai, J., Goldstein, R. S., Yang, F., 2020. Is the credit spread puzzle a myth? Journal of Financial Economics 137, 297–319.

Bakshi, G., Kapadia, N., Madan, D., 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. Review of Financial Studies 16, 101–143.

Bakshi, G., Madan, D., 2000. Spanning and derivative-security valuation. Journal of Financial Economics 55, 205–238.

Barone-Adesi, G., Whaley, R. E., 1987. Efficient Analytic Approximation of American Option Values. Journal of Finance 42, 301–320.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. Journal of Econometrics 183, 181–192.

Black, F., 1976. Studies of stock price volatility changes. In: *Proceedings of the 1976 Meeting of the Business and Economic Statistics Section*, American Statistical Association, Washington, D.C.

Bondarenko, O., 2014. Variance trading and market price of variance risk. Journal of Econometrics 180, 81–97.

Brownlees, C., Engle, R. F., 2017. Srisk: A conditional capital shortfall measure of systemic risk. Review of Financial Studies 30, 48–79.

Campbell, J. Y., Taksler, G. B., 2003. Equity Volatility and Corporate Bond Yields. Journal of Finance 58, 2321–2350.

Cao, C., Yu, F., Zhong, Z., 2010. The information content of option-implied volatility for credit default swap valuation. Journal of Financial Markets 13, 321–343.

Carr, P., Madan, D., 2001. Optimal positioning in derivative securities. Quantitative Finance 1, 19–37.

Carr, P., Wu, L., 2009. Variance Risk Premiums. Review of Financial Studies 22, 1311–1341.

Chen, S. S.-H., Doshi, H., Seo, S. B., 2020. Ex ante risk in the corporate bond market: Evidence from synthetic options. University of Houston.

Choi, H., Mueller, P., Vedolin, A., 2017. Bond variance risk premiums. Review of Finance 21, 987–1022.

Collin-Dufresne, P., Goldstein, R. S., Martin, J. S., 2001. The determinants of credit spread changes. Journal of Finance 56, 2177–2207.

Collin-Dufresne, P., Junge, B., Trolle, A. B., 2020a. How integrated are credit and equity markets? evidence from index options. Swiss Finance Institute Research Paper Series No 20-65.

Collin-Dufresne, P., Junge, B., Trolle, A. B., 2020b. Market structure and transaction costs of index cdss. Journal of Finance 75, 2719–2763.

Coval, J. D., Shumway, T., 2001. Expected option returns. Journal of Finance 56, 983–1009.

Cremers, M., Driessen, J., Maenhout, P., Weinbaum, D., 2008. Individual stock-option prices and credit spreads. Journal of Banking & Finance 32, 2706–2715.

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

Duarte, J., Longstaff, F. A., Yu, F., 2007. Risk and Return in Fixed-Income Arbitrage: Nickels in Front of a Steamroller? Review of Financial Studies 20, 769–811.

Duembgen, M., Doctor, S., Mehonic, A., Lamy, D., Bailey, M., 2015. Introducing VTRAC-X and VTRAC-X Swaps Tracking and Trading Credit Volatility European Credit Strategy and Derivatives Research. JP Morgan.

Ericsson, J., Jacobs, K., Oviedo, R., 2009. The Determinants of Credit Default Swap Premia. Journal of Financial and Quantitative Analysis 44, 109.

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

Gabaix, X., 2012. Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance. Quarterly Journal of Economics 127, 645–700.

Gourio, F., 2013. Credit risk and disaster risk. American Economic Journal: Macroeconomics 5, 1–34.

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

Huang, J.-Z., Huang, M., 2012. How Much of the Corporate-Treasury Yield Spread Is Due to Credit Risk? Review of Asset Pricing Studies 2, 153–202.

Huang, J.-Z., Nozawa, Y., Shi, Z., 2019. The global credit spread puzzle. PBCSF-NIFR Research Paper.

Jarrow, R. A., Li, H., Ye, X., Hu, M., 2019. Exploring Mispricing in the Term Structure of CDS Spreads. Review of Finance 23, 161–198.

Kaeck, A., 2018. Variance-of-variance risk premium. Review of Finance 22, 1549–1579.

Kelly, B. T., Manzo, G., Palhares, D., 2019. Credit Implied Volatility. Working Paper.

Kita, A., Tortorice, D. L., 2021. Same firm, two volatilities: How variance risk is priced in credit and equity markets. Journal of Corporate Finance 69, 101885.

Kostakis, A., Magdalinos, T., Stamatogiannis, M. P., 2015. Robust econometric Inference for stock Return predictability. Review of Financial Studies 28, 1506–1553.

Lo, A. W., 2002. The Statistics of Sharpe Ratios. Financial Analysts Journal 58, 36–52.

Longstaff, F. A., Pan, J., Pedersen, L. H., Singleton, K. J., 2011. How Sovereign Is Sovereign Credit Risk? American Economic Journal: Macroeconomics 3, 75–103.

Longstaff, F. A., Rajan, A., 2008. An empirical analysis of the pricing of collateralized debt obligations. Journal of Finance 63, 529–563.

Mele, A., Obayashi, Y., 2015. The Price of Fixed Income Market Volatility. Springer, Cham, Switzerland.

Merton, R. C., 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. Journal of Finance 29, 449–470.

Morini, M., Brigo, D., 2011. No-armageddon measure for arbitrage-free pricing of index options in a credit crisis. Mathematical Finance 21, 573–593.

Neuberger, A., 2012. Realized Skewness. Review of Financial Studies 25, 3423–3455.

Newey, W. K., West, K. D., 1987. A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

Norden, L., Weber, M., 2009. The Co-movement of Credit Default Swap, Bond and Stock Markets: an Empirical Analysis. European Financial Management 15, 529–562.

O'Kane, D., 2008. Modelling single-name and multi-name credit derivatives. John Wiley & Sons.

Pires, P., Pereira, J. P., Martins, L. F., 2015. The empirical determinants of credit default swap spreads: a quantile regression approach. European Financial Management 21, 556–589.

Prokopczuk, M., Symeonidis, L., Wese Simen, C., 2017. Variance risk in commodity markets. Journal of Banking & Finance 81, 136–149.

Rutkowski, M., 2012. Options on Credit Default Swaps and Credit Default Indexes. In: *Credit Risk Frontiers*, John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 219–279.

Seo, S. B., Wachter, J. A., 2018. Do rare events explain cdx tranche spreads? Journal of Finance 73, 2343–2383.

Toublan, D., 2017. Credit Derivatives Workshop High Grade Strategy and Credit Derivatives Research. JP Morgan.

Trolle, A. B., Schwartz, E. S., 2010. Variance Risk Premia in Energy Commodities. Journal of Derivatives 17, 15–32.

Trolle, A. B., Schwartz, E. S., 2014. The swaption cube. Review of Financial Studies 27, 2307–2353.

van Binsbergen, J. H., Koijen, R. S., 2017. The term structure of returns: Facts and theory. Journal of Financial Economics 124, 1–21.

Wang, H., Zhou, H., Zhou, Y., 2013. Credit default swap spreads and variance risk premia. Journal of Banking & Finance 37, 3733–3746.

Whaley, R. E., 2009. Understanding the VIX. Journal of Portfolio Management 35, 98–105.

Chapter 3

Commodity Tail Risks

*Manuel Ammann, Mathis Mörke, Marcel Prokopczuk, Christoph Matthias Würsig**

ABSTRACT

In this study, we investigate the cross-section of tail risks in commodity markets. In contrast to findings from equity markets, left and right tail risk implied by option markets are both large. Moreover, we find that both left and right tail risk are priced in the cross-section of commodity futures returns. The variance risk premium is the main determinant for the left tail and the right tail risk. We find strong links to the tail risk and equity markets, as well as to speculation in commodity markets. In general, commodity specific variables exert the largest influence on tail risk, while there is no evidence of systematic commodity factors that are linked to tail risk.

Keywords: Commodities, Tail Risks, Dependencies

JEL classification: G10, G17, G12, C58.

*We thank Dr. Fabian Hollstein and participants at the Finance Seminar at the Leibniz University of Hannover for their helpful comments and suggestions.

1. Introduction

Commodity markets exhibit regularly reoccurring "supercycles" and "busts".[1] These cycles are generally accompanied by sharp increases or declines in prices due to various events, such as supply disruptions, demand shocks, political instabilities, or natural catastrophes. For market participants these risk are very crucial, because they represent high marginal utility events. Moreover, commodities are important for the real economy as production and consumption goods. Hence studying these events and understanding their dynamics constitutes an important analysis.

Whereas concerns regarding left tail risk dominate in equity markets, left and right tail risks play an equally important role in commodity markets. For commodity producers, negative price jumps (left tail risk) might have devastating consequences, akin to the role of left tail risks for market participants in stock markets. For commodity consumers though, e.g., companies that process commodities, right tail risks matter more. This is especially true when they need to fulfill long term contracts with a lack of future market supply of raw materials. Tail risks in commodity markets are also important for countries whose exports mainly rely on commodities. Moreover, commodity markets, and especially tail events, have a large influence on inflation and consumer spending (Garratt and Petrella, 2019). Thus an understanding of tail risks in commodity markets is crucial to understanding movements in the cross-section of commodity market returns and overall extreme risks in the market.

The contribution of our paper is twofold: First, we seek to identify the determinants of tail risk in commodity markets. We consider left and right tail risk for a wide cross-section of 19 commodities. In doing do, we control for past tail risk in order to account for the autocorrelation in time series. We posit multiple possible factors that may influence the tail risk in commodity markets and order them into the following groups: commodity specific factors, commodity market factors, and equity market factors.

Second, we analyse whether tail risk is priced in the cross-section of the commodity market. We find that it is. Tail risk carries a statistically significant risk premium for both, the right and left tail of the return distribution. The effect is stronger for right tail risk, than for left tail risk, with a return of the long-short portfolio of 10.60% p.a. for the right tail risk and 8.50% p.a. for the left tail risk, which indicates that there are different risk premia for commodity producers and consumers to hedge tail risks.

The commodity specific factors we identify are: basis, speculation, and the variance

[1] The Economist – Commodity prices are surging Jan 12 2021

risk premium. The basis is the slope of the futures term structure and is, as such, related to the supply and inventory of commodity markets. The inventory of commodities might influence the tail risk of commodity markets because speculators and insurance providers face limited risk capacity and financing constraints (Bianchi, 2018). Tail risk might also be related to speculation. Excessive speculation might lead to tail risk if investors face liquidity or short selling constraints. Option-implied tail risk is part of the variance under the risk-neutral measure. Therefore we also investigate links to the variance risk premium (Prokopczuk, Symeonidis, and Wese Simen, 2017).

We also identify commodity market factors: the average return of the commodity market, the return on a commodity momentum strategy, and the return on a commodity carry strategy. These factors are motivated by Bakshi, Gao, and Rossi (2017), who find that the average return of the commodity market, the momentum return, and the carry return can explain up to 71% of the variation in commodity futures returns.

Some authors argue that commodity markets are increasingly integrated with equity markets due to the financialization of commodities and the following influx of retail investors (Henderson, Pearson, and Wang, 2014). Hence, we include equity market factors that might capture links between the equity market volatility, equity market tail risk, and tail risks in commodity markets. In particular, we consider the VIX, the left or right tail risk of the S&P 500, and the market return of the S&P 500 index.

We find that commodity specific factors have the largest influence on left, right, total, and asymmetry tail risk. Different commodity specific factors seem to influence the left and the right tail. First, we find that the variance risk premium is closely linked to both left and right tail risk, as the corresponding coefficient is significant for most commodities in our sample. Furthermore, we find that speculation is negatively linked to left tail risk and to right tail risk, for the agricultural market. Left tail risk of the S&P 500 seems to be positively linked for most commodities in the case of the left tail and for some cases of the right tail. This effect is more significant and at least twice as larger for right tail risk. For the right tail the VIX seems to matter more than the right tail risk of the equity market. Commodity market factors only matter for the right tail and indicate a link to the average return of the overall commodity market. We find that both right and left tail risk seem to be influenced by equity market factors, and this mainly for the agricultural sector. This link is positive and statistically significant, which indicates that a higher tail risk in the equity market indicates a higher tail risk for agricultural commodities. The VIX is only linked to right tail risk for the agricultural market.

We also include total tail risk and asymmetry in our study, computed as the sum of

right and left tail risk and the difference between left and right tail risk, both divided by implied volatility. For asymmetry, most of the links from the variance risk premium persist while these links do not persist for total tail risk. We find that speculation in the agricultural and energy market is negatively linked to total tail risk, which indicates that tail risk is large in periods where there is little speculation. For asymmetry this link is positive and present for agricultural and metal markets. This implies that the impact of speculation is mostly influential for the left tail risk, which indicates that speculative activity dampens left tail risk more, than right tail risk. Overall, speculation in general seems to reduce tail risk, rather than increase it by providing liquidity. The S&P 500 return is mostly linked to total tail risk. Which, indicates that the S&P 500 is equally influenced by left and right tail risk.

Our paper is related to various strands of the literature. First, it contributes to the growing literature on option-implied tail risks in financial markets. Bollerslev, Todorov, and Xu (2015) analyse the importance of tail risks for predicting the aggregate stock market return. Kelly and Jiang (2014) find that U.S. single stocks with high loadings on an aggregated market tail risk factor outperform stocks with low tail risk loadings. Gao, Lu, and Song (2019) consider the cross-sectional pricing of tail risk in a global setting. They find that a global option-based tail risk index drives the cross-sectional return variation across different asset classes. Moreover, the authors do not include commodities in their analysis. Fan, Londono, and Xiao (2021) document that an ex-ante U.S. equity tail risk factor is priced in the cross-section of currency returns. Ex-ante tail risks in commodity markets have been mainly studied for single commodities in isolation. Using the approach of Bollerslev et al. (2015), Ellwanger (2015) and Ebrahimi and Pirrong (2020) investigate the predictive power of tail risks in the WTI crude oil market. A notable exception is Gao (2017), who includes ex-ante tail risk as a control in her analysis of the pricing of implied volatility in the cross-section of commodities. Gao (2017) also finds that left tail risk is priced, but the analysis does not extend beyond this fact. By studying ex-ante tail risks, we also complement the literature on price jumps in commodity prices (e.g., Nguyen and Prokopczuk, 2018; van Oordt, Stork, and de Vries, 2021).

Second, our paper contributes to the literature on identifying priced factors in the cross-section of commodity returns (e.g., Bakshi et al., 2017), especially to the pricing of higher implied moments in commodity markets. Chabi-Yo, Doshi, and Zurita (2020) introduce the concept of entropy risk in commodity markets. The authors find that the difference of entropy under the physical and risk-neutral measures has predictive power

for the cross-section of commodity returns.

Finally, our paper also relates to the literature analysing drivers of commodity returns and volatility, such as Büyükşahin and Robe (2014) and Kim (2015), who investigate the links between speculation and volatility. Additionally, Symeonidis, Prokopczuk, Brooks, and Lazar (2012) and Szymanowska, De Roon, Nijman, and Van Den Goorbergh (2014) link returns and volatilites to the commodity futures basis. Finally, there are a number of papers that investigate possible links between the equity market, various macroeconomic variables, and commodity markets (e.g. Christoffersen, Lunde, and Olesen, 2019; Hollstein, Prokopczuk, and Würsig, 2020).

The remainder of the paper is organized as follows: in Section 2 we present the data and methodology. In Section 3, we discuss the empirical results. We conclude in Section 4.

2. Data and Methodology

2.1. Data

We obtain data on commodity futures and options from the Commodity Research Bureau (CRB). Our data cover 19 commodities: soybean oil, corn, oats, soybeans, soybean meal, wheat, cocoa, cotton, orange juice, coffee, sugar, milk, feeder cattle, livestock, gold, copper, silver, crude oil, and natural gas. These are grouped in 5 sectors: agricultural, softs, livestock, metal, and the energy sector. The selected commodities are in line with the literature on commodities and dictated by the availability and liquidity of options data. In order to construct the tail risk measures, we use options on futures prices. Since the options on commodity futures prices are American type options, we approximate the early exercise premium and calculate the implied volatility following Barone-Adesi and Whaley (1987). We use this implied volatility to convert all in-the-money options into out-of-the money options using the model of Black (1976). CRB records strike prices by means of four digits. In case there is space left, CRB fills them with zeros on the right hand side, i.e., a strike price of 10 is recorded as 1000. In order to infer the correct strike price, we follow the procedure described by Hollstein, Prokopczuk, and Tharann (2021).[2] Next, we apply several data filters to our option sample. First, we keep only options with a time to maturity of at maximum 365 days. Second, we remove options with prices lower than four times the minimum tick size. Third, we discard options violating simple no-arbitrage restrictions. Our data covers the period from January 1996 to April 2020.

[2]Details are presented in Appendix A.

Table 1 lists the selected commodities with their respective sector, maturity months, and the number of available options.

Throughout the analyses, 3-month T-Bills are taken as the risk-free asset. Furthermore, we source options data on the S&P 500 index from OptionMetrics, data on the futures and options positions of various market participants (Commitments of Traders report) from the Commodity Futures Trading Commission (CFTC), and the VIX index from the CBOE.

Table 1: Data Overview

This table reports the 19 commodities contained in our sample. It provides the identifier (*Mnemonic*), names (*Name*), sectors (*Sector*), time period (*Period*), and the number of observations (*# Obs*) in our sample.

Mnemonic	*Name*	*Sector*	*Period*	*# Obs*
BO	Soybean Oil	Agricultural	Jan 1996 - Apr 2020	1,778,463
C-	Corn	Agricultural	Jan 1996 - Apr 2020	1,538,713
O-	Oats	Agricultural	Jan 1996 - Apr 2020	678,336
S-	Soybean	Agricultural	Jan 1996 - Apr 2020	2,686,906
SM	Soybean Meal	Agricultural	Jan 1996 - Apr 2020	1,830,590
W-	Wheat	Agricultural	Jan 1996 - Apr 2020	1,619,442
CC	Cocoa	Softs	Jan 1996 - Apr 2020	2,506,768
CT	Cotton	Softs	Jan 1996 - Apr 2020	3,476,720
JO	Orange Juice	Softs	Jan 1996 - Apr 2020	1,783,598
KC	Coffee	Softs	Jan 1996 - Apr 2020	4,242,466
SB	Sugar	Softs	Jan 1996 - Apr 2020	3,033,164
DE	Milk	Livestock	Jan 1996 - Apr 2020	2,679,425
FC	Feeder Cattle	Livestock	Jan 1996 - Apr 2020	1,927,744
LC	Live Cattle	Livestock	Jan 1996 - Apr 2020	1,236,194
GC	Gold	Metal	Jan 1996 - Apr 2020	6,849,366
HG	Copper	Metal	Jan 1996 - Apr 2020	4,682,776
SI	Silver	Metal	Jan 2005 - Apr 2020	5,908,771
CL	Crude Oil	Energy	Jan 1996 - Apr 2020	5,654,093
NG	Natural Gas	Energy	Jan 1996 - Apr 2020	2,878,146

2.2. Tail Risk

To measure tail risk, we use the approach proposed by Bollerslev and Todorov (2011). Bollerslev and Todorov (2011) construct a measure of tail risk perceived by investors that is based on close-to-maturity deep out-of-the-money options. Dierkes, Hollstein,

Prokopczuk, and Würsig (2021) show that this measure outperforms other tail measures proposed in the literature. This approach uses insights of the quadratic variation to decompose the volatility into two separate parts in a model-free way. To isolate tail risks, only deep out-of-the-money options are used, since only a large event will affect the prices of these options significantly. Specifically, let F_t denote the commodities futures price. We assume a quite general form for the dynamics of F_t, given as

$$\frac{dF_t}{F_t} = \alpha_t dt + \sigma_t dW_t + \int_{\mathbb{R}} (e^x - 1)\, \tilde{\mu}(dt, dx), \tag{1}$$

where α_t and σ_t are locally bounded drift and instantaneous volatility processes and W_t is a standard Brownian motion. The last term in Equation (1) accounts for jumps by means of the compensated jump measure $\tilde{\mu}(dt, dx) = \mu(dt, dx) - \nu_t^{\mathbb{P}} dt$ with $\mu(dt, dx)$ being a counting measure for the jumps and $\nu_t^{\mathbb{P}} dt$ denoting the compensator of the jumps under the physical measure $\mathbb{P}$. We require that $\nu_t^{\mathbb{P}} dt$ is predictable and $\int_{\mathbb{R}} (x^2 \wedge 1)\, \nu_t^{\mathbb{P}} dt$ is locally integrable. To move from $\mathbb{P}$ to $\mathbb{Q}$, we denote by $\nu_t^{\mathbb{Q}}(dx)dt$ the compensator for the jumps under $\mathbb{Q}$ and let λ_t denote the change of drift that turns W_t in a Brownian motion under $\mathbb{Q}$, i.e., $W_t^{\mathbb{Q}} = W_t + \int_0^t \lambda_s ds$. Subsequently, the total, expected quadratic variation of F_t under $\mathbb{Q}$ is

$$QV_t^{\mathbb{Q}} = \frac{1}{T-t}\mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T \sigma_s^2 ds\right) + \frac{1}{T-t}\mathbb{E}_t^{\mathbb{Q}}\left(\int_t^T \int_{\mathbb{R}} x^2 \mu(ds, dx)\right). \tag{2}$$

As Equation (2) shows, the expected quadratic variation can be decomposed into diffusive and discontinuous price moves, the latter being of interest for tail risks. As Bollerslev and Todorov (2011) show, if $T \downarrow t$ and $K > F_{t-} = \lim_{s \uparrow t} F_s$, then the price of a call with strike price K can be approximated by

$$e^{r^f_{(t,T]}} C_t(K) \approx \int_t^T \mathbb{E}_t^{\mathbb{Q}}\Big(\int_{\mathbb{R}} 1_{F_{s-}<K} max(0, F_{s-}e^x - K) v_S^{\mathbb{Q}}(dx)\Big) ds, \tag{3}$$

while for a put with strike price $K < F_{t-}$ it holds that

$$e^{r^f_{(t,T]}} P_t(K) \approx \int_t^T \mathbb{E}_t^{\mathbb{Q}}\Big(\int_{\mathbb{R}} 1_{F_{s-}>K} max(0, K - F_{s-}e^x) v_S^{\mathbb{Q}}(dx)\Big) ds. \tag{4}$$

Based on Equations (3) and (4), Bollerslev and Todorov (2011) define the model-free risk-neutral left and right jump tail measures as

$$LT_t^{\mathbb{Q}}(k) \equiv \frac{1}{T-t}\int_t^T \int_{\mathbb{R}} max(0, e^k - e^x)\mathbb{E}_t^{\mathbb{Q}}(v_S^{\mathbb{Q}}(dx))ds \approx \frac{e^{r^f_{(t,T]}}P_t(K)}{(T-t)F_{t-}},$$

and

$$RT_t^{\mathbb{Q}}(k) \equiv \frac{1}{T-t}\int_t^T \int_{\mathbb{R}} max(0, e^x - e^k)\mathbb{E}_t^{\mathbb{Q}}(v_S^{\mathbb{Q}}(dx))ds \approx \frac{e^{r^f_{(t,T]}}C_t(K)}{(T-t)F_{t-}},$$

where $k = log(K/F_{t-,\tau})$ denotes the log-moneyness.

For the estimation, Bollerslev and Todorov (2011) use options with at least 8 days to expiration and interpolate the option price to the desired moneyness levels. Bollerslev and Todorov (2011) set k equal to 0.9 for the left tail and 1.1 for the right tail. Most commodity options have no regular monthly expiration cycle, thus we set the limits further, to ensure we consider mostly jump days. We set the limits for k equal to 0.8 for the left tail and 1.2 for the right tail. This threshold is exceeded only 3% of the time within 30 days. Therefore this seems to be a reasonable value.[3] To obtain tail risks with a time to maturity of 30 days, we linearly interpolate or extrapolate the tail risk measures from the adjacent time to maturities. We calculate total tail risk as the sum of both tail risk measure. The asymmetry of the tails is captured as the difference between left and right tail risk, standardized by the implied volatility.

2.3. Futures Returns

We use futures returns of the first nearby contract throughout our analyses. To construct them, we roll futures contracts at the end of the month that is one month prior to maturity. We hold futures contracts with fixed maturity date, that is, we do not calculate returns of hypothetical futures contracts with constant time to maturity. By regularly rolling the first nearby contract, our returns entail only futures spot premia, not term premia (Szymanowska et al., 2014).

Let $F_{t,T}$ denote the price of the first-nearby futures contract at time t and expiration at T. The simple return on a fully collateralized futures position is then

$$r_{t+1} = \frac{F_{t+1,T} - F_{t,T}}{F_{t,T}} + r_t^f, \tag{5}$$

where r_t^f denotes the risk-free rate. Consequently, the corresponding excess return on a

[3]But we will estimate these certainly with more noise, compared to the equity indices.

fully collateralized futures position is given as

$$xr_{t+1} = r_{t+1} - r_t^f. \tag{6}$$

As apparent from Equation (5), returns are always computed by comparing prices from the same contract at different time periods. This is especially important when rolling futures contracts from one expiration month into the following expiration month. Hence, our returns are always constructed in a way that they are investable.

2.4. Explanatory Variables

Commodity Specific Factors

Commodity Basis For this measure we follow Gorton, Hayashi, and Rouwenhorst (2012). The basis is associated with inventory, where low inventory months are associated with above average basis and vice versa. Gorton et al. (2012) calculate the basis with the following formula

$$\left(\frac{F_{1,t}}{F_{2,t-1}} - 1\right) \times \frac{365}{(D_{2,t} - D_{1,t})}, \tag{7}$$

where $F_{1,t}$ is the nearest futures contract with at least 30 days time to maturity. $F_{2,t}$ is the next nearest futures contract. $D_{2,t}$ and $D_{1,t}$ represent the number of trading days until the expiration of the underlying futures contract.

Speculation We employ Working's T following Working (1960)

$$WorkingT = \begin{cases} 1 + \frac{SS_t}{HL_t + HS_t} & , SS_t > HL_t \\ 1 + \frac{SL_t}{HL_t + HS_t} & , otherwise, \end{cases} \tag{8}$$

where HS_t is the open interest of commercials that are short, HL_t is the open interest of commercials which are long, SS_t is the open interest of noncommercials which are short, and SL_t is the open interest of noncommercials which are long. We use this measure for the total tail risk and asymmetry. Because we assume that speculation matters only in one direction for the left tail risk, we use $\frac{SS_t}{TP_t}$, where TP_t are the total open interest (long and short) at time t. For the right tail risk, we use: $\frac{SL_t}{TP_t}$ (Bakshi et al., 2017). This measure is only available at the monthly frequency, so we extrapolate the last month's observation over the subsequent month to facilitate our daily analysis.

Variance Risk Premium We follow Carr and Wu (2009) and Prokopczuk et al.

Table 2: Left and Right Tail Risk

This table reports the summary statistics of the left tail risk on the left hand side and right tail risk on the right hand side. We order the commodities by sector. First, we have the agricultural sector, with soybean oil (BO), corn ($C-$), oats ($O-$), soybeans ($S-$), soybean meal (SM), and wheat ($W-$). Second, we have the softs sector, with cocoa (CC), cotton (CT), orange juice (JO), coffee (KC), and sugar (SB). Third, we have livestock, with milk (DE), feeder cattle (FC), and live cattle (LC). Fourth, we have metals, with gold (GC), copper (HG), and silver (SI). Finally, we have the energy sector, with WTI crude oil (CL) and natural gas (NG). We present the mean ($Mean$), the median ($Median$), the standard deviation (Std), the skewness ($Skew$), and the excess kurtosis ($Kurtosis$).

	Left Tail					**Right Tail**				
	Mean	*Median*	*Std*	*Skew*	*Kurtosis*	*Mean*	*Median*	*Std*	*Skew*	*Kurtosis*
Soybean Oil	11.17	8.24	10.20	2.32	6.95	27.29	23.33	15.78	1.19	1.52
Corn	17.84	13.09	15.68	1.46	1.96	37.63	33.47	21.11	0.77	0.09
Oats	23.95	21.35	11.63	1.46	2.65	46.76	44.29	18.48	1.09	3.04
Soybeans	12.90	8.98	13.44	2.46	6.7	29.87	24.64	20.07	1.33	2.01
Soybean Meal	16.36	11.99	16.68	3.46	18.89	34.21	28.93	20.30	1.81	5.05
Wheat	19.69	13.66	16.89	1.75	3.06	41.98	36.76	20.05	1.44	3.12
Cocoa	26.53	21.86	17.80	0.97	0.61	45.85	42.08	23.45	0.82	0.384
Cotton	18.59	12.55	16.99	2.62	13.43	31.14	26.25	19.05	2.10	6.72
Orange Juice	27.50	22.85	20.58	3.58	40.53	49.77	43.08	27.13	1.20	1.83
Coffee	36.92	33.92	19.98	1.65	5.53	73.05	66.28	34.70	1.50	3.60
Sugar	34.31	29.95	22.38	1.21	2.26	53.21	47.93	28.30	1.26	3.15
Milk	7.29	5.75	6.12	2.57	11.55	12.37	10.05	9.71	1.57	3.40
Feeder Cattle	7.26	5.12	7.60	5.21	56.51	6.03	4.17	6.29	4.92	45.95
Live Cattle	10.96	8.72	10.06	4.36	38.25	10.97	8.79	9.37	4.29	35.25
Gold	8.26	5.46	9.52	3.42	14.91	14.41	10.74	13.44	2.57	8.86
Copper	23.90	17.36	22.33	2.59	8.27	34.02	27.06	27.56	2.08	5.90
Silver	30.22	25.12	23.42	1.75	4.26	42.75	34.69	28.82	1.51	2.55
Crude	43.41	36.41	33.61	4.15	34.15	54.23	47.94	36.82	3.44	27.65
Natural Gas	71.04	63.78	39.28	1.11	2.18	97.14	88.54	51.50	1.08	2.24

(2017) to create synthetic 30-day overlapping variance swap rates. Then we subtract the variance swap rates over 30 days from the realized variance to obtain an estimate for the variance risk premium

$$VRP_{t,T} = RV_{t,T} - E_t^{\mathbb{Q}}(V_{t,T}),$$

$$E_t^{\mathbb{Q}}(V_{t,T}) = MFIV_{t,T} = \frac{2e^{r_t(T-t)}}{T-t}\left[\int_0^{F_{t,T}} \frac{P(t,K,T)}{K^2}dK + \int_{F_{t,T}}^{\infty} \frac{C(t,K,T)}{K^2}dK\right], \quad (9)$$

where $RV_{t,T}$ is the realized variance, computed as the sum of daily squared returns over the next month. $MFIV_{t,T}$ refers to the model-free implied variance between t and T. The annualized risk-free rate is denoted r_t. $F_{t,T}$ denotes the futures contract observed at time t and expires at time T. We truncate the first and second integrals at $F_{t,T} * e^{-10\sigma T}$ and $F_{t,T} \times e^{10\sigma T}$. σ is the average implied volatility of all out-of-the-money options. For each maturity we linearly interpolate the available Black (1976) implied volatilities across moneyness. We create a grid of 1,000 equidistant implied volatilities between the truncation threshold. We convert the implied volatilities with the Black (1976) option formula and calculate Equation (9). We linearly interpolate between the two swap rates, to obtain the 30-day swap rates.

Commodity Market Factors

Average For this factor, we calculate the return of the equally weighted commodity portfolio for every day.

Carry For this factor, we calculate the slope of the two nearest maturity futures for each commodity. The carry factor is $log(\frac{nearest}{secondnearest})$, where *nearest* is the next time to maturity, which cannot have the first notice day in the next month and *secondnearest* is the contract after this one. To ensure for commodities in contango (backwardation) a downward (upward) sloping curve, we require that the log differences is above (below) zero. Each day, we construct four portfolios, two contango and two backwardation portfolios. The daily return of the carry portfolio is the portfolio in which we go long in the equally weighted portfolio that is most in backwardation and short in the equally weighted contango portfolio.

Momentum To construct this factor, we invest in an equally weighted portfolio going long the five past winners over the last 12 months and shorting the portfolio with the

past five losers over the same period. We re-balance the portfolio daily.

Equity Market Factors

VIX Is the volatility index of the S&P 500.

Tail Variation of the S&P 500 For the tail risk of the S&P 500, we follow the same methodology as outlined in Section 2.2.

3. Empirical Results

3.1. Summary Statistics

Table 2 shows the summary statistics for daily left and right tail risk measures, while Figure 1 depicts their weekly time series.[4]

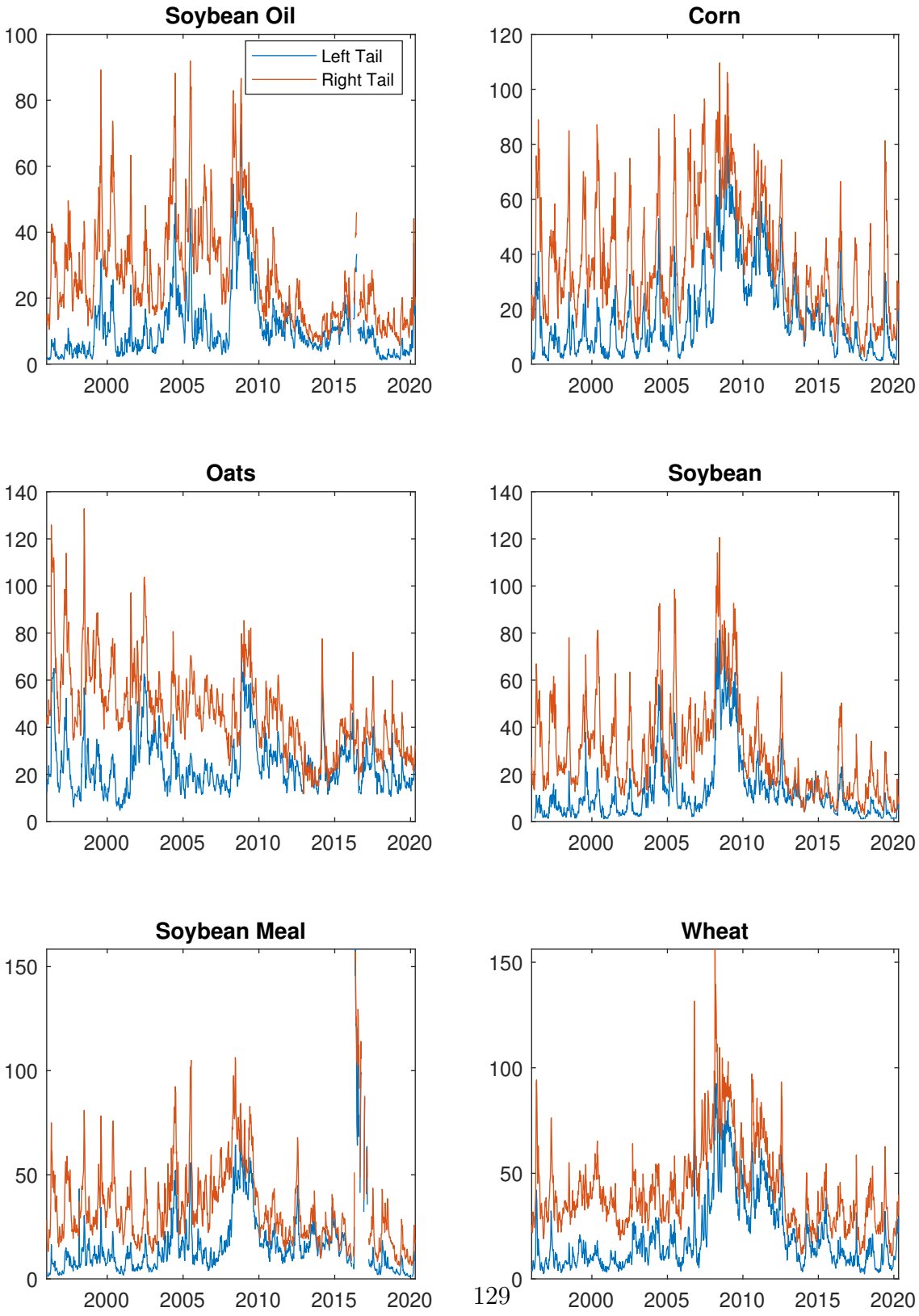
There are several noteworthy aspects. First, right tail risk is for most commodities larger than left tail risk. This contrasts findings for the equity index market. Bollerslev et al. (2015) show that right tail risk is negligible compared to left tail risk. However, our findings are more line with tail risk in single stock options, where left and right tail risk exhibit the same order of magnitude (Lin and Todorov, 2019). Second, right tail risk is not higher all of the time. As seen for example for livestocks, left and right tail risk might alternate. Third, the skewness for most tail risk measures is positive, which indicates that the distribution is left skewed. Fourth, the kurtosis is considerably larger for left than for right tail risk, indicating that crashes are larger and more immediate than booms in commodity markets.

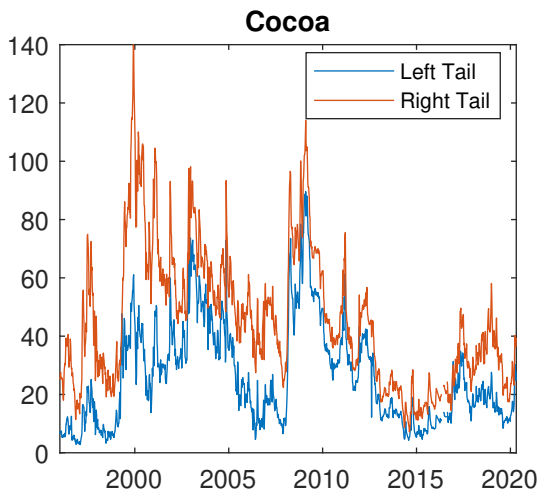
Moreover, the level of tail risks varies across commodities. Livestocks show comparably low readings of tail risk, contrary to energy commodities, with natural gas exhibiting the highest tail risk. For the softs market we observe higher levels, whereas agriculture and precious metals tend to exhibit lower degrees of tail risks.

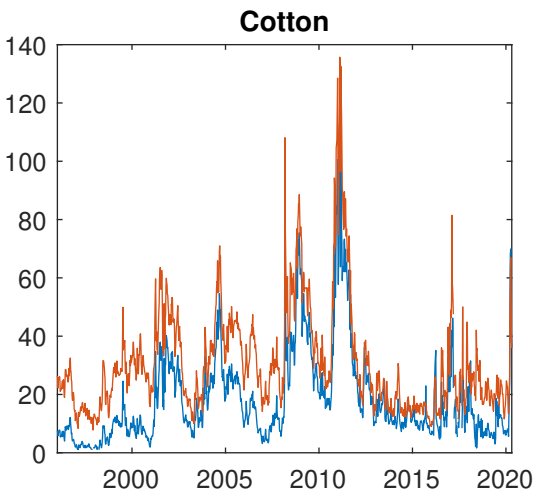
The crash in commodity prices after 2008 is mirrored in tail risks for most commodities as they exhibit a build-up of tail risk until 2009. This effect is especially pronounced for commodities which are more influenced by the business cycle, e.g., WTI crude oil and

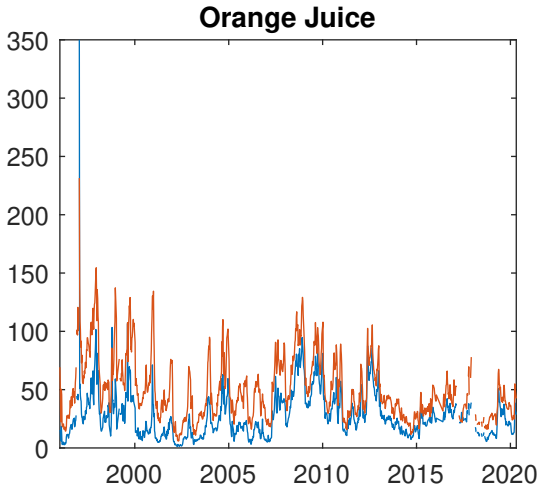
[4]As not all commodities have the same range of strike prices in their corresponding option market, we observe for very few occasions missing data for some commodities. In these occasions, there have not been enough options to yield a robust estimate for our tail risk measures.

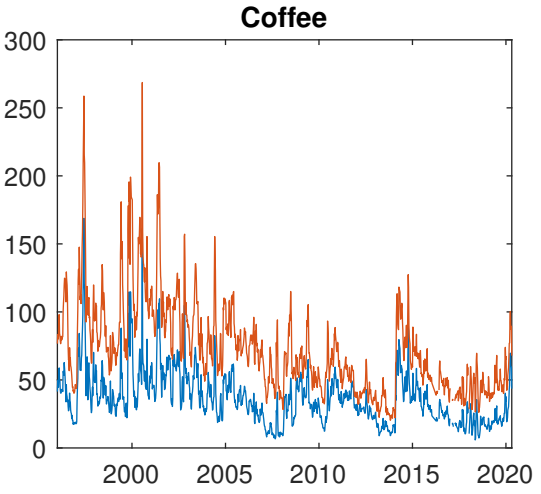
Fig. 1. Left and Right Tail Risk

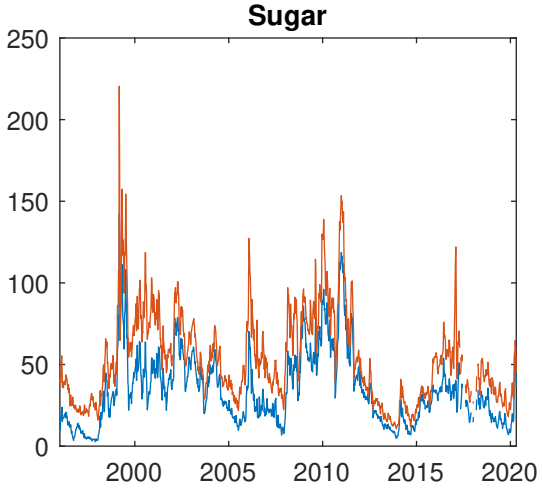


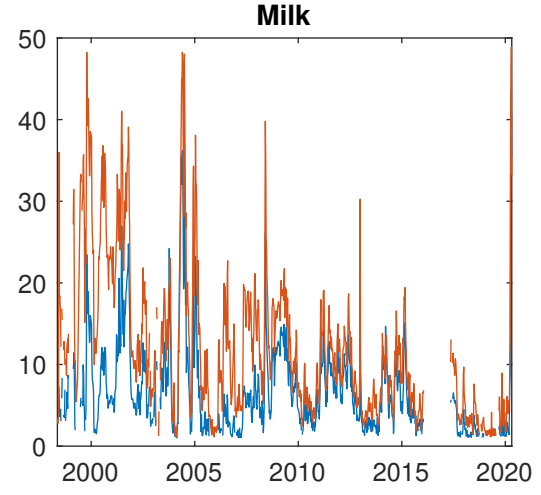

Cocoa
Left Tail
Right Tail
140
120
100
80
60
40
20
0
2000
2005
2010
2015
2020

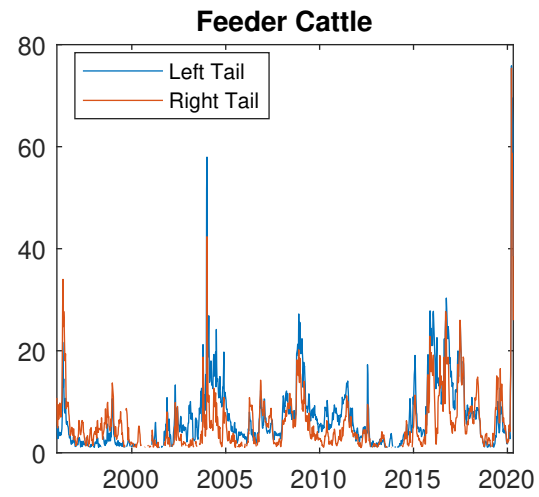

Cotton
140
120
100
80
60
40
20
0
2000
2005
2010
2015
2020

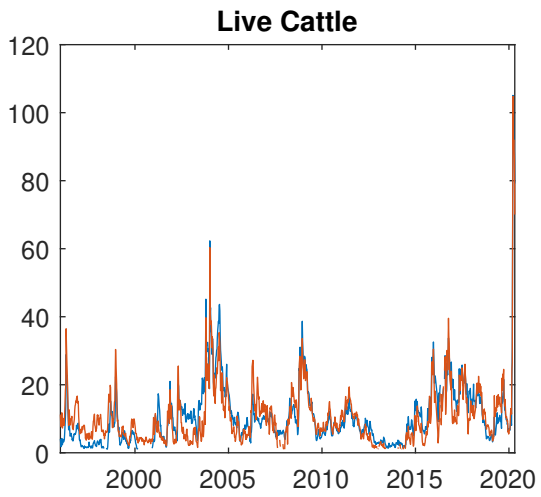

Orange Juice
350
300
250
200
150
100
50
0
2000
2005
2010
2015
2020

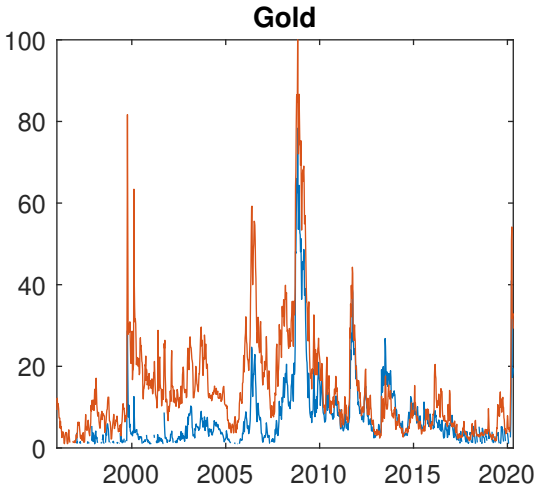

Coffee
300
250
200
150
100
50
0
2000
2005
2010
2015
2020

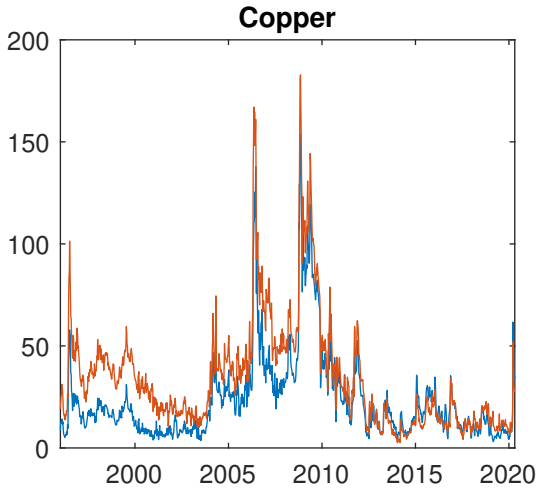

Sugar
250
200
150
100
50
0
2000
2005
2010
2015
2020

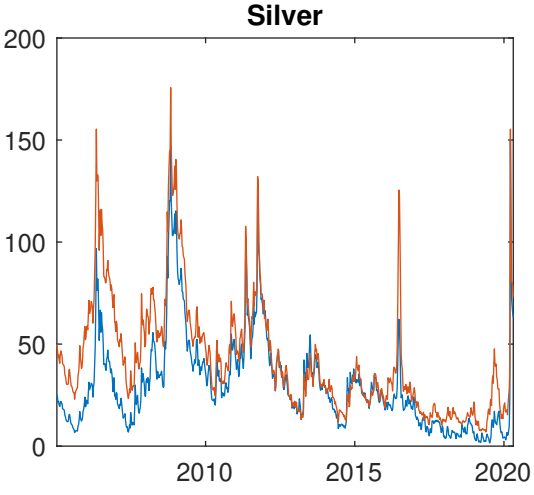

Milk
50
40
30
20
10
0
2000
2005
2010
2015
2020

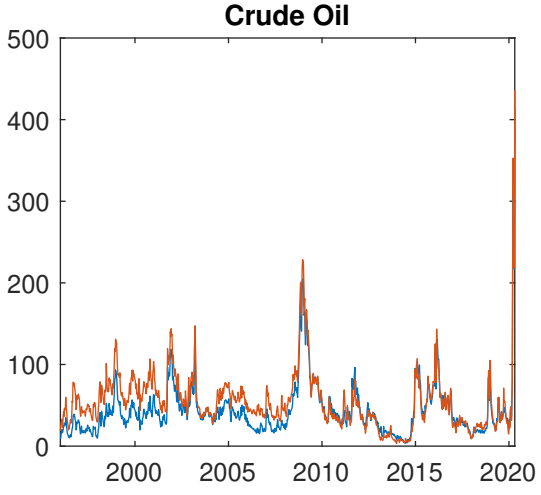

Feeder Cattle
Left Tail
Right Tail
80
60
40
20
0
2000
2005
2010
2015
2020

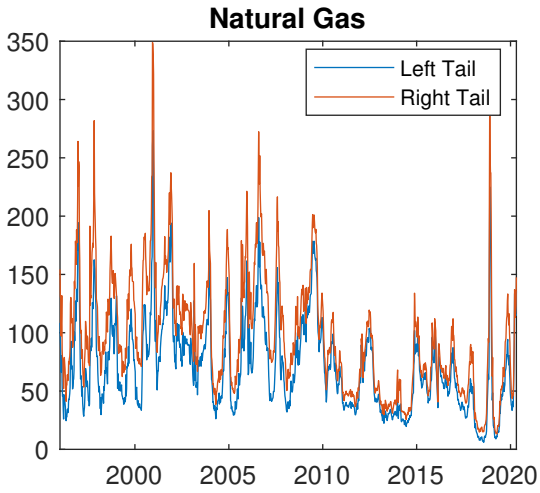

Live Cattle
120
100
80
60
40
20
0
2000
2005
2010
2015
2020


Gold
100
80
60
40
20
0
2000
2005
2010
2015
2020


Copper
200
150
100
50
0
2000
2005
2010
2015
2020


Silver
200
150
100
50
0
2010
2015
2020


Crude Oil
500
400
300
200
100
0
2000
2005
2010
2015
2020



copper. Additionally, though there seem to exist events affecting commodities at the same time, most of the time-series dynamics of tail risk appear to be commodity specific. Finally, and related to that, tail risks exhibit seasonality, presumably caused by the seasonal production/consumption cycle present in many commodity markets.

In Table 3, we present correlations of the left and right tail risk measures. We highlight in grey correlations within each sector. The correlations within the agricultural and metal sector are the highest, around 0.60 for left and right tail risk. Within the softs, livestock, and energy sector there seems to be substantially lower within sector correlations, they are around 0.30 in most cases. Cocoa, cotton, and orange juice seem to share a higher correlation with the agricultural market than within their own sector. Correlations for most tail risk measures across sectors seem to be moderately high. Gold and copper seem to have the largest correlations with the entire sector, which highlights their importance as economic predictors. The correlation between the economically important commodities in the metal market and crude oil are quite sizable at around 0.50. This might indicate that there are commodity-wide events that influence the risk of the entire market. Especially for the left tail, we also see correlations around 0.20 with crude oil.

3.2. Determinants of Tail Risk

In order to investigate the determinants of tail risk, we conduct regressions, that include commodity specific, commodity market, and equity market variables. As commodity specific variables we use: the basis to proxy for inventory risk, Workings T or the number of speculative long/short trading relative to the total trading to proxy for spec-

ulation, and the variance risk premium as proxy for the return on variance risk which should also be captured by tail risk. These variables are commonly used proxies for idiosyncratic behaviour of commodities.

As commodity market variables, we use the factors suggested by Bakshi et al. (2017): the average return of a commodity portfolio, the momentum return, and the return from a commodity carry strategy. These variables should capture any common variation between the tail risk measures in the overall commodity market. Lastly, we consider equity market specific variables, which are the volatility index, the tail risk of the S&P 500, and the overall market return.

With these variables, we estimate the following regression

$$TR_{j,t} = a + \gamma CS_{j,t} + \theta CM_t + \Omega EM_t + \beta TR_{j,t-1} + \sum_{k=1}^{11} \zeta_k S^D_{k,t} + \epsilon_{j,t}, \tag{10}$$

where $TR_{j,t}$ is the left or right tail risk of commodity j at time t, CS, CM, and EM are vectors of the commodity specific, commodity market, and equity market variables. $S^D_{k,t}$ represents a seasonal dummy variable, that is one if the observation is in month k and zero otherwise.

In Tables 4 to 7 we present the results, for the left tail risk, for right tail risk, and asymmetry (left minus right tail risk scaled by implied volatility).[5]

Overall, we find that commodity specific factors and the volatility of the equity market influence tail risk for all commodities or for particular sectors. Moreover, there seem to be some commodity market factors that influence specific commodities, but there does not seem to be a pattern across different commodity markets. For all tail risk measures, we see that the variance risk premium plays the by far most significant role to explain tail risks.

Focusing on left tail risk, in Table 4, we can observe that the coefficient estimates for the variance risk premium are between -0.19 and -3.89. This implies that a 1% increase in the variance risk premium can decrease the left tail risk by up to 3.89%. For the right tail risk, we see in Table 5 that the coefficient estimates range between -0.14 and -3.42, which implies that if the variance risk premium changes by 1%, we observe an decrease in the right tail risk of up to 3.42%. We see that the coefficient of the right tail risk has more links to the variance risk premium. While left tail risk has no link between the livestock market and the variance risk premium, right tail risk is statistically significant

[5]Total tail risk, i.e., the sum of left and right tail risk and tail asymmetry is the difference between left and right tail risk.

Table 3: Correlation of Tail Risk across Commodities

This table reports the correlation between multiple commodities. The left tail risk is on the lower half of the table and right tail risk on the upper half. We order the commodities by sector. First, we have the agricultural sector, with soybean oil (BO), corn ($C-$), oats ($O-$), soybeans ($S-$), soybean meal (SM), and wheat ($W-$). Second, we have the softs sector, with cocoa (CC), cotton (CT), orange juice (JO), coffee (KC), and sugar (SB). Third, we have livestock, with milk (DE), feeder cattle (FC), and live cattle (LC). Fourth, we have metals, with gold (GC), copper (HG), and silver (SI). Finally, we have the energy sector, with WTI crude oil (CL) and natural gas (NG). We present the correlation within a sector shaded in grey.

	BO	$C-$	$O-$	$S-$	SM	$W-$	CC	CT	JO	KC	SB	DE	FC	LC	GC	HG	SI	CL	NG
BO		0.51	0.49	0.77	0.70	0.33	0.37	0.22	0.12	0.25	0.18	0.21	0.08	0.09	0.26	0.33	0.50	0.13	0.06
$C-$	0.58		0.38	0.64	0.55	0.64	0.23	0.45	0.10	0.16	0.27	0.07	0.04	0.01	0.33	0.49	0.57	0.09	0.06
$O-$	0.26	0.33		0.46	0.34	0.24	0.29	0.15	0.11	0.34	0.18	0.24	−0.01	−0.04	0.10	0.24	0.36	0.12	0.20
$S-$	0.66	0.73	0.25		0.77	0.46	0.31	0.29	0.13	0.22	0.18	0.18	0.04	0.02	0.22	0.33	0.39	0.05	0.02
SM	0.47	0.33	0.12	0.41		0.43	0.17	0.22	0.13	0.17	0.12	0.17	0.11	0.06	0.19	0.28	0.33	0.05	0.01
$W-$	0.48	0.75	0.22	0.65	0.26		0.18	0.44	0.18	0.06	0.31	0.04	0.07	0.05	0.38	0.41	0.47	0.15	0.05
CC	0.29	0.24	0.20	0.25	0.07	0.21		0.13	0.25	0.36	0.25	0.35	−0.07	−0.02	0.38	0.20	0.49	0.23	0.19
CT	0.30	0.51	0.13	0.31	0.18	0.38	0.11		0.03	0.07	0.35	0.01	0.08	0.12	0.29	0.29	0.40	0.15	0.02
JO	0.13	0.15	0.01	0.17	0.11	0.11	0.08	0.04		0.15	0.11	0.21	−0.00	0.03	0.20	0.21	0.33	0.18	0.36
KC	0.00	−0.09	0.09	−0.07	−0.08	−0.07	0.22	−0.03	0.03		0.08	0.32	0.01	0.02	0.12	0.00	0.13	0.03	0.14
SB	0.19	0.36	0.18	0.21	0.09	0.27	0.34	0.35	0.06	0.01		0.13	−0.01	0.02	0.19	0.21	0.36	0.11	0.08
DE	0.02	−0.04	0.02	0.01	0.00	−0.02	0.11	0.03	0.12	0.15	0.09		0.06	0.10	0.07	0.03	0.08	0.08	0.09
FC	0.12	0.09	0.03	0.09	0.07	0.07	0.02	0.21	−0.01	−0.01	0.09	0.30		0.77	0.16	0.06	0.19	0.42	0.00
LC	0.17	0.07	0.00	0.09	0.07	0.06	0.00	0.19	0.00	0.01	0.05	0.28	0.59		0.20	0.08	0.26	0.53	0.06
GC	0.53	0.54	0.20	0.47	0.32	0.43	0.11	0.35	0.12	−0.03	0.15	0.12	0.31	0.34		0.58	0.80	0.45	0.23
HG	0.46	0.58	0.28	0.48	0.22	0.43	0.12	0.27	0.11	−0.09	0.14	0.00	0.16	0.13	0.51		0.75	0.28	0.30
SI	0.50	0.58	0.28	0.41	0.17	0.40	0.37	0.39	0.25	0.23	0.36	0.08	0.23	0.34	0.72	0.52		0.50	0.37
CL	0.33	0.35	0.22	0.26	0.14	0.27	0.21	0.27	0.08	0.03	0.19	0.22	0.48	0.59	0.55	0.32	0.54		0.25
NG	0.05	0.07	0.12	0.06	0.02	0.04	0.12	0.01	0.13	0.06	0.10	0.06	0.10	0.11	0.06	0.24	0.20	0.27	

Table 4: Determinants of Left Tail Risks

This table reports the results for the following regression

$$LT_{j,t} = a + \gamma CS_{j,t} + \theta CM_t + \Omega EM_t + \beta LT_{j,t-1} + \sum_{k=1}^{11} \zeta_k S^D_{k,t} + \epsilon_{j,t},$$

where CS is a vector of the commodity specific variables (basis, speculation, and variance risk premium). CM is a vector of commodity market factors (average portfolio return, return of a momentum portfolio, and return of a carry portfolio). EM is a vector for equity market factors (volatility index, left tail variation of the S&P 500, and the return of the S&P 500). Additionally we include the own lagged tail variation and seasonal dummies ($S^D_{k,t}$). We will not report the dummies for brevity. We scale the left tail variation to have a mean of zero and standard deviation of 1. We use Newey and West (1987) standard errors, the coefficient estimates in **bold** indicate a significance level of 5%. The t-statistic is indicated in brackets below the estimates.

	Agricultural						**Softs**				
	BO	*C−*	*O−*	*S−*	*SM*	*W−*	*CC*	*CT*	*JO*	*KC*	*SB*
Commodity Specific Factors											
Basis	0.22	**0.08**	**0.17**	0.04	0.03	0.17	0.14	**0.74**	**−0.45**	**0.75**	0.20
	(0.96)	(2.75)	(4.90)	(0.52)	(1.06)	(0.79)	(0.70)	(2.48)	(−2.42)	(3.98)	(1.49)
Spec	**−0.43**	**−0.81**	**0.34**	**−0.70**	−0.30	**−0.41**	0.35	0.10	−0.36	−0.27	−0.25
	(−2.16)	(−5.21)	(2.73)	(−2.77)	(−1.77)	(−2.28)	(1.00)	(0.26)	(−1.45)	(−1.31)	(−1.00)
VRP	−0.70	**−1.62**	**−0.24**	**−3.39**	−0.08	**−2.00**	**−1.06**	**−3.02**	**−3.83**	**−1.67**	−2.26
	(−0.86)	(−6.00)	(−4.05)	(−4.72)	(−1.04)	(−3.51)	(−2.61)	(−2.99)	(−2.91)	(−4.36)	(−1.63)
Commodity Market Factors											
Average	0.06	**2.71**	−0.61	1.81	0.97	**3.57**	−1.75	**5.22**	0.55	0.82	1.02
	(0.03)	(2.05)	(−0.66)	(0.82)	(0.59)	(2.23)	(−1.46)	(2.21)	(0.42)	(0.82)	(0.59)
Momentum	**1.99**	0.55	0.41	0.74	0.72	1.38	0.44	1.34	0.34	0.72	−0.05
	(2.11)	(1.39)	(0.85)	(1.20)	(1.24)	(1.84)	(0.70)	(1.38)	(0.67)	(1.41)	(−0.09)
Carry	**−2.03**	−0.24	0.18	−0.63	−1.44	−0.81	0.47	0.05	0.72	**−1.02**	0.08
	(−2.11)	(−0.42)	(0.29)	(−0.73)	(−1.48)	(−0.89)	(0.67)	(0.03)	(0.52)	(−2.24)	(0.17)
Equity Market Factors											
VIX	**−0.01**	0.00	**0.00**	0.00	**−0.01**	0.00	**0.02**	0.01	**0.01**	**0.01**	0.01
	(−2.25)	(0.59)	(2.07)	(−0.68)	(−2.14)	(−0.30)	(4.07)	(1.48)	(3.01)	(2.13)	(1.31)
LTSP	**3.95**	**0.88**	0.46	**2.45**	**1.86**	**1.76**	−0.23	**2.24**	−0.37	−0.47	−0.52
	(4.94)	(2.50)	(1.84)	(2.02)	(2.82)	(2.99)	(−0.61)	(4.17)	(−0.67)	(−1.07)	(−0.89)
*SP*500	2.87	−0.36	0.93	0.87	0.69	0.73	1.35	2.53	−0.24	−0.64	−0.36
	(1.75)	(−0.67)	(1.79)	(1.02)	(1.00)	(0.59)	(1.69)	(1.62)	(−0.28)	(−1.23)	(−0.95)
R^2	71.42	90.70	84.27	79.86	71.58	78.32	71.82	50.62	62.55	72.49	80.15
	Livestock			**Metal**			**Energy**				
	DE	*FC*	*LC*	*GC*	*HG*	*SI*	*CL*	*NG*			
Commodity Specific Factors											
Basis	0.04	**0.76**	0.13	0.00	0.12	−0.08	−0.02	**0.08**			
	(0.87)	(3.37)	(0.89)	(−0.03)	(1.30)	(−0.94)	(−1.11)	(2.43)			
Spec	−1.78	−0.71	−0.65	**0.37**	−0.05	−0.11	−0.15	**−0.59**			
	(−1.71)	(−1.59)	(−1.48)	(2.26)	(−0.49)	(−1.46)	(−0.42)	(−3.48)			
VRP	−1.16	−0.30	−1.73	**−3.89**	**−1.67**	**−0.19**	**−0.61**	**−1.61**			
	(−1.80)	(−0.38)	(−1.84)	(−3.36)	(−2.15)	(−2.79)	(−4.76)	(−5.61)			
Commodity Market Factors											
Average	−2.29	**2.02**	1.83	−1.64	−0.22	**−3.27**	**−2.82**	0.63			
	(−1.38)	(2.54)	(1.32)	(−1.76)	(−0.28)	(−3.41)	(−3.88)	(0.65)			
Momentum	3.58	−0.20	0.37	−0.14	−0.05	0.46	**0.62**	0.63			
	(1.93)	(−0.48)	(0.90)	(−0.31)	(−0.14)	(1.34)	(1.99)	(1.17)			
Carry	0.27	1.14	0.20	0.81	0.13	−0.28	−0.04	0.27			
	(0.40)	(1.92)	(0.39)	(1.40)	(0.41)	(−0.84)	(−0.16)	(0.56)			
Equity Market Factors											
VIX	**0.01**	−0.01	0.00	**−0.01**	0.00	0.00	0.00	0.00			
	(2.70)	(−1.88)	(−1.47)	(−4.52)	(0.59)	(1.02)	(1.85)	(0.44)			
LTSP	**−0.78**	**2.27**	**0.78**	**2.69**	0.38	0.46	**0.35**	−0.20			
	(−1.99)	(2.44)	(2.33)	(4.38)	(1.24)	(1.93)	(3.13)	(−0.74)			
*SP*500	1.56	0.40	0.29	1.35	−0.64	0.45	−0.65	−0.70			
	(1.10)	(0.90)	(1.14)	(1.56)	(−1.53)	(0.80)	(−1.87)	(−1.43)			
R^2	17.46	27.77	56.05	93.11	90.51	94.95	95.03	85.46			

Table 5: Determinants of Right Tail Risks

This table reports the results for the following regression

$$RT_{j,t} = a + \gamma CS_{j,t} + \theta CM_t + \Omega EM_t + \beta RT_{j,t-1} + \sum_{k=1}^{11} S^D_{k,t} + \epsilon_{j,t},$$

where CS is a vector of the commodity specific variables (basis, speculation, and variance risk premium). CM is a vector of commodity market factors (average portfolio return, return of a momentum portfolio, and return of a carry portfolio). EM is a vector for equity market factors (volatility index, right tail variation of the S&P 500, and the return of the S&P 500). Additionally we include the own lagged tail variation and seasonal dummies ($S^D_{k,t}$). We will not report the dummies for brevity. We scale the right tail variation to have a mean of zero and standard deviation of 1. We use Newey and West (1987) standard errors, the coefficient estimates in **bold** indicate a significance level of 5%. The t-statistic is indicated in brackets below the estimates.

	Agricultural						**Softs**				
	BO	*C−*	*O−*	*S−*	*SM*	*W−*	*CC*	*CT*	*JO*	*KC*	*SB*
Commodity Specific Factors											
Basis	0.14	−0.02	0.04	0.02	0.01	−0.05	**−0.27**	0.06	−0.21	**0.19**	0.12
	(0.56)	(−0.45)	(1.31)	(0.20)	(0.32)	(−0.35)	(−2.34)	(1.03)	(−1.83)	(2.80)	(1.51)
Spec	**−1.30**	**−2.86**	**−0.31**	**−1.02**	**−0.90**	**−1.03**	0.05	−0.04	**−0.21**	−0.23	**−0.72**
	(−5.27)	(−6.38)	(−3.29)	(−4.33)	(−5.58)	(−5.09)	(0.34)	(−0.45)	(−2.03)	(−1.85)	(−3.64)
VRP	0.30	**−1.81**	−0.06	**−2.37**	−0.03	**−2.09**	**−0.72**	**−0.75**	**−0.67**	**−0.94**	**−2.00**
	(0.47)	(−4.94)	(−1.88)	(−8.17)	(−0.96)	(−4.47)	(−4.10)	(−2.68)	(−4.88)	(−5.54)	(−1.97)
Commodity Market Factors											
Average	1.92	**6.48**	**2.18**	**6.36**	1.61	**4.67**	0.24	2.78	0.91	**3.94**	**1.98**
	(1.47)	(5.52)	(2.99)	(4.11)	(1.44)	(3.05)	(0.32)	(1.89)	(1.06)	(4.91)	(3.65)
Momentum	0.77	0.65	0.14	0.86	0.63	**1.28**	0.00	1.26	0.49	0.38	0.60
	(1.42)	(1.47)	(0.34)	(1.64)	(1.37)	(2.31)	(0.00)	(1.60)	(1.26)	(1.01)	(1.94)
Carry	−1.33	−1.14	−0.15	−0.03	**−1.26**	−1.50	0.33	−0.46	−0.28	**−1.56**	−0.23
	(−1.73)	(−1.62)	(−0.34)	(−0.04)	(−2.10)	(−1.42)	(0.63)	(−0.51)	(−0.49)	(−3.51)	(−0.56)
Equity Market Factors											
VIX	0.00	**0.01**	**0.00**	**0.00**	0.00	**0.00**	**0.01**	0.00	**0.00**	0.00	0.00
	(1.03)	(2.68)	(4.44)	(2.90)	(0.31)	(2.97)	(4.21)	(1.96)	(3.42)	(1.44)	(1.67)
RTSP	**1.57**	**1.11**	−0.10	0.27	**0.75**	**0.84**	−0.04	**0.51**	−0.08	−0.07	−0.58
	(4.32)	(4.09)	(−0.78)	(0.95)	(3.10)	(2.74)	(−0.20)	(2.62)	(−0.45)	(−0.49)	(−1.38)
*SP*500	0.16	0.13	−0.01	0.07	**1.42**	1.13	0.87	0.16	0.75	−0.53	0.10
	(0.27)	(0.17)	(−0.02)	(0.12)	(2.62)	(1.51)	(1.77)	(0.30)	(1.42)	(−1.13)	(0.23)
R^2	72.23	78.58	86.89	77.21	81.89	80.88	87.22	90.63	87.43	90.55	77.52

	Livestock			**Metal**			**Energy**	
	DE	*FC*	*LC*	*GC*	*HG*	*SI*	*CL*	*NG*
Commodity Specific Factors								
Basis	−0.01	0.05	0.02	−0.31	0.06	−0.05	0.03	**0.09**
	(−0.36)	(0.70)	(1.12)	(−1.94)	(1.07)	(−0.70)	(1.24)	(2.36)
Spec	−0.92	0.12	0.08	0.01	**−0.22**	**−0.42**	−0.08	**−1.14**
	(−1.16)	(0.53)	(1.29)	(0.12)	(−3.42)	(−3.99)	(−0.24)	(−2.79)
VRP	−0.51	**−1.31**	**−0.33**	**−3.42**	**−0.92**	**−0.14**	**−0.53**	**−1.54**
	(−1.34)	(−2.38)	(−2.95)	(−4.51)	(−5.78)	(−2.79)	(−3.95)	(−3.08)
Commodity Market Factors								
Average	1.64	**−3.39**	−0.47	1.27	−0.24	−0.53	**−2.22**	**2.29**
	(0.70)	(−2.61)	(−1.01)	(1.67)	(−0.35)	(−0.69)	(−3.48)	(2.88)
Momentum	0.08	0.15	0.39	−0.06	0.03	0.35	**1.00**	0.61
	(0.09)	(0.24)	(1.55)	(−0.16)	(0.10)	(1.14)	(3.62)	(1.30)
Carry	−0.83	0.88	−0.12	0.78	0.05	−0.19	0.29	−0.37
	(−1.10)	(1.54)	(−0.34)	(1.80)	(0.16)	(−0.66)	(1.11)	(−0.48)
Equity Market Factors								
VIX	**0.01**	0.00	0.00	0.00	**0.00**	0.00	**0.00**	0.00
	(2.00)	(−0.49)	(0.50)	(0.90)	(3.46)	(1.81)	(2.94)	(−1.54)
RTSP	−0.32	1.04	0.30	**1.58**	0.28	**0.55**	**0.46**	0.05
	(−1.46)	(1.89)	(1.60)	(6.19)	(1.51)	(2.31)	(4.00)	(0.25)
*SP*500	−0.35	**1.01**	0.40	0.08	−0.29	0.27	−0.32	**−1.24**
	(−0.38)	(2.14)	(1.90)	(0.12)	(−0.81)	(0.48)	(−0.89)	(−2.05)
R^2	70.49	57.84	86.86	94.25	95.35	95.86	94.56	89.02

Table 6: Determinants of Total Tail Risks

This table reports the results for the following regression

$$LTRT_{j,t} = a + \gamma CS_{j,t} + \theta CM_t + \Omega EM_t + \beta LTRT_{j,t-1} + \sum_{k=1}^{11} \zeta_k S^D_{k,t} + \epsilon_{j,t},$$

where CS is a vector of the commodity specific variables (basis, speculation, and variance risk premium). CM is a vector of commodity market factors (average portfolio return, return of a momentum portfolio, and return of a carry portfolio). EM is a vector for equity market factors (volatility index, total tail variation of the S&P 500, and the return of the S&P 500). Additionally we include the own lagged tail variation and seasonal dummies ($S^D_{k,t}$). We will not report the dummies for brevity. We scale the standardized total tail variation to have a mean of zero and standard deviation of 1. We use Newey and West (1987) standard errors, the coefficient estimates in **bold** indicate a significance level of 5%. The t-statistic is indicated in brackets below the estimates.

	Agricultural						**Softs**				
	BO	*C−*	*O−*	*S−*	*SM*	*W−*	*CC*	*CT*	*JO*	*KC*	*SB*
Commodity Specific Factors											
Basis	0.07	−0.05	−0.01	0.01	−0.08	**−0.62**	**−1.35**	0.18	−0.30	**0.41**	**0.16**
	(0.51)	(−1.44)	(−0.11)	(0.60)	(−1.31)	(−2.70)	(−3.02)	(1.07)	(−1.00)	(3.50)	(3.34)
Spec	0.30	**−0.77**	**−1.29**	−0.48	**−1.75**	**−2.69**	0.37	0.51	0.11	−0.01	**−0.76**
	(0.58)	(−4.84)	(−4.60)	(−1.60)	(−4.18)	(−5.08)	(0.55)	(0.95)	(0.74)	(−0.03)	(−3.64)
VRP	−0.14	**0.68**	**0.31**	0.79	−0.02	0.41	**−1.26**	−0.23	−0.31	−0.30	**0.44**
	(−0.26)	(3.96)	(3.07)	(1.21)	(−0.70)	(0.81)	(−2.44)	(−1.09)	(−1.05)	(−1.76)	(2.56)
Commodity Market Factors											
Average	3.95	−0.38	1.47	2.40	−0.45	−0.58	−1.27	−2.30	−5.98	−2.96	**1.69**
	(0.95)	(−0.64)	(1.05)	(1.47)	(−0.40)	(−0.56)	(−0.82)	(−0.84)	(−0.97)	(−0.90)	(2.84)
Momentum	−0.57	0.20	−0.55	−1.30	0.11	**1.41**	−1.77	0.38	−1.17	2.65	**1.30**
	(−0.96)	(0.85)	(−1.01)	(−1.06)	(0.21)	(1.99)	(−0.83)	(1.19)	(−1.18)	(1.37)	(2.62)
Carry	0.22	**−0.72**	2.46	−0.18	−0.52	0.14	−0.41	−1.28	3.26	−0.88	−0.49
	(0.87)	(−2.04)	(1.31)	(−0.14)	(−0.77)	(0.11)	(−0.50)	(−0.92)	(1.18)	(−1.72)	(−0.82)
Equity Market Factors											
VIX	0.00	**0.01**	0.01	0.00	**0.01**	**0.02**	**0.02**	0.01	0.00	**0.01**	**0.01**
	(1.06)	(3.98)	(1.91)	(0.49)	(2.97)	(3.61)	(5.04)	(1.10)	(0.52)	(2.68)	(3.79)
LTRTSP	0.07	**−0.30**	−0.46	0.03	−0.32	**−0.92**	**−0.87**	−0.33	−0.23	**−0.85**	**−0.61**
	(0.58)	(−2.68)	(−1.54)	(0.22)	(−1.61)	(−2.45)	(−3.37)	(−1.02)	(−1.11)	(−2.34)	(−3.26)
SP500	−0.99	0.41	0.50	0.23	**1.37**	0.77	0.65	0.02	−0.81	−0.38	−0.16
	(−0.98)	(1.59)	(0.88)	(0.26)	(2.11)	(0.64)	(1.46)	(0.16)	(−1.40)	(−0.84)	(−0.32)
R^2	0.22	11.53	10.14	0.35	43.98	27.62	3.33	4.12	0.60	1.23	80.56
	Livestock			**Metal**			**Energy**				
	DE	*FC*	*LC*	*GC*	*HG*	*SI*	*CL*	*NG*			
Commodity Specific Factors											
Basis	0.00	**0.56**	0.04	**−0.46**	0.03	−0.07	0.22	0.06			
	(−0.06)	(2.67)	(0.49)	(−3.12)	(0.39)	(−1.54)	(1.57)	(1.74)			
Spec	−0.23	−0.08	−0.51	0.17	**−0.59**	**−0.67**	**5.89**	**−2.30**			
	(−0.45)	(−0.25)	(−1.53)	(0.38)	(−4.04)	(−2.43)	(2.75)	(−6.15)			
VRP	0.54	0.67	**−0.50**	−0.12	−0.04	0.34	−0.05	−0.14			
	(0.75)	(1.58)	(−2.02)	(−0.21)	(−0.45)	(1.90)	(−0.16)	(−1.00)			
Commodity Market Factors											
Average	−0.33	1.00	0.13	0.62	−0.11	0.49	−3.14	0.35			
	(−0.25)	(0.77)	(0.09)	(0.75)	(−0.14)	(0.49)	(−1.30)	(0.37)			
Momentum	−1.50	−0.72	−0.02	0.32	−0.31	−0.70	1.28	−0.66			
	(−1.12)	(−1.14)	(−0.04)	(0.74)	(−0.75)	(−1.01)	(1.17)	(−1.19)			
Carry	3.40	1.54	0.58	0.53	0.18	−0.10	−0.23	−0.53			
	(1.27)	(1.41)	(0.62)	(1.24)	(0.26)	(−0.06)	(−0.36)	(−0.73)			
Equity Market Factors											
VIX	0.01	0.00	0.00	**0.01**	**0.00**	0.01	**0.04**	0.00			
	(1.94)	(−0.74)	(−1.52)	(3.52)	(3.17)	(1.69)	(3.81)	(1.47)			
LTRTSP	−0.38	0.54	**0.58**	−0.15	−0.13	0.01	**−1.35**	**−0.47**			
	(−1.31)	(1.84)	(2.58)	(−0.95)	(−1.33)	(0.03)	(−2.39)	(−2.16)			
SP500	−1.72	**1.35**	1.14	−0.06	0.71	−0.30	0.99	−0.40			
	(−1.02)	(2.07)	(1.82)	(−0.13)	(1.33)	(−0.38)	(1.50)	(−0.62)			
R^2	0.67	54.52	66.62	32.88	77.13	11.14	17.61	67.03			

Table 7: Determinants of Asymmetry

This table reports the results for the following regression

$$LTMRT_{j,t} = a + \gamma CS_{j,t} + \theta CM_t + \Omega EM_t + \beta LTMRT_{j,t-1} + \sum_{k=1}^{11} \zeta_k S^D_{k,t} + \epsilon_{j,t},$$

where CS is a vector of the commodity specific variables (basis, speculation, and variance risk premium). CM is a vector of commodity market factors (average portfolio return, return of a momentum portfolio, and return of a carry portfolio). EM is a vector for equity market factors (volatility index, tail asymmetry of the S&P 500, and the return of the S&P 500). Additionally we include the own lagged tail variation and seasonal dummies ($S^D_{k,t}$). We will not report the dummies for brevity. We scale the standardized asymmetry to have a mean of zero and standard deviation of 1. We use Newey and West (1987) standard errors, the coefficient estimates in **bold** indicate a significance level of 5%. The t-statistic is indicated in brackets below the estimates.

	Agricultural						**Softs**				
	BO	*C−*	*O−*	*S−*	*SM*	*W−*	*CC*	*CT*	*JO*	*KC*	*SB*
Commodity Specific Factors											
Basis	−0.05	**0.09**	**0.24**	0.00	0.10	**0.61**	**1.33**	−0.15	0.27	−0.15	−0.07
	(−0.36)	(2.06)	(3.16)	(−0.13)	(1.94)	(2.75)	(2.84)	(−0.85)	(0.95)	(−1.22)	(−1.83)
Spec	−0.32	**0.51**	**1.86**	0.46	**1.55**	**2.27**	0.36	−0.49	−0.04	0.21	**0.84**
	(−0.60)	(2.93)	(5.44)	(1.66)	(4.47)	(4.57)	(0.51)	(−0.95)	(−0.26)	(0.65)	(3.91)
VRP	0.22	**−0.87**	**−0.45**	−0.70	−0.01	**−0.83**	**1.17**	0.10	0.25	0.22	**−0.34**
	(0.43)	(−4.36)	(−3.83)	(−1.18)	(−0.34)	(−2.39)	(2.35)	(0.54)	(0.82)	(1.21)	(−2.27)
Commodity Market Factors											
Average	−3.98	0.32	−2.28	−2.17	**2.11**	0.33	0.67	2.27	5.12	1.55	**−2.15**
	(−0.97)	(0.55)	(−1.60)	(−1.51)	(2.24)	(0.34)	(0.37)	(0.91)	(0.90)	(0.51)	(−3.48)
Momentum	0.55	−0.24	0.58	1.32	−0.08	−1.06	2.54	−0.33	1.31	−2.53	**−1.02**
	(1.01)	(−0.96)	(1.03)	(1.05)	(−0.20)	(−1.73)	(1.15)	(−1.19)	(1.38)	(−1.37)	(−2.40)
Carry	−0.23	**0.76**	−2.09	0.36	0.35	−0.55	0.35	1.31	−3.56	1.13	0.26
	(−0.95)	(2.13)	(−1.08)	(0.31)	(0.59)	(−0.61)	(0.46)	(0.95)	(−1.35)	(1.73)	(0.52)
Equity Market Factors											
VIX	0.00	0.00	−0.01	0.00	0.00	**−0.01**	**−0.01**	0.00	0.00	−0.01	0.00
	(−1.25)	(−1.84)	(−1.56)	(−1.94)	(−1.44)	(−2.20)	(−2.92)	(−1.06)	(0.58)	(−1.76)	(−0.79)
LTRTSP	0.03	**0.55**	1.01	0.16	0.59	**1.72**	2.24	0.59	0.24	**2.42**	0.16
	(0.17)	(2.00)	(1.54)	(1.04)	(1.31)	(2.29)	(1.66)	(1.15)	(0.60)	(2.57)	(0.48)
*SP*500	1.04	−0.42	−0.52	−0.37	**−1.43**	−0.52	−0.17	−0.11	1.13	0.39	−0.41
	(1.01)	(−1.57)	(−1.08)	(−0.48)	(−2.56)	(−0.44)	(−0.37)	(−0.78)	(1.14)	(0.72)	(−0.82)
R^2	0.23	12.92	10.71	0.34	51.24	35.66	2.30	4.32	0.57	1.88	75.35
	Livestock			**Metal**			**Energy**				
	DE	*FC*	*LC*	*GC*	*HG*	*SI*	*CL*	*NG*			
Commodity Specific Factors											
Basis	0.01	**0.34**	0.18	**0.40**	−0.03	−0.02	−0.32	**−0.10**			
	(0.12)	(3.75)	(1.42)	(3.48)	(−0.40)	(−0.18)	(−1.66)	(−2.51)			
Spec	0.34	**−0.73**	**−1.52**	−0.18	**0.81**	**2.75**	−2.08	**3.65**			
	(0.84)	(−3.05)	(−4.00)	(−0.51)	(3.98)	(6.08)	(−0.91)	(9.44)			
VRP	−0.45	0.34	**−0.67**	−0.42	**−0.33**	−0.08	**−0.80**	**−0.29**			
	(−0.68)	(1.17)	(−2.34)	(−0.99)	(−2.62)	(−0.57)	(−1.99)	(−2.33)			
Commodity Market Factors											
Average	0.09	**3.07**	2.35	−0.99	0.45	**−2.02**	1.61	−2.08			
	(0.08)	(2.61)	(1.68)	(−1.38)	(0.73)	(−2.06)	(0.67)	(−1.64)			
Momentum	1.84	−0.93	−0.90	0.16	0.37	0.65	−2.18	0.70			
	(1.35)	(−1.36)	(−1.54)	(0.40)	(1.02)	(1.03)	(−1.91)	(1.16)			
Carry	−3.22	−0.02	0.12	−0.60	−1.04	0.91	0.33	0.20			
	(−1.23)	(−0.02)	(0.12)	(−1.51)	(−1.76)	(0.54)	(0.46)	(0.20)			
Equity Market Factors											
VIX	0.00	0.00	0.00	**−0.01**	**−0.01**	0.00	**−0.02**	0.00			
	(−1.18)	(0.00)	(−1.37)	(−3.64)	(−2.89)	(−0.82)	(−2.34)	(−1.72)			
LTRTSP	0.30	0.40	−0.13	**1.66**	0.17	0.15	2.41	**2.55**			
	(0.76)	(1.22)	(−0.36)	(2.66)	(0.47)	(0.26)	(1.66)	(3.40)			
*SP*500	1.44	−0.05	−0.62	0.06	**−0.92**	1.19	−0.94	0.25			
	(0.94)	(−0.09)	(−1.37)	(0.13)	(−2.00)	(1.76)	(−1.48)	(0.33)			
R^2	0.63	67.62	52.86	48.67	76.84	10.31	27.48	54.44			

for all markets and almost all commodities. This implies that tail risks are a significant part of the variance risk premium of commodity markets, which indicates that higher tail risk is associated with a significantly lower variance risk premium. Right tail risk is associated with higher coefficients for the variance risk premium, compared to left tail risk.

For both the left and right tail risk (in Tables 4 and 5) we observe links to the speculation in commodity markets. A higher proportion of short (long) speculators decrease left (right) tail risk substantially. This link, however, is only observable for the agricultural market for both left and right tail risk. For right tail risk this effect can also be observed for copper and silver. The effect is larger for right tail risk, compared to left tail risk, which indicates that speculative activity increases liquidity and therefore reduces uncertainty in the tails of the distribution. This finding is consistent with findings from Brunetti, Büyükşahin, and Harris (2016), who find that speculative activity reduces volatility for the agricultural market and natural gas.

Furthermore, we do find links to the VIX and the left and right tail risk of the S&P 500. These links exist for the agricultural, metal markets, and crude oil. Especially for left tail risk the coefficients are significantly larger; this indicates that especially in a period of a economic uncertainty this seems to occur in both the commodity market and equity markets. For right tail risk, we also observe a link between the average return on a commodity portfolio and some commodities. This link is especially present in the agricultural market, where it is positive. Hence, in tranquil periods, when returns are large, demand increases right tail risk substantially.

For the total tail risk and asymmetry of the tails in Tables 6 and 7, we still observe a link to speculation for both asymmetry and total tail risk. The asymmetry of the tails is still weakly linked to the variance risk premium, while the total tail risk is not. The total tail risk still has links to the total tail risk of the equity market and the VIX, while asymmetry has no such links. For speculation this shows that both left and tight tail risk are reduced by speculation, except for crude oil, but this effect is stronger for the right tail. The variance risk premium seems to be mainly driven by the left tail risk. While the link to the equity market disappears for asymmetry, it is still present for total tail risk, which shows that the common factor of the commodity left and right tail risk is crucial to understanding the linkages with the equity market.

In this section we have gained an understanding about the links of tail risk. In the following section, we investigate the economic importance of tail risk in the cross-section.

Table 8: Return of Portfolios Sorted by Tail Risk

This table reports descriptive statistics of commodity portfolios sorted on the tail risk measure at $t-1$ detrended by the previous 12-month average $\frac{1}{12}\sum_{\tau=1}^{12} TR_{t-\tau}$ (see Equation (5)). The "High" ("Low") portfolio entails commodities with highest (lowest) innovation in tail risk. Portfolios are rebalanced monthly. Returns are displayed in percentage and annualized. t-stats are based on Newey and West (1987) standard errors with optimal bandwidth with *, **, *** indicating significance at the 10%, 5%, and 1% level, respectively. Furthermore, we report the annualized standard deviation (Sdev), skewness (Skew), kurtosis (Kurtosis), Sharpe ratio (SR) and first lag auto-correlation coefficient (AC(1)). Panel A shows results for detrended left tail risk, whereas Panel B shows statistics for detrended right tail risk. The data sample is from January 1996 to April 2020.

	Mean	*t-stat*	*Sdev*	*Skew*	*Kurtosis*	*SR*	*AC*(1)
Panel A: Detrended Left Tail Risk							
High	3.16	0.93	14.89	0.07	0.64	0.21	0.04
2	0.43	0.12	15.25	−0.09	2.76	0.03	−0.01
3	2.04	0.63	14.21	−0.21	0.92	0.14	0.07
Low	−5.34	−1.40	19.32	−0.04	1.37	−0.28	−0.10
High–Low	8.50**	1.96	20.31	0.14	1.01	0.42	0.01
Panel B: Detrended Right Tail Risk							
High	1.71	0.51	14.87	0.21	1.59	0.12	0.00
2	4.60	1.50	14.45	−0.14	1.26	0.32	0.00
3	3.43	0.94	16.16	−0.24	1.90	0.21	0.11
Low	−8.89**	−2.27	18.49	−0.05	1.35	−0.48	−0.06
High–Low	10.60**	2.46	19.32	0.36	0.92	0.55	0.03

3.3. Portfolio Sorts

After having provided a descriptive analysis of tail risk, we now test if tail risk is priced in the cross-section of commodity futures. Therefore, we sort commodities each month into portfolio quartiles based on right and left tail risk, respectively. The commodities in our sample are quite heterogeneous with respect to their production region, period, storage suitability, and designated use. This might not only have an impact on commodity prices, but also on higher moments of the risk-neutral and physical return distributions. Figure 1 provides evidence in favour of heterogeneous risk-neutral tail risk. Consequently, we detrend the tail risk measures before comparing them cross-sectionally.[6] Thereby, we account for variation in their levels, but also their time-variation (see also Gao, 2017). At the end of each month, the detrended tail risk measure is given as

$$DTR_{j,t} = TR_{j,t} - \frac{1}{12}\sum_{\tau=1}^{12} TR_{j,t-\tau}, \tag{11}$$

where $TR_{j,t}$ denotes either the right or left tail risk measure of commodity j at time t. At the end of each month t, we allocate commodities into quartile portfolios. Portfolio 1 comprises commodities with highest right (left) tail risk innovation, whereas portfolio 4 contains commodities with lowest right (left) tail risk innovation. Commodities are equally weighted within each portfolio. We then construct a long-short strategy. The strategy buys commodities in portfolio 1 and shorts commodities in portfolio 4. We calculate the excess return for each portfolio and the return of the long-short portfolio at $t+1$.

Table 8 reports results. Panel A of Table 8 shows results when sorting by left tail risk innovations. In contrast to the right tail, the results suggest that innovation in left tail risk is also cross-sectionally priced. Though the long-short strategy yields an annualized return of 8.50%, it is significant at the 5% level. As Panel B shows, the long-short strategy that buys commodities with highest right tail risk innovation and sells commodities with lowest right tail risk innovation yields an annualized return of 10.60%, which is statistically significant at the 5% level, as well. The Sharpe Ratio is 0.55. Most of the strategy's return originates from the short leg. The portfolio with the lowest tail risk innovation yields an annualized loss of 8.89%, which is significant at the 5% level. Conversely, the portfolio with highest tail risk innovation returns 1.71% annually, which

[6]As Appendix B shows, our results remain quantitatively and qualitatively the same when we sort commodities according to the change in left and right tail risk, respectively.

is not significantly different from zero.

Next, we seek to determine whether the long-short returns can be explained by known risk factors in commodity markets. Therefore, we regress long-short portfolio returns on the set of risk factors proposed by Bakshi et al. (2017), comprising an average commodity factor, a commodity momentum factor, and a commodity carry factor. If the factors can price returns of our long-short strategy, the results should exhibit a small and statistically insignificant intercept, statistically significant slope coefficients, and a high adjusted R^2. Panel A in Table 9 details the results for the long-short portfolio based on left tail risk innovations. None of the three risk factors is significantly related to the strategy's returns. Panel B in Table 9 shows that the same is true for the long-short strategy based on right tail risk innovations. Only the carry factor is statistically significantly related to tail risk. This indicates that cross-sectionally tail risk seems to be related to inventory risk. Moreover, the intercept is significantly different from zero, economically large, and approximately matches its sample average. Finally, the adjusted R^2 is close to zero. Concluding, the long-short strategy cannot be explained by the three risk factors.

In Table B1 and B2, we report the same results for the monthly change in tail risk measures, showing that the detrending did not contribute to the cross-sectional predictive power.

Table 9: Risk Factors in the Tail Risk Strategies

This table reports descriptive statistics of regressing the monthly return of the tail risk measure strategy on a constant (Intercept), an average commodity factor (Commodity), a commodity momentum factor (Momentum), and a commodity carry factor (Carry). The factors are described in Bakshi et al. (2017). Panel A shows results for detrended left tail risk, whereas Panel B shows statistics for detrended right tail risk. The data sample is from January 1996 to April 2020. t-stats are calculated via Newey and West (1987) standard errors with optimal bandwidth with *, **, *** indicating significance at the 10%, 5%, and 1% level, respectively.

	Intercept	*Commodity*	*Momentum*	*Carry*
Panel A: Detrended Left Tail Risk				
beta	8.54^{*}	0.09	−0.01	−0.01
t-stat	1.94	0.74	−0.21	−0.16
Panel B: Detrended Right Tail Risk				
beta	11.21^{***}	0.07	0.05	-0.11^{*}
t-stat	2.59	0.76	1.00	−1.88

4. Conclusion

In this study, we study tail risk in commodity markets. We find that tail risk is priced in the cross-section of the commodity market. This is the case for right and left tail risk. Moreover, we investigate possible links with tail risk and numerous commodity factors. We find that the variance risk premium has the greatest influence for left and right tail risk. Speculation is only linked to tail risk in the agricultural, metal market and crude oil. We also document links to the equity market which seem to be driven by the common variation in left and right tail risk, which is general uncertainty. These links are particularly strong for the agricultural, metal, and energy market, and are likely exacerbated by the higher impact of speculative activity, which might lead to a transmission of uncertainty throughout markets.

Appendix

Appendix A CRB Data Handling

We follow the algorithm by Hollstein et al. (2021) that implements the algorithm to determine the "fair" strike price for each option, because CRB denotes every strike price with four digits; if there is space left, it fills them with zeros on the right hand side. For a commodity you might have the choice between a strike price of \$1.4, \$14, \$140, and \$1400. To calculate the value of an American option, we need to calculate two components, the early exercise payoff and an early exercise premium. Especially for short-term options early exercise premia are negligible, because we also have options up to 365 days into the term structure. But we argue that this approach is still feasible, because the difference between the potential strike prices is very large (by a factor of 10x), and the choice of the fair strike price is unlikely to be influenced by the early exercise premium even for longer maturity options.

To determine the fair strike, we only need to minimize the distance between the early exercise payoff and the price for an option:

$$\epsilon_{C,K_i} = |C - max(S - K_i, 0)|$$
$$\epsilon_{P,K_i} = |P - max(K_i - S, 0)|,$$

where ϵ_{K_i} refer to a call and put option. C, and P are the prices for call and put options. S and K_i represent the underlying futures and strike prices. With this data source you can make four guesses about potential strike prices, you can divide the price by 1000, 100, 10, or 1. The algorithm compares successively (from in-the-money to out-of-the-money) pricing errors, until the next pricing error is larger. The strike guess with the smallest error is the fair strike price.

In order to provide some robustness (as the strike price might be time varying), we compute the mode of the strike price for each contract. Further we ensure that the resulting options comply with monotonicity and no-arbitrage restrictions. We apply the following restrictions:

$$max(K - S_t, 0) \leq P_t \leq K$$
$$max(S_t - K, 0) \leq C_t \leq S_t$$

We calculate the implied volatility with the method of Barone-Adesi and Whaley (1987). As a further robustness check, we delete options with an implied volatility greater than

3 times the median implied volatility that day.

Appendix B Change in Tail Risk as Cross-sectional Predictor

Table B1: Return of Portfolios Sorted by Tail Risk

The table reports descriptive statistics of commodity portfolios sorted on the change in the tail risk measure at $t-1$. The "High" ("Low") portfolio entails commodities with the highest (lowest) innovation in tail risk. Portfolios are rebalanced monthly. Returns are displayed in percentage and annualized. t-stats are based on Newey and West (1987) standard errors with optimal bandwidth with *, **, *** indicating significance at the 10%, 5%, and 1% level, respectively. Furthermore, we report the annualized standard deviation (Sdev), skewness (Skew), kurtosis (Kurtosis), Sharpe ratio (SR) and first lag autocorrelation coefficient (AC(1)). Panel A shows results for changes in left tail risk, whereas Panel B shows statistics for changes in right tail risk. The data sample is from January 1996 to April 2020.

	Mean	*t-stat*	*Sdev*	*Skew*	*Kurtosis*	*SR*	*AC*(1)
Panel A: Change in Left Tail Risk							
High	1.93	0.65	15.92	0.10	0.56	0.12	−0.15
2	2.40	0.82	13.46	0.64	1.31	0.18	0.07
3	4.57	1.15	16.61	−0.18	2.19	0.28	0.11
Low	−7.54*	−1.75	18.68	−0.38	1.45	−0.40	−0.01
High–Low	9.47**	2.02	20.19	0.09	1.20	0.47	0.08
Panel B: Change in Right Tail Risk							
High	2.28	0.72	15.95	0.19	0.87	0.14	−0.12
2	0.67	0.24	12.97	0.63	1.66	0.05	0.02
3	3.52	0.87	17.08	−0.50	2.18	0.21	0.10
Low	−5.87	−1.38	17.97	−0.26	1.18	−0.33	0.02
High–Low	8.15*	1.81	18.86	0.27	0.94	0.43	0.09

Table B2: Risk Factors in the Tail Risk Strategies

The table reports descriptive statistics of regressing the monthly return of the tail risk measure strategy on a constant (Intercept), an average commodity factor (Commodity), a commodity momentum factor (Momentum), and a commodity carry factor (Carry). The factors are described in Bakshi et al. (2017). Panel A shows results for changes in left tail risk, whereas Panel B shows statistics for changes right tail risk. The data sample is from January 1996 to April 2020. t-stats are calculated via Newey and West (1987) standard errors with optimal bandwidth with *, **, *** indicating significance at the 10%, 5%, and 1% level, respectively.

	Intercept	*Commodity*	*Momentum*	*Carry*
Panel A: Change in Left Tail Risk				
beta	9.66**	−0.10	0.04	−0.03
t-stat	2.04	−0.94	0.77	−0.52
Panel B: Change in Right Tail Risk				
beta	8.75*	−0.11	0.02	−0.06
t-stat	1.94	−1.07	0.49	−0.99

Bibliography

Bakshi, G., Gao, X., Rossi, A. G., 2017. Understanding the Sources of Risk Underlying the Cross Section of Commodity Returns. Management Science 65, 619–641.

Barone-Adesi, G., Whaley, R. E., 1987. Efficient Analytic Approximation of American Option Values. Journal of Finance 42, 301–320.

Bianchi, D., 2018. Carry Trades and Tail Risk: Evidence from Commodity Markets Working Paper.

Black, F., 1976. The Pricing of Commodity Contracts. Journal of Financial Economics 3, 167–179.

Bollerslev, T., Todorov, V., 2011. Tails, Fears, and Risk Premia. Journal of Finance 66, 2165–2211.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail Risk Premia and Return Predictability. Journal of Financial Economics 118, 113–134.

Brunetti, C., Büyükşahin, B., Harris, J. H., 2016. Speculators, Prices, and Market Volatility. Journal of Financial and Quantitative Analysis 51, 1545–1574.

Büyükşahin, B., Robe, M. A., 2014. Speculators, Commodities and Cross-market Linkages. Journal of International Money and Finance 42, 38–70.

Carr, P., Wu, L., 2009. Stock Options and Credit Default Swaps: A Joint Framework for Valuation and Estimation. Journal of Financial Econometrics 8, 409–449.

Chabi-Yo, F., Doshi, H., Zurita, V., 2020. Never a Dull Moment: Entropy Risk in Commodity Markets Working Paper.

Christoffersen, P., Lunde, A., Olesen, K., 2019. Factor Structure in Commodity Futures Return and Volatility. Journal of Financial and Quantitative Analysis 19, 1083–1115.

Dierkes, M., Hollstein, F., Prokopczuk, M., Würsig, C., 2021. Measuring Tail Risk Working Paper.

Ebrahimi, N., Pirrong, C., 2020. Oil Jump Risk. Journal of Futures Markets 40, 1282–1311.

Ellwanger, R., 2015. Driven by Fear? The Tail Risk Premium in the Crude Oil Futures Market Working Paper.

Fan, Z., Londono, J. M., Xiao, X., 2021. US Equity Tail Risk and Currency Risk Premia Working Paper.

Gao, G. P., Lu, X., Song, Z., 2019. Tail Risk Concerns Everywhere. Management Science 65, 3111–3130.

Gao, L., 2017. Commodity Option Implied Volatilities and the Expected Futures Returns Working Paper.

Garratt, A., Petrella, I., 2019. Commodity Prices and Inflation Risk Working Paper.

Gorton, G. B., Hayashi, F., Rouwenhorst, K. G., 2012. The Fundamentals of Commodity Futures Returns. Review of Finance 17, 35–105.

Henderson, B. J., Pearson, N. D., Wang, L., 2014. New Evidence on the Financialization of Commodity Markets. Review of Financial Studies 28, 1285–1311.

Hollstein, F., Prokopczuk, M., Tharann, B., 2021. Anomalies in Commodity Futures Markets. Quarterly Journal of Finance forthcoming.

Hollstein, F., Prokopczuk, M., Würsig, C., 2020. Volatility Term Structures in Commodity Markets. Journal of Futures Markets 40, 527–555.

Kelly, B., Jiang, H., 2014. Tail Risk and Asset Prices. Review of Financial Studies 27, 2841–2871.

Kim, A., 2015. Does Futures Speculation Destabilize Commodity Markets? Journal of Futures Markets 35, 696–714.

Lin, H., Todorov, V., 2019. Aggregate Asymmetry in Idiosyncratic Jump Risk Working Paper.

Newey, W. K., West, K. D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 53, 1047–1070.

Nguyen, D. B. B., Prokopczuk, M., 2018. Jumps in Commodity Markets. Journal of Commodity Markets 13, 55–70.

Prokopczuk, M., Symeonidis, L., Wese Simen, C., 2017. Variance Risk in Commodity Markets. Journal of Banking & Finance 81, 136–149.

Symeonidis, L., Prokopczuk, M., Brooks, C., Lazar, E., 2012. Futures Basis, Inventory and Commodity Price Volatility: An Empirical Analysis. Economic Modelling 29, 2651–2663.

Szymanowska, M., De Roon, F., Nijman, T., Van Den Goorbergh, R., 2014. An Anatomy of Commodity Futures Risk Premia. Journal of Finance 69, 453–482.

van Oordt, M. R., Stork, P. A., de Vries, C., 2021. On Agricultural Commodities' Extreme Price Risk. Extremes 24, 531–563.

Working, H., 1960. Speculation on Hedging Markets. Food Research Institute Studies 30, 185–220.

Chapter 4

Option Return Predictability with Machine Learning and Big Data

Turan G. Bali, Heiner Beckmeyer, Mathis Mörke, Florian Weigert[*]

ABSTRACT

Drawing upon more than 12 million observations over the period from 1996 to 2020, we find that allowing for nonlinearities significantly increases the out-of-sample performance of option and stock characteristics in predicting future option returns. Besides statistical significance, the nonlinear machine learning models generate economically sizeable profits in the long-short portfolios of equity options even after accounting for transaction costs. Although option-based characteristics are the most important standalone predictors, stock-based measures offer substantial incremental predictive power when considered alongside option-based characteristics. Finally, we provide compelling evidence that option return predictability is driven by informational frictions and option mispricing.

Keywords: Machine learning, big data, option return predictability

JEL classification: G10, G12, G13, G14.

[*]We thank Jie Cao, Ilias Filippou, Andreas Neuhierl, Seth Pruitt, Yinan Su, Guofu Zhou, and Sebastian Stöckel for their insightful and constructive comments. We also benefited from discussions with seminar participants at the Virtual Derivatives PhD Workshop organized by the University of Illinois Urbana-Champaign and Michigan State University, at the University of Liechtenstein, an internal seminar at Goldman Sachs, an internal seminar at Hull Tactical Asset Allocation, the annual meeting of the German Academic Association of Business Research 2022, and the BVI-CFR 2021 seminar. We also thank Andrea Barbon for sharing open-close CBOE C1 exchange volume data with us.

1. Introduction

The importance of option markets has gained momentum over the past decade. According to data from the Futures Industry Association (FIA)'s annual statistical review, options trading on exchanges worldwide has increased from \$9.42 billion contracts in 2013 to \$21.22 billion contracts in 2020 – a growth rate of more than 125%. Approximately 60% of these contracts are written on individual stocks and stock indices, making equity the most popular underlying asset of financial market participants. Given the high popularity of options trading by investors, the question arises whether individual option returns are predictable and, if yes, which characteristics can give rise to such predictability. Our paper is devoted to answer these questions.

While classical option pricing models assume that options are redundant assets (Black and Scholes, 1973), more recent research rejects this idea and shows that option prices depend on other risks but the underlying's exposure (Buraschi and Jackwerth, 2001; Garleanu, Pedersen, and Poteshman, 2009). As an example, Goyal and Saretto (2009) document that the cross-section of option returns reflects a premium for variance risk, computed as the difference between historical realized volatility and at-the-money implied volatility. In this paper, we follow the idea of characteristics-based asset pricing and link future delta-hedged option returns to ex-ante characteristics drawn from both options and stocks. As we eliminate the directional impact of stock prices through our hedging procedure, we focus on risks which are inherently nonlinear and are likely to interact with each other in complex ways. Hence, the described setup is ideally suited for the application of machine learning models which are not only able to capture the impact of non-linearities and interactions between a large set of option and stock characteristics, but also mitigate the risk of in-sample model overfitting.

We study the cross-section of individual U.S. equity option returns using data from *OptionMetrics IvyDB* over the period from January 1996 to December 2020. To abstract from the directional exposure to the underlying, we follow Bakshi et al. (2003) and perform daily delta-hedges for each option as the market closes. Our main variable of interest is the monthly excess delta-hedged option return. After accounting for different filtering techniques, our dataset consists of more than 12 million option-month return observations of calls and puts, all written on individual U.S. stocks.

To predict future option returns we use a total of 270 variables composed of 77 option-based characteristics (e.g., option illiquidity, time-to maturity, or the implied shorting fee)

and 193 stock-based characteristics.[1] The stock characteristics include the 94 predictor variables proposed by Green, Hand, and Zhang (2017) to predict the cross-section of stock returns, 90 industry dummies, and additional characteristics that have been shown to be significantly associated with future stock returns (such as the bear beta proposed by Lu and Murray (2019), default risk of Vasquez and Xiao (2021), and the underlying's close price following Eisdorfer, Goyal, and Zhdanov (2020)). In the same fashion as Gu, Kelly, and Xiu (2020), we apply different linear and nonlinear machine learning models to form optimal predictions based on these option- and stock-based characteristics. Linear models included are penalized regression models (*ridge*, *lasso*, and *elastic-net*) and dimensionality reduction regressions (*principal component* and *partial least squares*). Nonlinear models comprise *gradient-boosted regression trees* with and without dropout, *random forests*, and fully-connected *feed-forward neural networks*. We also compute equal-weighted ensembles of all linear and all nonlinear models to leverage the informational content of the individual models.

To assess the predictive power of the different models for individual option returns, we follow Gu et al. (2020) and use the out-of-sample R^2-statistic, which benchmarks the R^2 against a forecast of zero excess returns.[2] To make pairwise comparisons of the forecast accuracy of different machine learning models, we utilize the model-free Diebold and Mariano (1995) test statistic.

Our empirical results advance the knowledge on predictability of the cross-section of individual option returns in various dimensions: First, we show that complexity of the prediction model matters. While none of the *linear* models manages to produce positive out-of-sample R^2s for the entire testing sample, all *nonlinear* models do. Our results reveal that the best-performing models are gradient-boosted regression trees with and without dropout (*GBR* and *Dart*) that show out-of-sample R^2s of 2.58% and 2.72%.[3] Moreover, the equal-weighted ensemble of all nonlinear models (denoted N-En) beats the ensemble of all linear models (denoted L-En) by more than 1.8% in out-of-sample R^2

[1]Option characteristics operate on three different levels: First, they can be the same for all options on the same underlying stock (e.g., the variance risk premium by Goyal and Saretto (2009)). Second, they can be classified on the individual option contract level (e.g., the options maturity). Third, they can be categorized on a bucket-level (e.g., the option bucket's trading volume), where buckets are formed based on the moneyness and time-to-maturity of the option.

[2]In addition, we apply the Han, He, Rapach, and Zhou (2021) cross-sectional out-of-sample R^2 which focuses on the prediction of cross-sectional option return spreads, instead of also including the time-series variation in returns.

[3]Note that the magnitude of these R^2s is considerably higher than the corresponding numbers for the cross-section of stock returns, e.g., Gu et al. (2020) find out-sample R^2 of approximately 0.6% for nonlinear machine learning models.

prediction power. Our results are confirmed when we compare pairwise forecast accuracy using Diebold and Mariano (1995) tests: The ensemble of all nonlinear models beats most other models significantly at the 5% level. The only exceptions to this rule are *GBR* and *Dart*, which both produce forecasts highly correlated with the nonlinear ensemble model (correlations amount to 0.95 and 0.94, respectively). The outperformance of nonlinear models compared to linear models is stable over time with a higher predictability for future option returns in 71.2% of the months in our sample.[4] Interestingly, we also find better predictions for the nonlinear models during December 2019 to April 2020, i.e., the period in which the COVID19 pandemic shook financial markets worldwide.[5] We can also show that our result that nonlinear models significantly outperform linear ones is not driven mechanically by predicting returns of all option contracts of an underlying.

Second, we inspect whether predictability of option returns through machine learning models can be exploited in an economically profitable trading strategy. Our results indicate that long-short portfolios based on L-En's and N-En's forecasts of expected returns generate economically significant return spreads of 1.28% and 2.02% per month, respectively, both statistically significant at the 1% level. The respective monthly Sharpe ratios amount to 1.03 and 1.38. The long-short return spread of the nonlinear ensemble outperforms the return spread of the linear ensemble by statistically significant 0.74% per month, stressing the importance of nonlinearities. We find that the short leg of the arbitrage portfolio, formed by the nonlinear ensemble, contains more short-term and fewer call options, while the long leg includes more long-term call options. The long-minus-short return spreads of the nonlinear ensemble are robust to risk adjustments of established asset pricing models, and remain significant across different states of the economy.

Ofek, Richardson, and Whitelaw (2004) show that transaction costs in the options market are high and that these costs can substantially reduce economic profits of option-based trading strategies. Hence, to understand in how far the machine learning trading strategy based on the nonlinear ensemble is implementable, we examine its profitability after accounting for transaction costs. Since actual transaction costs of trades are not observable in the OptionMetrics IvyDB database, we assume that investors have to pay a percentage of the quoted bid and ask spread, which we denote as the effective spread (Eisdorfer et al., 2020). We consider effective spreads of 25% – 100%. Our results

[4]Considering the cross-sectional out-of-sample R^2, this number increases to 85.6% of all months.

[5]Dew-Becker and Giglio (2020) show that the coronavirus epidemic is marked by an extraordinary high level of cross-sectional uncertainty, as measured by stock options on individual firms. Similar levels of cross-sectional uncertainty have been only witnessed during the tech boom in the early 2000s and the financial crisis.

show that the monthly returns of the machine learning trading strategy remain sizeable for 50% effective spreads. For short-term at-the-money options, which have been the focus in most previous studies, we find that the resulting trading strategies survive large effective spreads of more than 75%. Moreover, we show that an investment strategy which carefully weighs the initial transaction costs against the expected profits from trading on the model predictions survives 100% effective spreads.

As our third main empirical result, we quantify the relative importance of different characteristics for the prediction of option returns. We follow recent advances in computer science and estimate SHAP values (Lundberg and Lee, 2017), which approximate changes in the model predictions had we excluded certain characteristics in its estimation. First, we classify our 270 option and stock predictor variables into twelve sub-groups: *Accruals*, *industry*, *investment*, *profitability*, *quality*, *value*, *contract*, *frictions*, *illiquidity*, *informed trading*, *past prices*, and *risk*. Our results reveal that the *contract*-group contains the most important predictors, which includes information about the option's location on the underlying's implied volatility surface. *Illiquidity* and *risk* measures follow as the second and third most important variable group, respectively.

With respect to the relative importance of single characteristics, we find that *implied volatility* plays by far the most important role, followed by the *bid-ask spread* of the underlying stock and *industry momentum*. *Theta*, *delta*, and the *variance risk premium* are also highly important. We can furthermore assess the functional form of the impact of a single characteristic on the model-predicted delta-hedged return. We find that higher implied volatility contributes negatively to delta-hedged return predictions. Higher levels of the variance risk premium also negatively impact return predictions. In contrast, higher bid-ask spreads of the underlying increase the predicted delta-hedged return. Investigating this functional form also highlights the importance of nonlinearities: each of the ten most important characteristics impacts predicted returns in a nonlinear fashion. Additionally, we assess in how far proxies for volatility and jump risks impact return predictions. A high current variance risk premium is associated with negative predicted returns, higher gamma with positive predicted returns. We find that proxies for both, volatility and jump risks, are important and impact the predictions in a nonlinear fashion.

Our empirical setting enables us to answer whether option or stock characteristics are more important to accurately predict future option returns. Hence, we re-estimate the machine learning models (*i*) using only option-based characteristics, (*ii*) stock-based characteristics, as well as (*iii*) option-based characteristics that operate on the bucket- or contract level, and compare the out-of-sample forecasting results with the full information

set of *all* option and stock characteristics.[6] We observe that the models only based on a subset of information show severely lower out-of-sample R^2s compared to the models that incorporate all option and stock characteristics. When comparing different subsets of information, our results indicate that restricting information to only option-based characteristics yields substantially higher predictive R^2s than information of only stock-based characteristics. The benefit of option-derived characteristics is large and should not be neglected when making informed forecasts on future option returns.

Finally, we determine possible sources of return predictability. To do so, we first test if informational frictions are underpinning option return predictability. The presence of informed traders might cause faster impounding of information, more efficient price formation, and, hence, less predictability for option returns. To test this conjecture, we apply two proxies for the presence of informed investors in security positions: the share of institutional ownership and analyst coverage. In line with our prediction, we find that predictability of option returns is higher for underlying stocks which are characterized by low institutional ownership and low analyst coverage. In numbers, our results reveal that the out-of-sample R^2 for the nonlinear ensemble model equals 4.61% (2.22%) for options whose underlyings fall within the lowest (highest) quintile of institutional ownership. Options whose underlying stocks are followed by the fewest (most) analysts show an R^2 of 4.44% (0.58%). Moreover, when applying an additional dataset consisting of option transaction data for various types of market participants from four NASDAQ exchanges and the CBOE C1 exchange from January 2011 to December 2020, we observe that option return predictability is highest for options with little demand by professional traders, but substantial interest by public customers. We also find high levels of predictability when public customers are long the option and institutional investors take the short side of the contract.

We also estimate the level of mispricing in a given underlying's options. Consistent with the intuition that machine learning models manage to identify both situations of over- *and* under-pricing, we find high levels of R^2_{OS}, exceeding 2.5% in both cases. Turning to the profitability of the machine learning portfolios conditional on the level of mispricing, we show that betting on the portfolios for the subset of mispriced options is significantly more profitable than following the predictions for the least mispriced options. Thus, we conclude that option return predictability is largely driven by informational frictions and mispricing in the options market.

[6] In contrast to the feature importance by means of SHAP values, this approach has the benefit of correctly accounting for interaction effects between different feature groups.

The remainder of the paper is organized as follows. Section 2 reviews the literature and outlines our contribution. Section 3 describes alternative machine learning methods implemented in this study and explains how we evaluate the models' predictive power. In Section 4, we introduce the option return data and describe the option and stock characteristics used for prediction. Section 5 presents the main empirical results. Section 6 investigates the sources of option return predictability. We conclude in Section 7.

2. Related Literature

Our paper contributes to the literature on predicting and explaining the cross-section of individual option returns. Dennis and Mayhew (2002) document the importance of various factors, such as beta, size, and trading volume in explaining the risk-neutral volatility skew observed in stock option prices, whereas Bollen and Whaley (2004) investigate the relationship between net buying pressure and the shape of the implied volatility function of stock options. Garleanu et al. (2009) theoretically model and empirically investigate demand-pressure effects on option prices. Roll, Schwartz, and Subrahmanyam (2010) examine trading volume in option markets relative to the volume in underlying stocks and relate it to contemporaneous returns. Examining volatility risk in the options market, Goyal and Saretto (2009) find that options with a large positive difference between realized and implied volatility have low future delta-hedged returns. Cao and Han (2013) find that delta-hedged option returns decrease monotonically with an increase in the idiosyncratic volatility of the underlying stock. In a similar vein, Byun and Kim (2016) show that call options written on the most lottery-like stocks underperform otherwise similar call options on the least lottery-like stocks by more than 10% per month. An, Ang, Bali, and Cakici (2014) show, among others, that the cross-section of stock returns predicts option implied volatilities. Christoffersen, Goyenko, Jacobs, and Karoui (2018) include illiquidity premia in option valuation models and Kanne, Korn, and Uhrig-Homburg (2020) find that these premia are negative (positive) if there is net buying (selling) pressure. Ramachandran and Tayal (2021) report a monotonic relationship between various measures of short-sale constraints and delta-hedged returns of put options on overpriced stocks. Cao, Han, Tong, and Zhan (2021) uncover return predictability in the cross-section of delta-hedged equity options based on stock-based information, such as prices, profit margins, and firm profitability. Finally, Goyenko and Zhang (2021) apply machine learning techniques to analyze which characteristics drive option and stock returns. Our paper differs from Goyenko and Zhang (2021) in several important aspects: First, our

sample of option returns is substantially larger and comprises more than 33 times as many option-month observations. Instead of analyzing option portfolios as in Goyenko and Zhang (2021), we focus on single option contracts. Second, our paper focuses on the predictability of the cross-section of option returns, while Goyenko and Zhang (2021) also consider option determinants to predict the cross-section of stock returns. Finally, we provide a comprehensive investigation of the economic underpinnings of option return predictability and offer important insights on the cross-sectional pricing of equity options with machine learning and big data.

Moreover, we add to the empirical asset pricing literature employing machine learning techniques. So far, empirical asset pricing has mainly focused on using machine learning to predict the cross-section of individual stock returns.[7] Rapach, Strauss, and Zhou (2013) are the first to use Lasso in predicting market returns across countries. Moritz and Zimmermann (2016) apply tree-based conditional portfolio sorts to examine the relation between past and future stock returns. Kelly, Pruitt, and Su (2019) use the instrumental principal component analysis of Kelly, Pruitt, and Su (2020b) to study the cross-section of stock returns employing a latent factor model with time-varying factor loadings. Gu et al. (2020) compare several machine learning techniques for predicting cross-sectional stock returns based on a large set of stock characteristics. Whereas Gu et al. (2020) use a broad set of stock characteristics (see Green et al., 2017), Murray, Xiao, and Xia (2021) focus solely on historical price data. The authors find that machine learning based forecasts on past cumulative returns have economically and statistically significant predictive power for the cross-section of future stock returns. Neuhierl, Tang, Varneskov, and Zhou (2021) examine the predictive power of option characteristics for the cross-section of stock returns by jointly using broad sets of firm and option characteristics. Kozak, Nagel, and Santosh (2020) use a Bayesian estimator to construct a characteristic-sparse stochastic discount factor. Chen, Pelger, and Zhu (2021) combine three neural network structures to build an asset pricing model which enforces the no-arbitrage pricing condition. Lettau and Pelger (2020) propose a generalized principal component analysis including an additional penalty term to jointly explain the cross-section and time-series of stock returns. Freyberger, Neuhierl, and Weber (2020) use the adaptive group Lasso to identify important cross-sectional stock characteristics. Feng, Giglio, and Xiu (2020) and Giglio, Liao, and Xiu (2021) suggest different machine learning based techniques to evaluate the importance of a new asset pricing factor above and beyond already documented factors. Martin

[7]Nagel (2021) provides an overview of machine learning methods and the challenges involved when applying them to questions in empirical asset pricing.

and Nagel (2020) raise the importance of out-of-sample tests. In contrast to out-of-sample tests, in-sample return predictability tests might lose their economic meaning in the presence of a large set of predictor variables. Grammig, Hanenberg, Schlag, and Sönksen (2020) contrast theory-based and machine learning methods for measuring stock risk premia. The aforementioned studies have mainly focused on the cross-section of U.S. stocks. Leippold, Wang, and Zhou (2021), instead, employ machine learning algorithms to analyze return prediction factors in the Chinese stock market. Recently, the empirical asset pricing literature has applied machine learning methods beyond stock markets. Kelly, Palhares, and Pruitt (2020a) use instrumented principal component analysis to model corporate bond returns. Bali, Goyal, Huang, Jiang, and Wen (2021) successfully apply machine learning to predict the cross-section of corporate bond returns based on stock and bond characteristics, whereas Bianchi, Büchner, and Tamoni (2021) focus on Treasury securities. Filippou, Rapach, Taylor, and Zhou (2020) expand the literature by studying foreign exchange rates. DeMiguel, Gil-Bazo, Nogales, and Santos (2021) use machine learning models to predict mutual fund returns, whereas Li and Rossi (2020) apply machine learning in a fund holdings-based approach to optimally select mutual funds. Wu, Chen, Yang, and Tindall (2021) concentrate on the predictability of hedge funds via various machine learning methods. Our article is the first to predict the cross-section of *individual* option returns using a large set of linear and nonlinear machine learning models and the largest set of characteristics to date.

3. Methodology and Performance Evaluation

In its most general form, we express future option returns as the sum of expected returns and error term with zero mean:

$$r_{i,s,t+1} = \mathbb{E}_t[r_{i,s,t+1}] + \varepsilon_{i,s,t+1}. \tag{1}$$

The central element we aim to estimate is a functional representation $g(z_{i,s,t})$, which links expected future returns $\mathbb{E}_t[r_{i,s,t+1}]$ to characteristics $z_{i,s,t}$ of option i on underlying s:

$$\mathbb{E}_t[r_{i,s,t+1}] = g(z_{i,s,t}). \tag{2}$$

Methods considered: Following the growing literature on machine learning algorithms for predicting asset returns (Gu et al., 2020), we compare a variety of machine

learning methods with increasing complexity, and contrast the implications of linear and nonlinear models. For penalized linear models, we consider Lasso (Tibshirani, 1996), Ridge (Hoerl and Kennard, 1970) and Elastic Net regressions (Zou and Hastie, 2005, ENet). For linear dimension reduction techniques, we use principal component (PCR) and partial least squares regressions (PLS). To model nonlinearities, we consider tree-based methods: random forests (Breiman, 2001, RF), gradient boosted tree regressions (Friedman, 2001, GBR) and gradient boosted tree regressions with dropout (Gilad-Bachrach and Rashmi, 2015, Dart), as well as deep feed-forward neural networks (FFN) as universal function approximators (Hornik, Stinchcombe, and White, 1989). Online Appendix A provides a detailed description of these methods.

Forecast ensembles: We furthermore form ensembles of the five linear (L-En) and the four nonlinear models (N-En) to leverage the predictive power of multiple models (Goyal and Welch, 2008). We consider a simple ensemble, which equally weights the predictions of each method included. Building on the insights of Bates and Granger (1969), Rapach, Strauss, and Zhou (2010) document large benefits in economic forecasts using this type of ensemble. Denote the forecast of a given model by $\hat{r}_{i,s,t+1}$. Then, the ensemble forecasts for $t+1$ will be:

$$\hat{r}^{En}_{i,s,t+1} = \frac{1}{J} \sum_{j \in \mathcal{J}} \hat{r}^{(j)}_{i,s,t+1}, \tag{3}$$

where $\mathcal{J}$ contains the target models and J denotes the number of models in the set.

The design decision to create an ensemble of linear and one of nonlinear models allows us to directly analyze the informativeness of modeling nonlinear interactions and leverage the predictive power of multiple methods.

Assessing predictive power: We use the standard out-of-sample R^2 statistic to gauge the predictive power over single-equity option returns (Gu et al., 2020):

$$R^2_{OS} = 1 - \frac{\sum_{(i,t) \in \mathcal{T}_3} (r_{i,s,t+1} - \hat{r}_{i,s,t+1})^2}{\sum_{(i,t) \in \mathcal{T}_3} r^2_{i,s,t+1}}. \tag{4}$$

R^2_{OS} measures the reduction in the mean squared forecast error (MSFE) compared to a naive benchmark of zero excess returns for all options. We evaluate the predictive power on a testing sample, which is disjoint from the data used to estimate the model parameters and hyperparameters (such as the magnitude of the Lasso penalty). More specifically, we start by estimating model parameters on a training sample $\mathcal{T}_1$ of five

years (January 1996 – December 2000). We then perform an extensive hyperparameter optimization validating the method's fit in the next two years $\mathcal{T}_2$ (2001 – 2002). Lastly, for each method, we use an equal-weighted ensemble of the eight models with those hyperparameter combinations yielding the best fit in the validation sample to assess the predictive power in the one-year testing sample $\mathcal{T}_3$ (2003). We keep the models fixed for one year and replicate this procedure extending the number of years in the training sample by one year in each iteration, for a total of 18 out-of-sample years (2003 – 2020). Online Appendix B details the procedure we use to estimate the models, the libraries used for each model type, and the setup of the hyperparameter optimization.

In cross-sectional asset pricing tests, our main objective is not to forecast time-series variation in future returns, but rather cross-sectional return spreads in the testing sample. To focus on this cross-sectional variation, Han et al. (2021) propose a cross-sectional out-of-sample R^2,

$$R^2_{OS;XS} = 1 - \frac{\sum_{(i,t)\in\mathcal{T}_3}\left[(r_{i,s,t+1} - \bar{r}_{i,s,t+1}) - (\hat{r}_{i,s,t+1} - \bar{\hat{r}}_{i,s,t+1})\right]^2}{\sum_{(i,t)\in\mathcal{T}_3}(r_{i,s,t+1} - \bar{r}_{i,s,t+1})^2}, \tag{5}$$

which effectively compares the resulting cross-sectional return spread of a candidate model, $(\hat{r}_{i,s,t+1} - \bar{\hat{r}}_{i,s,t+1})$, with the realized return spread in the testing sample $(r_{i,s,t+1} - \bar{r}_{i,s,t})$. $R^2_{OS;XS}$ focuses on relative expected returns across options, for which accurate predictions result in profitable long-short trading strategies.

We test the statistical significance of each model's forecasts following Clark and West (2007), by comparing the resulting forecasts with a naive benchmark of always forecasting an excess return of zero:

$$\text{CW}^{(j)} = \frac{\bar{c}^{(j)}}{\hat{\sigma}_c^{(j)}}, \tag{6}$$

where $\bar{c}^{(j)}$ and $\hat{\sigma}_c^{(j)}$ denote the time-series average and Newey and West (1987) standard error of the mean difference between squared forecast errors:

$$c_{t+1}^{(j)} = \frac{1}{n_{\mathcal{T}_3}} \sum_{(i,t)\in\mathcal{T}_3} \left[r_{i,s,t+1}^2 - (\hat{e}_{i,t+1}^{(j)})^2\right]. \tag{7}$$

Here, $n_{\mathcal{T}_3}$ is the number of observations in the testing sample and $\hat{e}_{i,t+1}^{(j)}$ is the forecast error on option i at time $t+1$ for method j. We use 12 lags for the standard errors, coinciding with the number of months we keep model parameters fixed for each slice of

the testing sample.

Forecast comparison: To compare the forecasts of two methods, we use the modified Diebold and Mariano (1995) (DM) test, which accounts for potential cross-sectional dependence in equity option returns. The DM test-statistic for a comparison between methods 1 and 2 is defined as:

$$\mathrm{DM}^{(1,2)} = \frac{\bar{d}^{(1,2)}}{\hat{\sigma}_d^{(1,2)}}, \tag{8}$$

where $\bar{d}^{(1,2)}$ and $\hat{\sigma}_d^{(1,2)}$ denote the time-series average and Newey and West (1987) standard error of the mean difference between squared forecast errors $d^{(1,2)}$:

$$d_{t+1}^{(1,2)} = \frac{1}{n_{\mathcal{T}_3}} \sum_{(i,t)\in\mathcal{T}_3} \left[(\hat{e}_{i,t+1}^{(1)})^2 - (\hat{e}_{i,t+1}^{(2)})^2 \right]. \tag{9}$$

We also use correlations as a secondary method to assess how similar the forecasts of different methods are. Formally, the forecast correlation is defined as:

$$\rho_{t+1}^{(1,2)} = \frac{Cov(\hat{e}_{t+1}^{(1)}, \hat{e}_{t+1}^{(2)})}{\sigma(\hat{e}_{t+1}^{(1)})\sigma(\hat{e}_{t+1}^{(2)})}, \tag{10}$$

where $\sigma(\hat{e}_{t+1}^{(1)})$ and $\sigma(\hat{e}_{t+1}^{(2)})$ denote the standard deviations of forecast errors for models 1 and 2, respectively, and $Cov(\hat{e}_{t+1}^{(1)}, \hat{e}_{t+1}^{(2)})$ denotes their covariance. Divergence in the forecasting power of two methods provides a high-level view on why some methods outperform.

4. Data and Variable Definitions

We first outline the data sources used and then provide summary statistics for the option sample and the sample of underlying optionable stocks. Our primary data source is OptionMetrics IvyDB, which provides historical prices for all U.S. single equity options. We also use the interpolated volatility surface data from OptionMetrics. Due to the starting date of this database, our sample covers the period from January 1996 through December 2020.

Historical prices and accounting data for underlying stocks are obtained from CRSP and Compustat. We retain only underlyings with share codes 10 or 11 and exchange codes 1, 2, 3, 31, 32, 33; i.e., stocks listed on the NYSE, NYSE American (formerly AMEX)

or NASDAQ. Contrary to previous studies (see Cao et al., 2021), we purposely do not remove stocks with nominal prices below \$5 per share, as Eisdorfer et al. (2020) find that options trading on stocks with a low nominal price tend to be overpriced. Information on stock splits and dividend payments is taken from OptionMetrics and cross-checked with CRSP. We match these databases using the linking algorithm developed by WRDS. Daily risk-free rates are taken from Kenneth French's online data library.[8]

Option returns are notoriously noisy, especially for underlyings with few outstanding option contracts and less option trading activity. We therefore rely on a variety of standard filters established in the literature to assure the consistency of our analyses (Goyal and Saretto, 2009; Cao and Han, 2013; Cao et al., 2021).[9] First, we exclude all options for which OptionMetrics does not provide an implied volatility and Greeks. Second, we disregard options on stocks which have a dividend scheduled during the investment period. Third, we eliminate options with zero volume over the last seven calendar days. Fourth, to avoid any biases due to microstructure noise, we remove options for which the bid price is zero, the ask is smaller or equal to the bid, the mid price is below \$0.125, or the relative bid-ask-spread is above 50%. As a fifth step, we make sure that American option bounds are not violated. Sixth, we retain only options with a standard third-Friday expiration, such that we exclude short-term options with less than two weeks to expiration. Finally, we check for the convexity of option prices per underlying following Bollerslev, Todorov, and Xu (2015). Specifically, we retain only those observations for a given maturity τ, for which the difference between the prices of two neighboring call (put) options with strike price $K_1 < K_2$ is ≤ 0 (≥ 0).

4.1. Option Returns

Our main variable of interest is the excess return of buying an option that we delta-hedge on a daily rebalancing schedule. We consider delta-hedged option gains following Bakshi et al. (2003) as the value of a self-financing portfolio consisting of a long option, hedged by a position in the underlying such that the portfolio is locally immune to changes in the stock price. To establish notation, consider the partition $\Pi = \{t = t_0 < \cdots < t_N = t + \tau\}$ of the interval from t to $t + \tau$. Assume that the long option position is hedged discretely N times at each of the dates $t_n, n = 0, \ldots, N - 1$. The discrete delta-hedged

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

[9] To make the investment process as realistic as possible, we apply the filters only at the start of the trade, and assume that we have to use prevailing market quotes when we unwind the position or regard the position as worthless.

option gain over the period $[t, t+\tau]$ is then given by:

$$\Pi(t, t+\tau) = V_{t+\tau} - V_t - \sum_{n=0}^{N-1} \Delta_{V,t_n} \times [S(t_{n+1}) - S(t_n)] - \sum_{n=0}^{N-1} \frac{a_n r_n}{365} [V(t_n) - \Delta_{V,t_n} S(t_n)], \tag{11}$$

where V_t denotes the price of the option at time t, r_n is the risk-free rate at t_n, a_n is the number of calendar days between rehedging dates t_n and t_{n+1}, which we set to $a_n = 1$, and Δ_{V,t_n} is the observed delta of the option. We consider gains for investment horizons of one calendar month, or until maturity if the option expiration falls within the investment month. When an individual stock exhibits large price movements over the investment horizon, establishing initial delta-neutrality may still expose the investor to substantial sensitivity to future movements in the underlying stock. Tian and Wu (2021) estimate that one-time delta-hedging at initiation removes an average of about 70% of the directional risks embedded in the option, whereas daily delta-hedging manages to eliminate upwards of 90% of these risks. Hence, we opt for delta-hedging at the end of each trading day. Moreover, daily delta-hedging enables us to understand how characteristics relate to the pricing of higher-order risks embedded in options, detached from predictors for stock returns. Finally, we define option returns following Cao and Han (2013) and Cao et al. (2021) as:

$$r_{t,t+\tau} = \frac{\Pi(t, t+\tau)}{|\Delta_t S_t - V_t|}. \tag{12}$$

4.2. Summary Statistics

After the data filters discussed earlier, our sample comprises 4,932,372 options on 7,046 unique underlyings, for a total of 12,129,844 option-month observations for the period January 1996 – December 2020. Our sample is made up of roughly 53% call and 47% put options. Panel A of Table 1 shows that the average monthly delta-hedged option return is −0.01%, whereas the median monthly delta-hedged option return is −0.34%. The average moneyness is 1.03 and the average implied volatility is 47.67%. The average (median) days to maturity are 193 (113), while every fourth option exhibits a time to maturity of less than 50 days. As Panels B and C depict, the median monthly delta-hedged option returns are slightly positive for call options at an average 0.1% per month, but strongly negative for put options, −0.13%. The median return, however, is negative

for both puts and calls. Panel D in Table 1 shows summary statistics for the years of 1996 through 2002 which are used in the training step of the machine learning models, while Panel E gives the summary statistics only for the testing subsample from 2003 through 2020. Average monthly delta-hedged option returns tend to be slightly more negative for the more recent time period. Moreover, implied volatility and moneyness are lower and days to maturity higher for the testing period from 2003 to 2020.

Table E.1 in the Online Appendix reports summary statistics of the 7,046 stocks in our sample. Our sample includes on average 1,747 optionable stocks per month, which comprise 83% of the total market capitalization of the U.S equity market. Moreover, our sample comprises large stocks with representative volatility, given that the average size and volatility percentile within the total stock universe are 72 and 45, respectively. Finally, the Fama-French 12-industry distribution in our sample is comparable to the total stock sample, as evident from Panel C in Table E.1.

4.3. Option and Equity Characteristics

Throughout this paper, we differentiate between different parts of the time-to-maturity and moneyness domain of options, which we refer to as "buckets". Specifically, we separately consider predictability for short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (ITM: $K/S > 1.1$ for puts, $K/S < 0.9$ for calls), out-of-the-money (OTM: $K/S < 0.9$ for puts, $K/S < 1.1$ for calls) calls and puts, and at-the-money options (ATM: $0.9 \leq K/S \leq 1.1$), as well as time-to-maturity and moneyness combinations. The moneyness and time-to-maturity of option contracts change rapidly. Therefore, it is unreasonable to assume that flow measures, such as option volume, derived from a particular option contract over a historical period will be valid for the same contract in the next month. The defined buckets allow us to compute these flow measures for option contracts, as we abstract from the impact of changing moneyness and fleeting time-to-maturity. Table E.3 in the Online Appendix shows that on average, 15 long term options and 11 short term options belong to each underlying stock per month, with the bulk located at the current price of the underlying.

We build a comprehensive set of option-based characteristics, motivated by earlier studies on the cross-section of option and/or stock returns. Out of the 77 we compute, 42 characteristics operate on the level of the underlying (e.g., iv-rv-spread; Goyal and Saretto, 2009), 19 on the level of option buckets (e.g., the Amihud, 2002, illiquidity measure) and 16 on the level of individual option contracts (e.g., the option's time-

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: All Options (N=12,129,844)							
Delta-Hedged Return	-0.01	5.17	-4.37	-1.94	-0.34	1.29	4.47
Days to Maturity	172.24	179.44	21.0	50.0	113.0	204.0	444.0
Moneyness	1.03	0.31	0.77	0.89	1.0	1.11	1.28
Implied volatility	47.67	25.62	23.46	30.2	41.18	58.0	80.07
Absolute Delta	0.46	0.25	0.13	0.25	0.44	0.66	0.81
Panel B: Call Options (N=6,553,753)							
Delta-Hedged Return	0.1	5.62	-4.72	-2.01	-0.29	1.53	5.14
Days to Maturity	176.0	181.95	21.0	50.0	113.0	204.0	448.0
Moneyness	1.06	0.27	0.82	0.93	1.03	1.15	1.32
Implied volatility	47.01	24.26	23.23	29.99	40.98	57.61	78.85
Absolute Delta	0.51	0.24	0.18	0.32	0.52	0.71	0.84
Panel C: Put Options (N=5,576,091)							
Delta-Hedged Return	-0.13	4.59	-4.0	-1.87	-0.39	1.04	3.7
Days to Maturity	167.82	176.34	21.0	50.0	112.0	204.0	417.0
Moneyness	0.98	0.35	0.72	0.85	0.96	1.06	1.21
Implied volatility	48.46	27.1	23.74	30.44	41.41	58.48	81.6
Absolute Delta	0.4	0.24	0.1	0.19	0.36	0.57	0.77
Panel D: All Options 1996-2002 (N=2,026,636)							
Delta-Hedged Return	0.48	7.45	-5.67	-2.28	-0.11	2.39	7.04
Days to Maturity	152.5	174.2	22.0	50.0	108.0	176.0	330.0
Moneyness	1.08	0.44	0.78	0.9	1.02	1.16	1.4
Implied volatility	65.16	29.18	32.17	42.9	60.6	82.14	103.56
Absolute Delta	0.49	0.23	0.18	0.3	0.48	0.68	0.82
Panel E: All Options 2003-2020 (N=10,103,208)							
Delta-Hedged Return	-0.1	4.58	-4.14	-1.89	-0.37	1.12	3.97
Days to Maturity	176.2	180.21	21.0	50.0	113.0	206.0	448.0
Moneyness	1.01	0.28	0.77	0.89	0.99	1.1	1.26
Implied volatility	44.16	23.31	22.69	28.88	38.71	52.8	71.22
Absolute Delta	0.45	0.25	0.12	0.24	0.44	0.65	0.81

Table 1: Delta-Hedged Option Return Summary Statistics

The table reports descriptive statistics for delta-hedged option returns. Panel A reports statistics of returns and option characteristics over the period from 1996 to 2020. Delta-hedged option returns are measured over a period of one calendar month, or until option maturity. Delta-hedging is performed daily. Days-to-maturity is defined as the number of calendar days until option expiration. Moneyness is the ratio between the underlying's stock price and the option's strike price. Option implied volatility is provided by OptionMetrics. Absolute delta follows the model of Black and Scholes (1973). Panels B and C depict statistics for call and put options, respectively, whereas panel D reports summary statistics for the period 1996 to 2002 used exclusively in the training step of the models. Panel E reports summary statistics for the entire out-of-sample period, comprising the years 2003 through 2020.

to-maturity). Online Appendix C provides a detailed description of the option-based

characteristics.

As we are also interested in the performance of stock-based characteristics for predicting option returns, we include the 94 stock characteristics proposed by Green et al. (2017). We enrich this set by adding 90 industry dummies, based on the first two digits of the SIC code, four seasonal returns for each underlying (Heston and Sadka, 2008; Keloharju, Linnainmaa, and Nyberg, 2016), the bear-beta factor proposed by Lu and Murray (2019), and the previous period's return. Finally, we add stock-based characteristics motivated by the literature on the cross-section of option returns, but which are not included in Green et al. (2017). These comprise default risk (Vasquez and Xiao, 2021), the underlying's close price (Eisdorfer et al., 2020), and realized volatility (Cao, Vasquez, Xiao, and Zhan, 2019).[10]

We are left with a total of 270 characteristics, which can be broadly classified into 12 groups. *Accruals, Industry, Investment, Profitability, Quality* and *Value* exclusively include stock-based characteristics and loosely follow the classification proposed in Jensen, Kelly, and Pedersen (2021) and Green et al. (2017). Classification group *Contract* contains five option-based characteristics, the time-to-maturity, moneyness, implied volatility, and put and call identifier, and thus combines information about the location of the respective option on the underlying's implied volatility surface. We therefore assume that these characteristics play a pivotal role in predicting future option returns as a sort of reference point, much as the market return in traditional factor models (Fama and French, 1993). Groups *Frictions, Illiquidity, Informed Trading, Past Prices* and *Risk* contain both stock- and option-based information and highlight the predictions' dependence on informational frictions. Online Appendix D provides a detailed list of all 270 characteristics, their origin from the literature, the primary information source (option- vs. stock-based) and the feature group we have assigned them to.

5. Predicting Option Returns

5.1. Predictability Comparison

Figure 1 shows the out-of-sample R^2 for the pooled testing sample from January 2003 through December 2020 using the nine machine learning methods and the two

[10]The 94 stock-characteristics in Green et al. (2017) also include factors shown to have not only predictive power for the cross-section of individual stocks, but also option returns (e.g., idiosyncratic volatility as documented by Cao and Han, 2013).

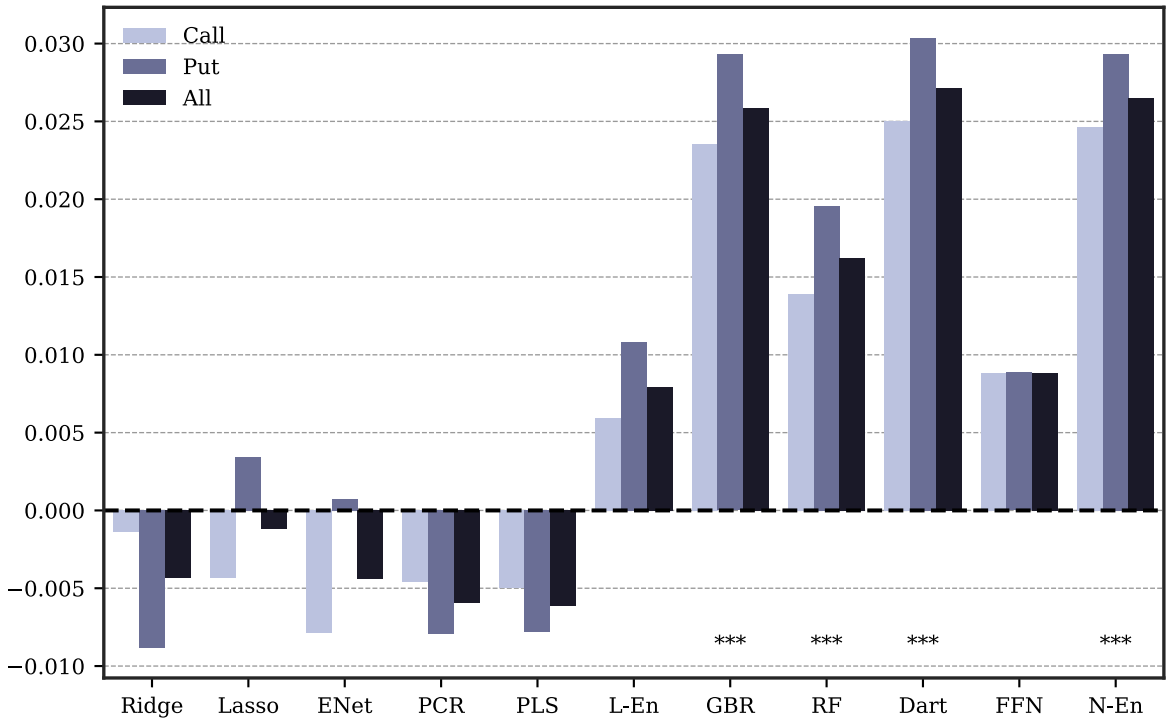


Fig. 1. R^2_{OS} Model Comparison

The figure shows out-of-sample R^2_{OS} as defined in Equation (4) for the nine models considered, as well as the linear (L-En) and nonlinear (N-En) ensemble methods. We separately document the predictive power for all options and for calls and puts. ***, **, * below the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7) for the sample of "all" options. The testing sample spans the years 2003 through 2020.

ensembles. Nonlinear models routinely outperform the predictability uncovered by linear models for option returns. The R^2_{OS} for the linear models on the full sample are all negative, ranging from −0.61% for PCR to −0.12% for Lasso regressions. The Clark and West (2007) test statistics indicate that none of these predictions outperforms a naive benchmark of predicting delta-hedged returns of zero.[11] Nonlinear models, in contrast, manage to uncover substantial predictability in single-equity option returns. GBR and Dart generate the highest out-of-sample R^2, above 2.5% for the pooled sample. All but FFN generate forecasts statistically better than the naive benchmark.

In addition to the full sample, we examine predictability for future call and put option returns separately. For most models we find that forecasts of future put returns yield

[11]We find that our results remain intact when the models' relative performance is evaluated against naive benchmarks predicting zero excess returns or the historical mean, either for the entire sample or subsamples by the respective bucket to which an option belongs at the time of the investment.

higher R^2_{OS}. The best call and put return predictions are both made by Dart, with $R^2_{OS} \geq 2.5\%$, the worst call predictions are generated by ENet, and the worst put predictions by Ridge. Linear models fail to adequately uncover a relationship between characteristics and future returns. While FFN is the most promising model in Gu et al. (2020) for stock returns, we find that it uncovers low predictability in the case of delta-hedged option returns. Given that FFN is the method with the highest potential complexity, this suggests that this complex structure does not generalize well to the testing sample of option returns. In contrast, tree-based methods tend to outperform, suggesting that histogram-based estimation including *nonlinear interactions* trumps model complexity in this market.[12]

Ensembles have been shown to improve the accuracy and consistency of the predictions and are a staple in modern machine learning estimation (e.g., Krizhevsky, Sutskever, and Hinton, 2012; Lakshminarayanan, Pritzel, and Blundell, 2016).[13] The linear ensemble L-En produces significantly better forecasts than any of the individual linear models, notably managing to produce positive R^2_{OS} for call and put returns. The predictability uncovered is comparable to that of FFN, but again not statistically significant. The level of predictability at roughly 0.8% is about twice as high as the levels of predictability of stock returns found by Gu et al. (2020). Within the class of nonlinear methods, Dart tends to outperform N-En for the full sample, as well as separately for puts and calls. Just as all tree-based methods, the resulting predictions comfortably beat the naive benchmark of zero excess returns.

We are ultimately interested in how far the models uncover cross-sectional return spreads in our option sample. For this, Figure 2 compares the cross-sectional $R^2_{OS;XS}$ defined in Equation (5). Interestingly, while none of the individual linear models generated predictability on the full sample, especially penalized regressions are able to generate realistic return spreads. For Lasso and ENet these are even statistically significant at the 5% and 1% level, respectively. The cross-sectional pricing power of all nonlinear models is highly significant, and surprisingly FFN manages to produce $R^2_{OS;XS}$ comparable to the tree-based methods, despite failing to adequately predict the average level of future returns. This exercise clearly highlights the benefits of using ensemble models, in that the cross-sectional predictive power increases both for L-En and N-En. The nonlinear ensemble generates the highest cross-sectional predictability for all subsamples considered

[12]It is well-known in the machine learning literature that tree-based methods outperform neural networks on tabular data (Arik and Pfister, 2019). Tree-based methods routinely win corresponding Kaggle competition.

[13]Steel (2020) discusses its uses in economics.

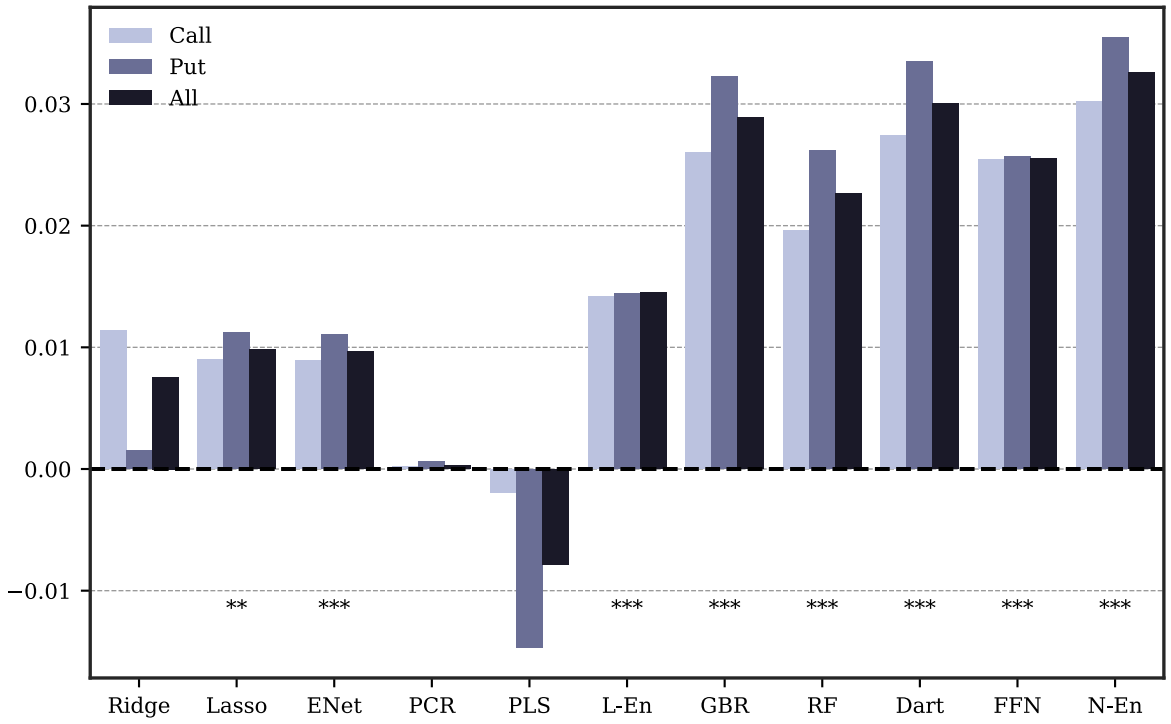


Fig. 2. $R^2_{OS;XS}$ Model Comparison

The figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) for the nine models considered, as well as the linear (L-En) and nonlinear (N-En) ensemble methods. We separately document the predictive power for all options and for calls and puts. ***, **, * below the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7) for the sample of "all" options. The testing sample spans the years 2003 through 2020.

with an $R^2_{OS;XS}$ of 3.3%.

Most studies up-to-date overlook the ability to zoom in on the predictive power of the models considered and provide an intuition of how stable the resulting predictions are. The focus mostly lies on pooled out-of-sample predictability as in Figure 1 and Figure 2. Instead, we also show the dispersion of R^2_{OS} and $R^2_{OS;XS}$ over time. While the pooled approach weights the predictions of each year by the number of option-month observations contained, we now provide estimates of the predictive power per year, which allows us to investigate the stability of the forecasts over time. Stable predictability is imperative for investors who wish to use the model forecasts in their investment decisions.

The upper panel of Figure 3 adds three main points to the pooled R^2_{OS} consideration above. First, the predictive power of all models fluctuates significantly over time. Linear models produce the largest dispersion, with the possibility of highly negative R^2. Second, most models exert an interquartile range of their predictive power that is either symmetric

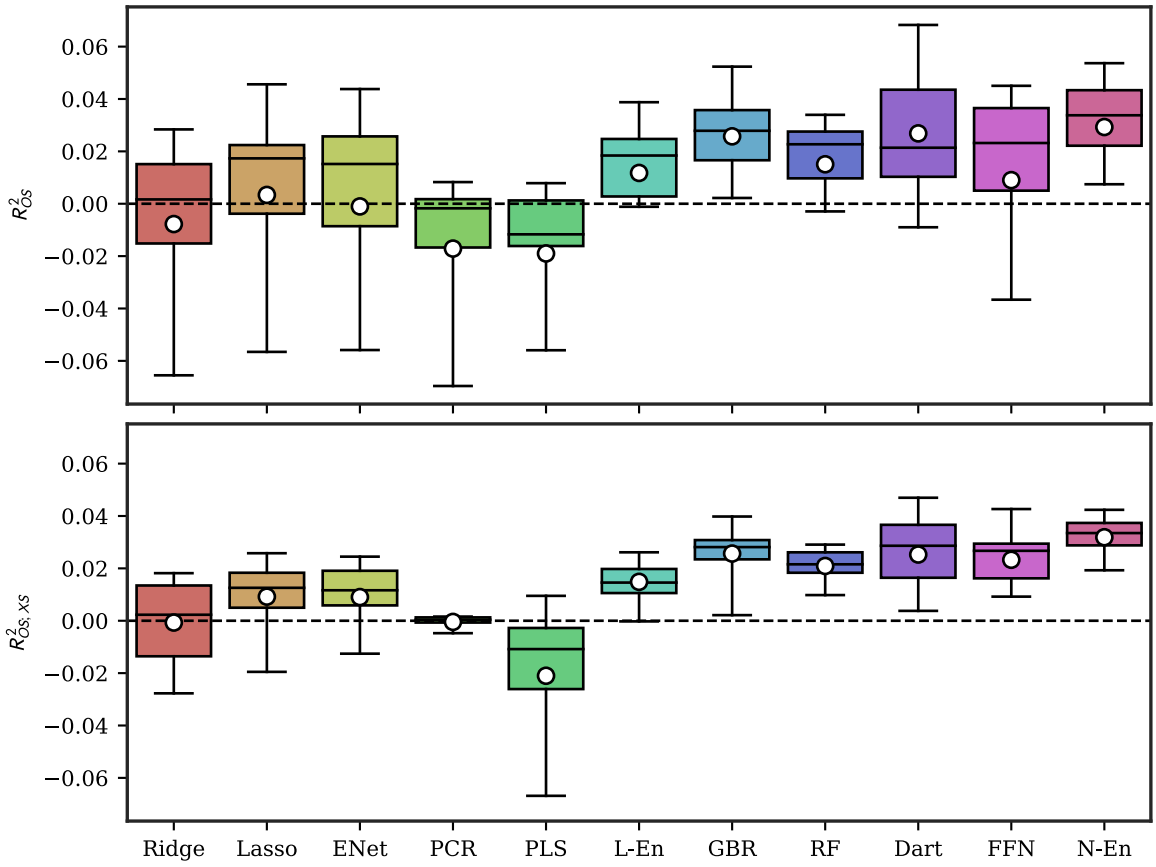


Fig. 3. Time-Series Dispersion of R^2_{OS} and $R^2_{OS;XS}$

The figure shows the dispersion of annual out-of-sample R^2 defined in Equation (4) in the upper panel and cross-sectional out-of-sample R^2 defined in Equation (5) in the lower panel. We show the 5th and 95th percentile R^2 in the whiskers, the interquartile range in the boxes, as well as the mean (circles) and median (bar).

around the median, or more exposed to the downside. A notable exception from this is Dart, which manages to produce the best predictability most consistently, with an interquartile range of 1% to 4.4%. Leveraging the informational content uncovered by Dart may thus grant large benefits to investors. Following this intuition, we find that ensemble models produce more stable forecasts. The minimum to maximum predictability for L-En ranges from −8% to +3.9% and for N-En between −2.7% and +5.7%. The forecasts of N-En are the most stable over time, producing significant predictive power for all years in the testing sample. Cross-sectional predictability is much less disperse (lower panel of Figure 3) and is generally increasing in the model complexity and the ability to model nonlinear interaction effects between characteristics. N-En is once more

the most stable model with the highest median and mean level of predictability over time.

Forecast comparison. We now turn to Diebold and Mariano (1995) tests to iteratively compare the forecasts of two competing models following Equation (8). Statistical significance at the 1% (5%) level is highlighted in light blue (blue). The first row of Table 2 shows that only L-En manages to statistically outperform the predictions made by Ridge regressions within the class of linear models. In comparison, all nonlinear models manage to beat the predictions by Ridge. Overall we find that the forecasts by L-En are significantly better than the forecasts any of the other linear models produce, highlighting the necessity to adequately pool the forecasts by multiple methods.

Panel A: Diebold and Mariano (1995) Forecast Comparison

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	1.76	0.27	−0.52	−0.56	**3.64**	**3.12**	**2.11**	**2.52**	**3.24**	**3.76**
Lasso		−1.24	−0.94	−0.98	**2.13**	**2.31**	1.42	1.93	1.85	**2.79**
ENet			−0.48	−0.51	1.94	**2.15**	1.43	1.84	**2.90**	**2.49**
PCR				−0.19	**2.60**	**7.17**	**7.03**	**5.49**	1.59	**8.08**
PLS					**2.61**	**7.78**	**8.02**	**5.71**	1.61	**8.64**
L-En						**2.09**	0.79	1.62	−0.42	**2.86**
GBR							**-4.40**	0.40	−1.60	1.04
RF								**2.52**	−0.72	**6.27**
Dart									−1.34	0.05
FFN										1.94

Panel B: Forecast Correlation

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	**0.84**	**0.83**	−0.10	0.53	**0.91**	0.51	0.44	0.50	**0.78**	0.64
Lasso		**0.95**	−0.14	0.51	**0.92**	0.45	0.41	0.44	**0.79**	0.59
ENet			−0.16	0.47	**0.90**	0.42	0.39	0.40	**0.79**	0.57
PCR				0.18	0.02	0.12	0.23	0.10	−0.06	0.10
PLS					**0.75**	0.52	0.53	0.51	0.57	0.60
L-En						0.56	0.53	0.55	**0.83**	0.70
GBR							**0.86**	**0.92**	0.61	**0.95**
RF								**0.79**	0.54	**0.88**
Dart									0.61	**0.94**
FFN										**0.79**

Table 2: Forecast Comparison

Panel A of the table shows Diebold and Mariano (1995) test statistics following Equation (8) for the nine models and two ensembles considered in the paper. A positive number indicates that the model in the column outperforms the row model. If it is highlighted in light blue (blue), this outperformance is statistically significant at the 1% level (5% level). Panel B shows forecast correlations defined in Equation (10). Here, highlighting in light blue (blue) denotes large values with a cutoff at 90% (70%).

Within the class of nonlinear models, only GBR manages to outperform L-En com-

fortably with a t-statistic of 2.09. Interestingly, the forecasts of GBR and Dart, which have produced the highest single-model predictability, are indistinguishable from one another. FFN manages to beat Ridge and ENet, but none of the nonlinear models.[14] The nonlinear ensemble N-En produces forecasts that beat any of the other models but Dart and GBR, which are its most vital inputs. Comparing the performance of L-En and N-En, we find that the forecasts of the latter are more informative with a t-statistic of 2.86. While the forecasts of GBR and Dart do not manage to statistically beat those of FFN, N-En manages to do so at the 10% level (t-stat = 1.94).

The forecast correlation analysis in Panel B of Table 2 confirms these insights. First, penalized linear regression methods yield similar predictions, which is expected given that Lasso and Ridge are special cases of ENet. Notably, L-En produces forecasts highly correlated to either of these methods ($\rho > 90\%$), as does FFN ($\rho > 70\%$), suggesting that the underlying process proposed by FFN is quite similar to a regularized linear function. Second, tree-based methods are a class of their own, showing correlations of $\rho > 70\%$ only among themselves. Given their unique setup of identifying quantile splits in the input characteristics to relate to option returns, this is not surprising. At the same time, these methods, especially boosted tree-based methods (GBR and Dart), outperform all other models. Consequently, predictions by N-En share many of their properties ($\rho > 90\%$). However, the ensemble predictions are also highly correlated with the two other nonlinear methods with $\rho^{\text{N-En,FFN}} = 79\%$ and $\rho^{\text{N-En,RF}} = 88\%$.

Impact of Nonlinearities. The results highlight the usefulness of forming forecast ensembles of many models. Therefore, from now on, we will compare the linear (L-En) and nonlinear (N-En) ensemble methods to understand the implications of modeling nonlinear interaction effects.

In this section, we highlight the usefulness of modeling nonlinear interaction effects between option characteristics. The left panel of Figure 4 compares the monthly R^2_{OS} for N-En in dark blue and for L-En in light blue. The figure provides multiple insights: First, both ensembles tend to beat a naive benchmark of predicting zero delta-hedged returns in most months. Second, N-En is less prone to experience predictability crashes. While L-En does a poor job of predicting future option returns during and after turbulent times (2008-2012 for example), N-En has predicted returns much more consistently. Third, for

[14]The Diebold and Mariano (1995) test provides a statistical measure of the differences in the total forecast errors. Comparing cross-sectional forecast errors in Section F.1 in the Online Appendix shows that FFN manages to outperform all linear models, as well as the linear ensemble, in this setting. Furthermore, all nonlinear models manage to beat L-En in uncovering cross-sectional dispersion in option returns.

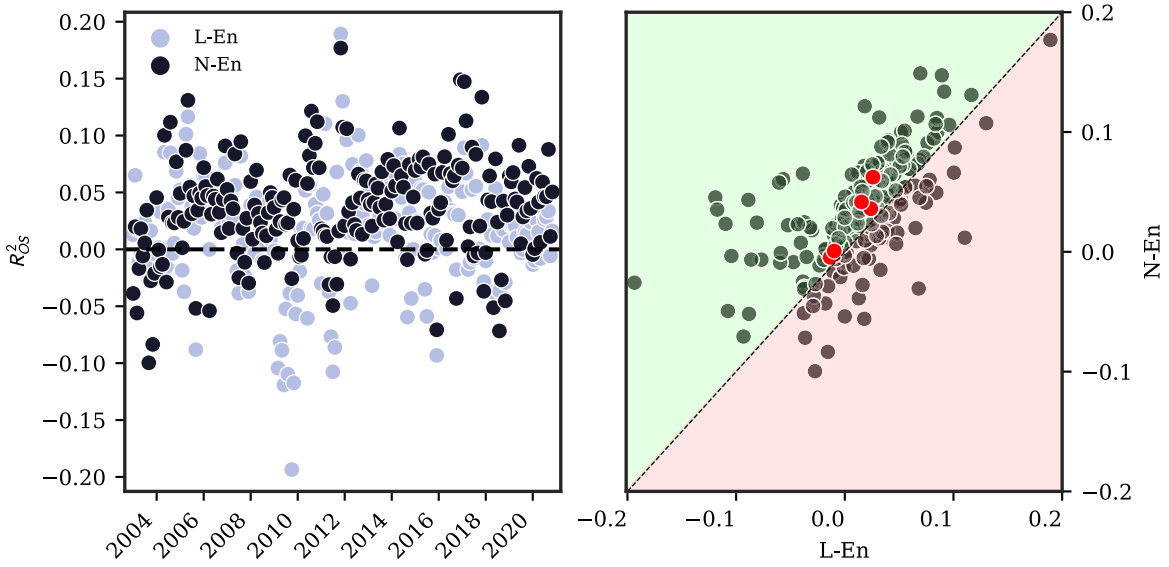


Fig. 4. Comparing Linear and Nonlinear Ensembles

The left panel of the figure shows monthly R^2_{OS} for the testing sample from 2003 through 2020 for the linear (L-En) and nonlinear (N-En) ensembles. The right panel compares the two by showing the resulting R^2_{OS} for L-En on the x-axis and for N-En and the y axis. The green-shaded area represents a relative outperformance in terms of predictability for N-En, while the red-shaded area represents the opposite. The five red circles represent the Coronavirus selloff from December 2019 through April 2020.

N-En, the predictive power has stayed roughly constant over time, confirming not only that nonlinear models leverage the information embedded in option characteristics, but also that the resulting predictability patterns are highly persistent over time.[15] At the same time, this suggests that our methods pick up more than just plain mispricing, especially in recent years, under the assumption that modern options markets have become informationally more efficient.

The panel on the right shows a scatter plot of the resulting R^2_{OS} of L-En on the x-axis and N-En on the y-axis. The green-shaded area represents a relative outperformance of the nonlinear ensemble, whereas the red-shaded area represents the opposite. For 71.2% of the months in our sample, we find that N-En outperforms L-En and if so, quite comfortably. In the figure, we have also highlighted in red circles the period of December 2019 through April 2020, which are the months surrounding the Coronavirus selloff in February and March 2020. Nonlinear ensembles were better able to deal with this huge exogenous shock, beating L-En in all of the five months. The R^2_{OS} for N-En dips slightly

[15]We keep increasing the training period by one year each time we roll forward, such that no historical information is ever discarded.

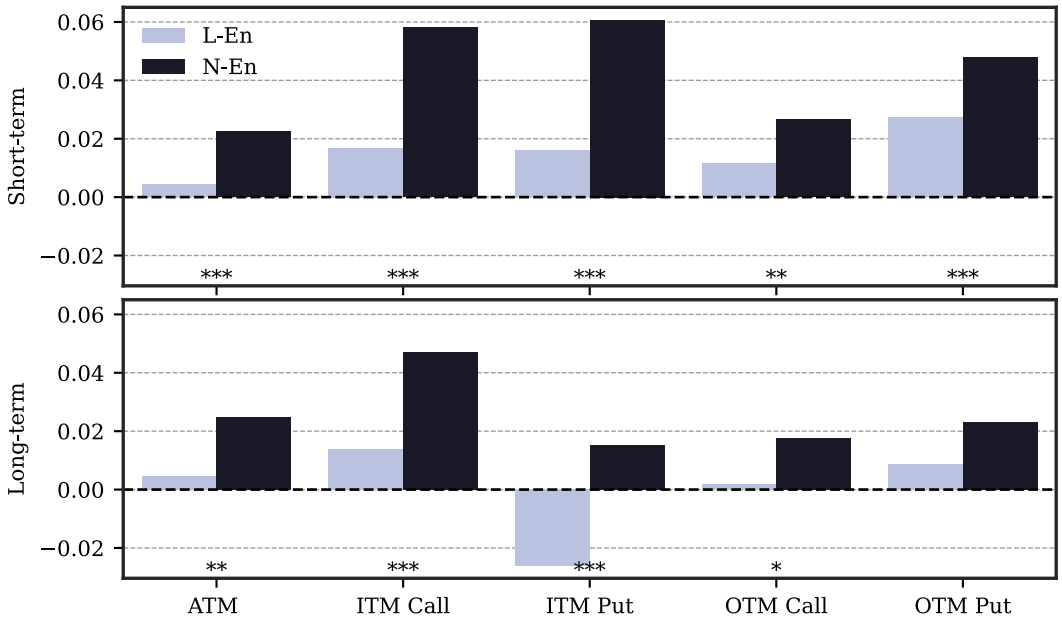


Fig. 5. Predictive Power and Impact of Nonlinear Models for Option Buckets

The figure shows monthly R^2_{OS} for the testing sample from 2003 through 2020 for the linear (L-En) and nonlinear (N-En) ensembles using options that fall in different moneyness and maturity buckets, as defined in Section 4. ***, **, * below the bars denotes a statistically significant outperformance of N-En at the 0.1%, 1% and 5% level as defined in Equation (8).

below zero only in January and reaches pre-crisis levels of 3 to 6% in March. While the pandemic-driven selloff constituted a large exogenous shock to financial markets, the relationship between option characteristics and future returns quickly went back to normalcy after the initial reaction. This speaks in favor of using nonlinear machine learning methods to identify persistent predictability patterns in the options market, especially given that the sample we used to train the models to uncover predictability patterns in option returns during the Coronavirus selloff ended in December of 2017. In Figure F.1 we repeat the exercise using the cross-sectional out-of-sample $R^2_{OS;XS}$ defined in Equation (5). N-En beats the linear ensemble in 85.6% of the months and predictability does not dip below 1.5% of variation of return spreads explained during the Coronavirus selloff.

Predictions for Option Buckets. A possible objection to our result that nonlinear models outperform linear models is our intent to predict *all* options across the moneyness-maturity spectrum. This potentially introduces nonlinear interactions mechanically that would be irrelevant when predicting returns of for example only at-the-money and short-term contracts.

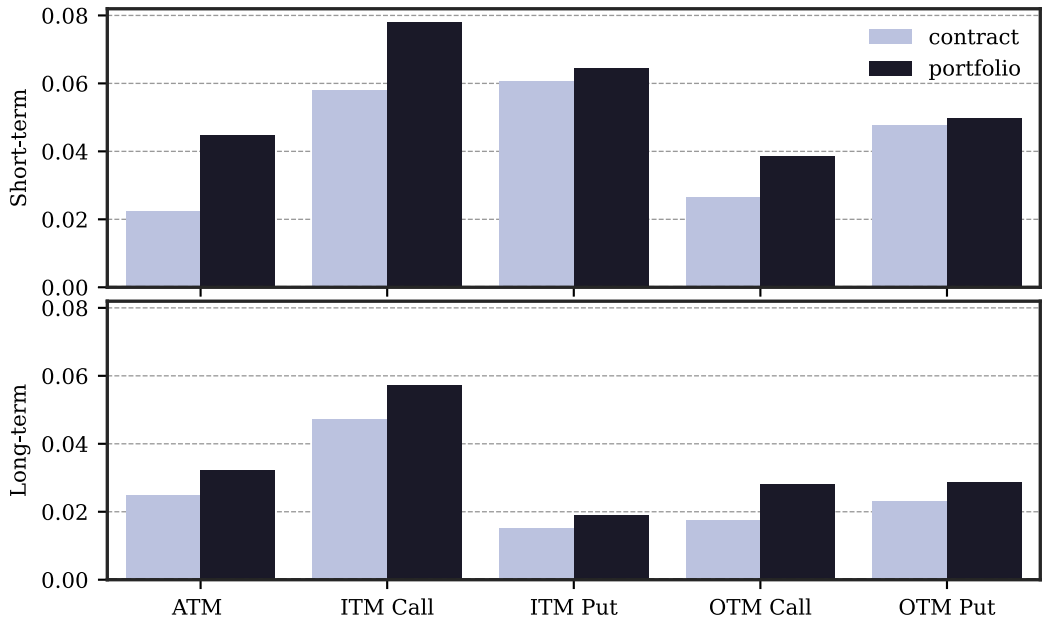


Fig. 6. Predictive Power and Impact of Nonlinear Models for Option Portfolios

The figure shows monthly R^2_{OS} for the testing sample from 2003 through 2020 for the nonlinear (N-En) ensembles using option portfolios. Specifically, we form dollar open interest-weighted portfolios using all options of a given underlying that fall in a respective bucket (defined in Section 4) and predict future returns using N-En. This approach closely mimics recent studies on option return predictability (Cao and Han, 2013; Cao et al., 2021; Goyenko and Zhang, 2021). We compare the resulting predictability levels with those obtained when predicting the individual contracts included in each bucket. Figure F.3 repeats this exercise using the cross-sectional out-of-sample $R^2_{OS;XS}$.

To address this potential concern, Figure 5 shows the R^2_{OS} comparison between the linear and nonlinear ensemble for each option bucket considered. Across all specifications, we find that N-En produces higher levels of predictability than L-En. In all cases for short-term and most for long-term options, this outperformance is statistically significant. Nonlinearities also play a vital role to predict option returns within a certain option bucket.[16] Deterministic differences between the maturity and moneyness are limited within these buckets. Adding the ability to model nonlinear interactions between characteristics is essential when predicting option returns, regardless of the contract's moneyness and maturity.

Additional evidence is provided in Figure 6. It shows the predictive power of N-En for predicting the returns of option portfolios and compares it with predicting returns for the individual contracts within that portfolio. The construction follows the bucket defi-

[16]Repeating the exercise using the $R^2_{OS;XS}$ yields similar results, see Figure F.2 in the Online Appendix.

nition of Section 4 and creates one option portfolio per bucket and underlying, weighting each contract by its dollar open interest. The literature commonly uses portfolios or a single most-liquid (call) contract to assess the predictability of option returns to limit the influence of noise and drastically shrink the size of the estimation problem at the loss of generalizability (see Cao and Han, 2013; Cao et al., 2021; Goyenko and Zhang, 2021). Instead, we seek to provide the best possible explanation for future returns of *all* options on U.S. equities. We can make three observations from this exercise: first, the predictive power for portfolios is higher for all buckets. Note that the models are trained on individual contracts, such that this increase in predictability reflects the benefit of a lower dispersion of portfolio returns, and a smaller impact of noise. Second, the difference in predictability is quite small for some buckets, and especially so for out-of-the money and long-term options. Third, especially for short-term at-the-money options, the benefit of forming portfolios is large, doubling R^2_{OS}. These numbers exceed the predictability for short-term at-the-money option portfolios previously found by the literature.[17]

As a last bit of evidence, Online Appendix G provides a full-on comparison between N-En estimated for all options in our sample and only for short-term at-the-money options, which have enjoyed the largest academic interest. We generally find comparable levels of overall predictability and little drawbacks of predicting returns of all options simultaneously.

5.2. Machine Learning Portfolios

To gauge if the predictability of machine learning methods is also economically significant, we follow Gu et al. (2020) and form portfolios using machine learning forecasts. Specifically, each month, we sort individual equity options into 10 decile portfolios based on the machine learning models' (L-En and N-En) expected return forecasts. Then, we calculate the one-month-ahead average realized return of individual equity options in each decile. Finally, we compute the average long-short portfolio return of a zero-net investment portfolio by buying options with the highest expected return forecast (decile 10) and financing this investment by writing options with the lowest expected return forecast (decile 1). Each contract is weighted by its dollar open interest at the time of investment.

Table 3 shows the average predicted and one-month-ahead realized portfolio returns.

[17] If we form a dollar open interest-weighted portfolio for all options of a given underlying, the resulting R^2_{OS} ($R^2_{OS;XS}$) amounts to 4.59% (5.33%).

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.308	−0.967	1.335	−0.725	−1.612	−1.556	1.845	−0.843	***
2	−0.722	−0.482	1.450	−0.332	−0.709	−0.637	1.564	−0.408	
3	−0.491	−0.319	1.434	−0.222	−0.397	−0.393	1.393	−0.282	
4	−0.321	−0.220	1.432	−0.153	−0.223	−0.220	1.272	−0.173	
5	−0.178	−0.154	1.520	−0.101	−0.098	−0.125	1.260	−0.099	
6	−0.047	−0.096	1.488	−0.065	0.006	−0.077	1.321	−0.059	
7	0.082	−0.026	1.485	−0.017	0.109	−0.038	1.444	−0.027	
8	0.219	0.027	1.475	0.018	0.228	0.042	1.494	0.028	
9	0.385	0.084	1.450	0.058	0.393	0.135	1.607	0.084	
Hi	0.695	0.314	1.580	0.199	0.854	0.460	1.852	0.248	
H-L	2.004	1.281	1.240	1.033	2.466	2.016	1.466	1.375	***
		(11.52)		(8.64)		(14.28)		(9.57)	

Table 3: Trading on Machine Learning Predictions

The table shows the returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods. Each contract is weighted by its dollar open interest at the time of investment. Pred denotes the average predicted return within the respective portfolio, Avg the average realized return, SD the standard deviation of realized returns and finally SR the realized Sharpe ratio. All values are given per month. The last column (N vs. L) gives the significance of comparing the mean realized returns for N-En and L-En. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

For both ensemble classes, the average predictions are slightly larger than the returns they actually realize, but the predicted and realized return spreads between the lowest and highest predicted return portfolio is much greater for N-En. The per-month realized high-minus-low return generated by N-En for the testing sample of 2003 through 2020 is 2.02% with a Sharpe ratio of 1.38. The realized payoff is also close to the predictions made by N-En (2.47% per month, on average). To validate the usefulness of allowing for nonlinearities when predicting future option returns, we perform a comparison of realized return spreads in the last column of Table 3, which highlights that N-En is better than L-En at identifying particularly overpriced options. This translates into the large H-L spread between the two models based on the difference-in-differences (diff-in-diff) test results; i.e., the diff-in-diff return spread generated by N-En vs. L-En is economically and statistically significant at 0.74% per month, showing that the cross-sectional variation in option returns is more accurately captured by the nonlinear models.[18]

Portfolio Decomposition. Table 4 provides summary statistics for the decile portfolios based on the nonlinear ensemble N-En. The high and low portfolio tend to include

[18]We have repeated this exercise using equal weights for each contract. The results are in Table H.1 in the Online Appendix and confirm the findings discussed here.

	Lo	2	3	4	5	6	7	8	9	Hi
N^U	634.65	797.43	820.74	808.51	794.23	784.62	769.00	738.25	672.45	519.95
m	1.03	1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.05
ttm	124.66	156.54	172.37	178.87	182.27	184.51	184.72	185.19	187.54	205.52
$\hat{r} > 0$	0.31	0.35	0.38	0.39	0.41	0.42	0.44	0.45	0.47	0.52
$s(r) = s(\hat{r})$	0.69	0.65	0.62	0.58	0.55	0.53	0.48	0.47	0.48	0.52
% Calls	0.43	0.42	0.44	0.46	0.47	0.51	0.55	0.60	0.65	0.75
Spread	0.20	0.16	0.13	0.12	0.11	0.11	0.11	0.12	0.13	0.15
OI	0.06	0.08	0.10	0.12	0.12	0.12	0.11	0.11	0.09	0.08
delta	0.02	0.02	0.03	0.03	0.03	0.05	0.08	0.13	0.17	0.27
gamma	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
vega	0.15	0.16	0.18	0.19	0.20	0.20	0.20	0.20	0.20	0.21
theta	−0.31	−0.19	−0.15	−0.13	−0.12	−0.11	−0.11	−0.11	−0.12	−0.12

Table 4: N-En Portfolio Decomposition

The table shows summary statistics for the ten machine learning portfolios following the nonlinear ensemble N-En. All measures are averaged over time. N^U denotes the number of individual stocks underlying the options in the portfolios, m the moneyness, ttm the time-to-maturity. $\hat{r} > 0$ denotes the share of positive returns in the portfolio and $s(r) = s(\hat{r})$ the share for which N-En correctly predicted the realized return's sign. % Calls is the share of call options in the portfolio, Spread the average option bid-ask spread, OI the relative open interest and *delta*, *gamma*, *vega*, and *theta* the respective option Greeks. *gamma* is expressed for a 1% move in the underlying stock ($gamma \times \frac{S}{100}$) and *vega* and *theta* in terms of the underlying stock price ($\frac{x}{S}$ for $x \in [vega, theta]$).

options from $N^U = 520$ and $N^U = 634$ underlyings, respectively. The portfolios use options from the least number of individual stocks, but still rely on average on options from about a third of the stocks included in the sample. While we find little change in the average moneyness of the included options, the Low portfolio includes more short-term and fewer call options (% Calls = 0.43), while the High portfolio includes more long-term call options (% Calls = 0.75). The relative bid ask spreads of the options are also highest for the extreme portfolios, while the outstanding open interest is lowest for options in these portfolios. The *delta* (of the unhedged option), *theta* and *vega* are increasing in the expected option return, while *gamma* is roughly comparable across portfolios. Focusing on whether the predicted return direction is correct, $s(r) = s(\hat{r})$, we find the highest accuracy for Low with 69% of the return directions correctly predicted. The ratio first decreases and then starts to increase once more for portfolios 9 and 10. For portfolio 10 (High), it stands at 52%. Figure H.1 to Figure H.3 in the Online Appendix provide a visual representation of the portfolio decomposition, split by the put and call contracts included.

Figure H.4 shows the relative share of call options in the machine learning portfolios for each year in our testing sample. We find that it increased after the financial crisis.

Over time, the call share of the High portfolio has decreased, especially after the financial crisis. For the Low portfolio, it has stayed roughly constant over time. Another interesting aspect of the machine learning portfolios is to analyze in how far N-En places all options trading on a given stock in the same bucket, or whether it distinguishes between over- and under-priced contracts *within* a single underlying. For this, Figure H.5 in the Online Appendix shows the share of options of a given underlying that ended up in the same portfolio. We find that this measure is highest for the High and Low portfolios, but still regularly below 50%. This shows that N-En does not simply identify underlying stocks for which all options are considered mispriced, but also picks up on the variation *within* the option chain of a given stock.

Prediction Persistence & Turnover. How likely is it that securities selected in portfolio i at month t remain in the same portfolio at month $t+1$? While we cannot answer this question for individual option contracts, given that their moneyness and time-to-maturity change rapidly, we can provide indicative evidence for this on the options-bucket level. We focus on how the portfolio mode for all stock-bucket combinations in our sample changes from one month to another.[19] With this, we can understand more about the persistence of the machine learning predictions. Figure H.6 in the Online Appendix provides the results to this exercise: we first note that the diagonal is the lightest-shaded area, highlighting that moving from one portfolio to the same or a neighboring is the most likely transition. Second, the lightest areas overall are at the High and especially the Low portfolio. If options of a given stock-bucket combination are classified as being highly overpriced (Low portfolio), they tend to stay in this portfolio in the next months. The same applies to a lesser degree to the most underpriced stocks (High portfolio).

Subsamples. In a next step, we investigate in how far the profitability of our machine learning portfolios changes with the state of the economy, and whether N-En manages to beat L-En regardless of this state. The results in Table H.2 in the Online Appendix shows that realized returns are generally amplified during bad states of the economy and that N-En manages to beat L-En across different market phases. This holds for a wide range of measures approximating the state of the economy, market volatility, uncertainty, and investor sentiment.

Option Buckets. Table H.3 in the Online Appendix provides the H-L results for different option buckets. Predictability is concentrated in short-term options, for which we also find the highest payoffs when following the investments proposed by our ma-

[19]That is, each stock-bucket combination is assigned the portfolio that most of the options in that combination fall into. Results remain intact when we consider the average portfolio instead.

chine learning models. Throughout all option buckets but short-term otm calls we find significantly higher raw (Avg) and risk-adjusted (SR) returns for N-En than for L-En.

Risk Attribution. A possible explanation for these results is that the machine learning models are best at predicting the most risky option positions, which should translate to higher future realized returns. To understand whether this is the case, we compute risk-adjusted returns for the H-L portfolios for N-En, either for the pooled sample of all options, or split by option buckets. We consider a wide range of candidate models, which have been proposed by earlier studies to explain the returns of a variety of financial instruments. To date, we are still lacking a concise factor model for option returns. Therefore, we resort to the most prominent models from other asset classes including the CAPM, the Fama and French (2015) five factor model enhanced by the momentum factor of Carhart (1997) (FF6, Fama and French, 2018) and additionally by the liquidity factor of Pástor and Stambaugh (2003) (FF6+PS), a model using the Agarwal and Naik (2004) option-market factors (AN), a model with the intermediary leverage bearing capacity of Grünthaler, Lorenz, and Meyerhof (2020), and finally the factor model on optionable stocks proposed by Bali and Murray (2020) (BM).

The results are provided in Table H.4 in the Online Appendix. Raw and risk-adjusted returns for all option types are virtually the same. The risk exposure picked up by the candidate models does not suffice to explain the return spreads generated by the nonlinear machine learning methods.

Impact of Transaction Costs. So far we have assumed that the investor can buy and sell each option at the mid-point between bid and ask – that is with zero effective spreads. Prior research has shown that transaction costs in option markets can be large (Ofek et al., 2004). We now turn to the impact of trading at different transaction prices by changing the ratio of effective to quoted spreads. For that, Table 5 shows the full-sample payoffs to the high-minus-low trading strategy assuming that investors have to pay an effective spread of 50% or 75% of the quoted bid-ask spread. As a simple example, with an effective spread of 50%, we would buy an option with $bid = \$10; ask = \$15$ at \$11.25 and sell it at \$13.75. We consider H-L payoffs, such that we go long options in decile 10 and short options in decile 1. For options that expire after our holding period of one month, we consider the effective spreads for trading in and out of the position; for options that expire within the holding period, we only do so for engaging into the position and cash in the payoff at maturity. We run these analyses with short-term options to be consistent with previous studies in the option-return literature (e.g., Cao and Han (2013) and Cao et al. (2019)). Moreover, short-term options are demanded more by investors

			Eff. Spread = 50			Eff. Spread = 75		
			H-L	t	SR	H-L	t	SR
Pooled Sample			0.706	(4.98)	0.484	0.075	(0.47)	0.049
Different Moneyness Levels								
$\tau \leq 90$	atm		0.835	(7.10)	0.565	0.354	(2.89)	0.236
	itm	C	−0.616	(−2.31)	−0.239	−1.772	(−5.03)	−0.601
	itm	P	0.587	(3.35)	0.336	−0.100	(−0.54)	−0.056
	otm	C	0.655	(1.83)	0.153	−0.744	(−1.94)	−0.171
	otm	P	2.011	(6.54)	0.543	1.069	(3.38)	0.288

Table 5: Machine Learning Predictions and Effective Spreads

The table shows the realized returns of the high-minus-low portfolio following the predictions by N-En using all short-term options after accounting for transaction costs through effective spreads, which we define as a fraction of the quoted spreads provided OptionMetrics. H-L denotes the average returns controlling for spreads of 50% and 75%, t denotes the corresponding t-statistic and SR the resulting Sharpe ratio. We also provide the information for short-term option buckets as defined in Section 4.

(Garleanu et al., 2009) and generally enjoy tighter spreads.

As presented in Table 5, the strategy using all available short-term option contracts remains highly profitable for 50% effective spreads. The return is indistinguishable from zero for spreads of 75%. The average monthly returns are 0.706% or 0.075% with Sharpe ratios of 0.484 and −0.05. Breaking this predictability down by the moneyness of the contracts, we find that at-the-money options, as well as out-of-the-money puts manage to produce highly significant return spreads of 0.354% and 1.069% per month even for large effective spreads of 75%.

In Figure 7 we furthermore impose a more realistic trading behavior of our investor using the best-performing option bucket of short-term ATM options: before entering a trade, she considers the current level of transaction costs relative to the trade's expected profits. Specifically, we assume that the investor excludes options with the x% largest bid-ask spreads at trade initiation and bases her investment decisions only on the remaining sample. We consider $x \in [25, 50, 75, 100]$%. For low to moderate levels of transaction costs we find it to be worthwhile to consider the entire sample of short-term ATM options. In contrast, at around 75% effective spreads we find a relative outperformance of restricting her attention to those options with the lowest spreads at trade initiation. While all strategies generate significantly positive returns for 75% effective spreads, we find that the sample restricted to the 25% most-liquid options still manages to generate a positive return spread significant at the 5% level under the assumption of paying the full end-of-day bid-ask spread quoted by OptionMetrics.

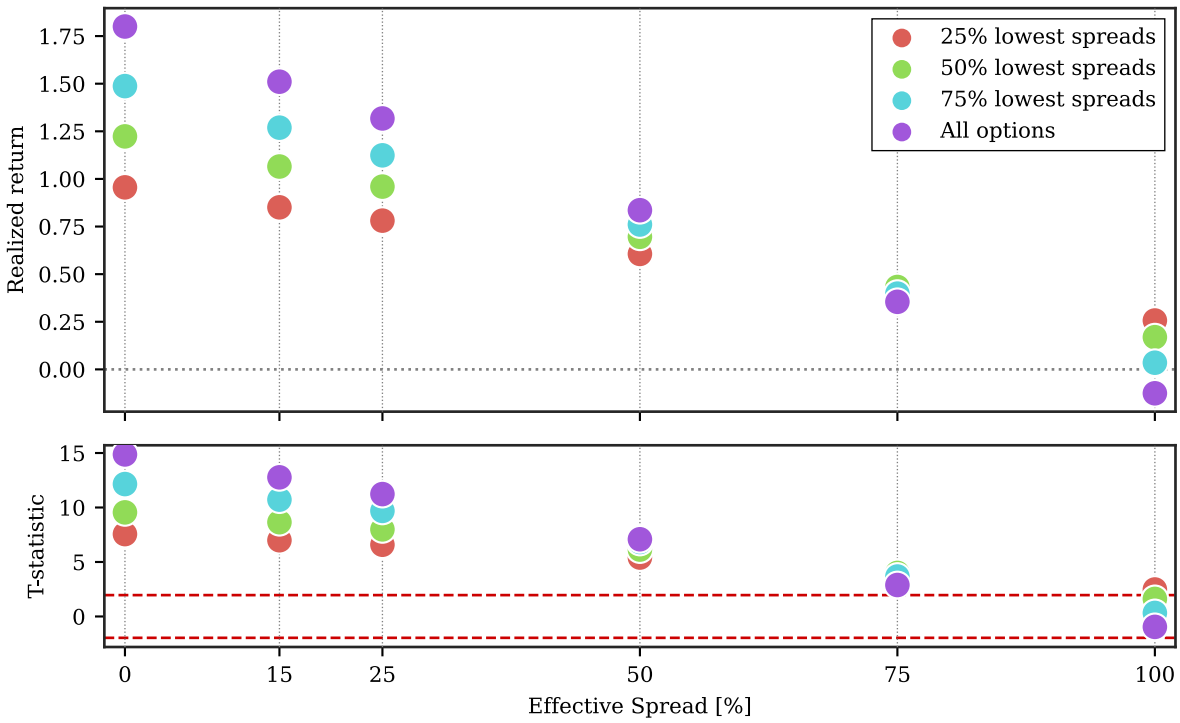


Fig. 7. Impact of Spreads on Profitability Conditional the Option Universe

The figure shows the profitability of trading on the predictions of the nonlinear ensemble N-En using only short-term at-the-money options, conditional on an assumed level of effective spreads the investor has to be pay for each transaction and a restriction of the considered option sample using the level of option bid-ask spreads at trade initiation. The upper panel shows the resulting long-short returns, the lower panel the corresponding Newey and West (1987) t-statistics with a lag of twelve months.

Muravyev and Pearson (2020) argue that effective spreads are much lower than the ones we measure here using OptionMetrics' end-of-day quoted spreads, since investors have the ability to time trades over the entire trading day. Using high-frequency data on single-stock options, the authors find that effective transaction costs are about a quarter smaller than their conventional estimates. Moreover, considering the impact of execution timing, the reduction can be increased to more than 60%. We are therefore confident that the signals N-En generates are valuable and require further discussion on the sources of these highly profitable trading strategies.

5.3. Which Characteristics Matter?

We next analyze the importance of option characteristics and groups built thereof. Optimally, we would re-estimate the model after excluding the characteristics in each

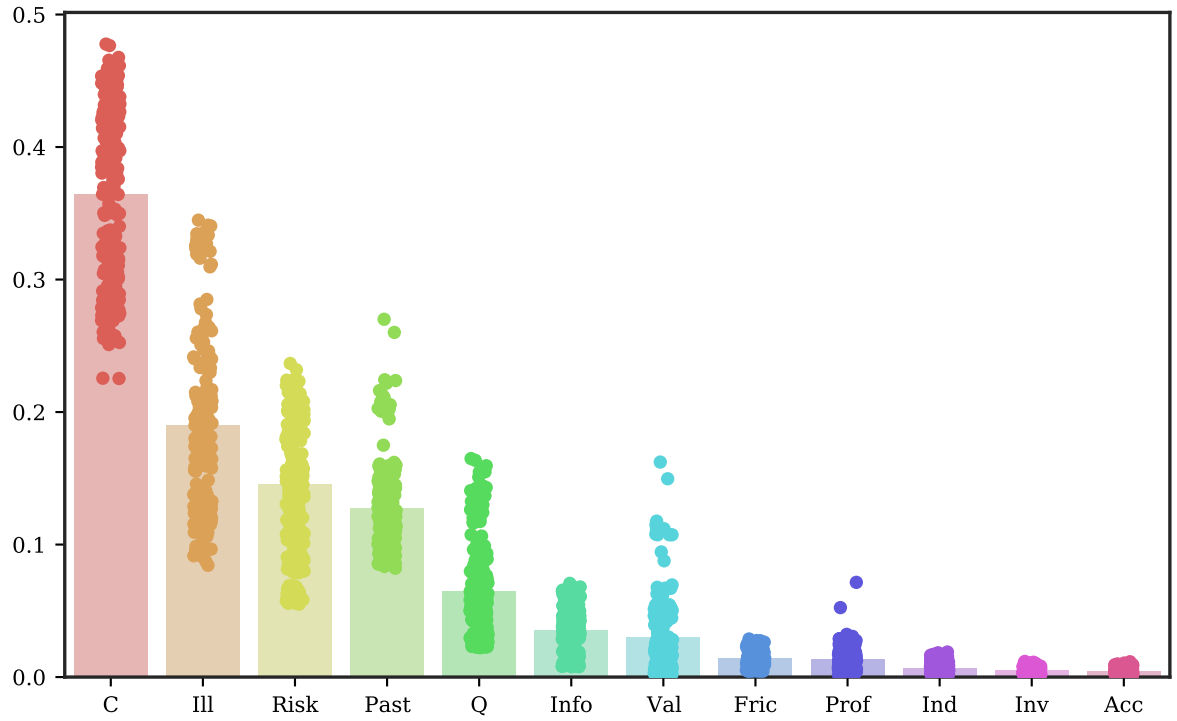


Fig. 8. Feature Group Importance for Nonlinear Forecast Ensemble

The figure shows the feature group importance for the twelve feature groups defined in Section 4.3 for the nonlinear (N-En) ensemble. We measure the importance using SHAP values following Lundberg and Lee (2017). The group importance is the sum of the resulting SHAP values for all features included in a given group. The values are scaled such that they sum to one. The bars represent the mean feature group importance for the entire testing sample, the dots the dispersion of the group importance for the months in the testing sample. The abbreviations used: Acc=Accruals, Prof=Profitability, Q=Quality, Inv=Investment, Ill=Illiquidity, Info=Informed Trading, Val=Value, C=Contract, Past=Past Prices, Fric=Frictions, Ind=Industry.

group sequentially. This approach, however, is infeasible, given the large number of characteristics (a total of 270) in our estimation and hence, the large computational burden required.[20] Instead, we adopt the approach of Lundberg and Lee (2017) who propose the use of SHAPs (SHapley Additive exPlanations), which approximate the effect of this feature exclusion and are based on cooperative game theory.

Characteristics Groups. The relative feature group importance for N-En is provided in Figure 8, in which the groups are sorted by their total importance over the entire testing sample. Contract-based characteristics are the most important predictors of future option returns. Knowing where an option lies on the underlying's implied volatility

[20]We use this approach in Section 5.4 and re-estimate each model for three subsamples of the input characteristics.

surface and knowing where that implied volatility surface lies relative to the market is essential when making option return forecasts. Measures about illiquidity and risk are the next-most important predictors. Of secondary importance, but nevertheless aiding in the prediction process for N-En are characteristics in the groups Past Prices, which includes measures of stock and option momentum.[21] The dots in the figure represent the group importance during each month in the testing sample. Occasionally, we find that illiquidity-based features are the most important. For most months, however, contract-based information exerts the highest influence on N-En's predictions. Finally, N-En relies mostly on information drawn from options to make its predictions. The impact of most groups comprised solely of stock characteristics is very small in comparison.

Single Characteristics. Figure 9 shows the ten most important characteristics, along with the dispersion of their relative importance across the testing sample. The most influential characteristic by far is the implied volatility of the option (iv), followed by the bid-ask spread of the underlying stock (baspread). Industry momentum (indmom), the option's theta (theta), the stock's realized variance (rv), and the variance risk premium (ivrv) are also highly important. The reliance also on stock-based characteristics corroborates the findings by Cao et al. (2019).[22]

Feature Sensitivity. Next, we seek to understand how the most important features affect the predictions made by N-En. For this, Figure 10 shows a bee-swarm plot highlighting the change in the return prediction due to a given characteristic. For example, a higher implied volatility leads to a more negative prediction of delta-hedged returns. Negative relationships are also present for the difference between implied volatility and realized volatility and the volatility of liquidity on the stock level (std_dolvol). In contrast, higher values of the underlying's bid-ask spread or industry momentum positively affect return predictions. Figure I.4 in the Online Appendix shows that the above relationships are stable over time in their directional assessment. However, the magnitude to which changes in individual characteristics impact the return predictions varies substantially over time. Moreover, Figure I.5 in the Online Appendix provides further insights on the functional from of the impact of single features on N-En's predicted delta-hedged return. The impact is highly non-linear for all of the ten most important characteristics.

Sensitivity to Volatility and Jump Risk. A large part of the literature man-

[21]Figure I.1 and Figure I.2 in the Online Appendix provide evidence that the above findings are robust if focusing on single option buckets and on single years in our sample, respectively.

[22]Figure I.3 in the Online Appendix shows that these findings are robust to focusing on single option buckets. There are only few occasions where features not ranked among the ten most important characteristics of the full sample are entering the top ten of a certain option bucket.

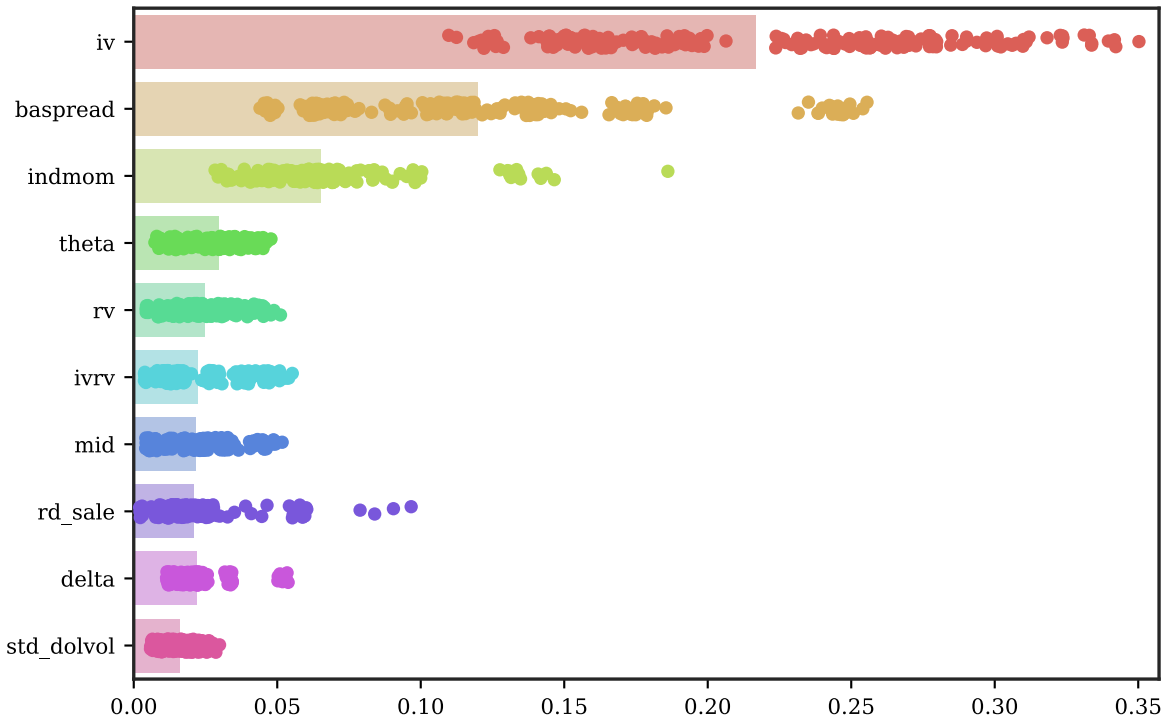


Fig. 9. Most Important Features for N-En

The figure shows the ten most influential features for the predictions of the nonlinear ensemble (N-En) following the importance using SHAP values (Lundberg and Lee, 2017). The values are scaled such that they sum to one across all 270 characteristics. The bars represent the mean feature importance for the entire testing sample, the dots the dispersion of the importance for the months in the testing sample.

ifests that options incorporate volatility and jump risk premia. Cremers, Halling, and Weinbaum (2015) and Dew-Becker, Giglio, and Kelly (2021) show that some options are driven by volatility risk, while others are driven by jump risks, conditional on their time-to-maturity and moneyness. Following the previous literature, we identify four (five) proxies for volatility (jump) risks in our set of characteristics.[23] We compute the importance of the nine characteristics in terms of their ranking using absolute SHAP values for the entire sample, as well as the ranking for individual option buckets, relative to the rank of the full sample. Table I.1 in the Online Appendix reports these results. In general we find that jump risk proxies impact short-term and especially out-of-the-money

[23]We proxy volatility risk with the volatility of implied at-the-money volatility (ivvol, as in Baltussen, van Bekkum, and van der Grient, 2018), volatility uncertainty as defined in Cao et al. (2019), the option's vega following Cremers et al. (2015) and ivrv as in Goyal and Saretto (2009). For tail risk we use option implied tail risk (tlm30, as defined in Vilkov and Xiao, 2012), the difference in implied volatility between out-of-the-money put and at-the-money call options (skewiv as defined in Xing, Zhang, and Zhao, 2010), risk-neutral skewness (rns30) and kurtosis (rnk30), as well as the option's gamma.

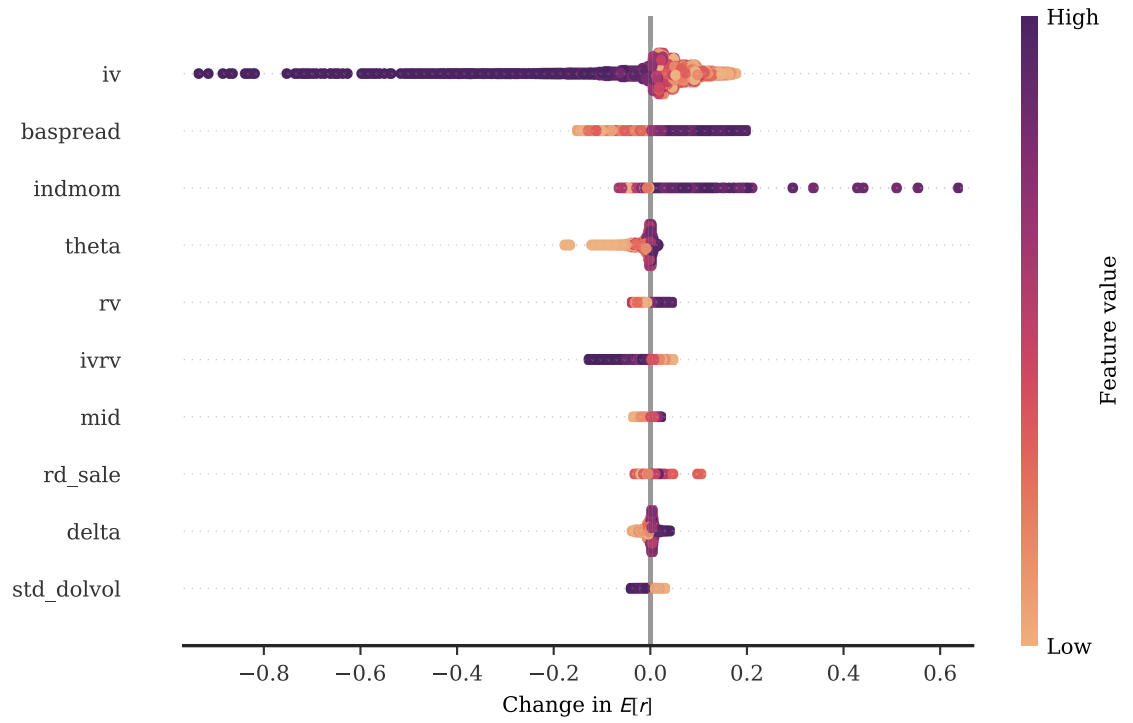


Fig. 10. Impact of the Most Important Characteristics on Predicted Delta-Hedged Returns for N-En

The figure shows the impact of the ten most influential characteristics on the predictions of the nonlinear ensemble (N-En). Feature impact is measured by SHAP values following Lundberg and Lee (2017).

options the most. In contrast, the impact of many volatility risk proxies is comparable across buckets, with an increased impact of vega for short-term options. Knowing about a stock's variance risk premium (ivrv) is vital to predict all options. Note that this proxy is also part of the ten most important features that N-En identifies.

To provide an intuition for how options with a different moneyness-maturity combination depend on volatility and jump risks, Figure 11 visualizes the functional form of how sensitive N-En's predictions are to changes in the most important volatility (ivrv) and jump risk proxy (gamma). The predicted delta-hedged returns are uniformly decreasing for higher values of ivrv. Interestingly, we find a strong tail dependence in this pattern: For the largest values of ivrv, we find an amplified impact on the return prediction, depressing predicted returns by more than -4% for the stocks with the five percent highest ivrv. We find little variation in this functional form across option buckets.

In contrast, the relationship between predicted delta-hedged returns and tail risk shows more variation across different buckets. We generally find that low gamma values

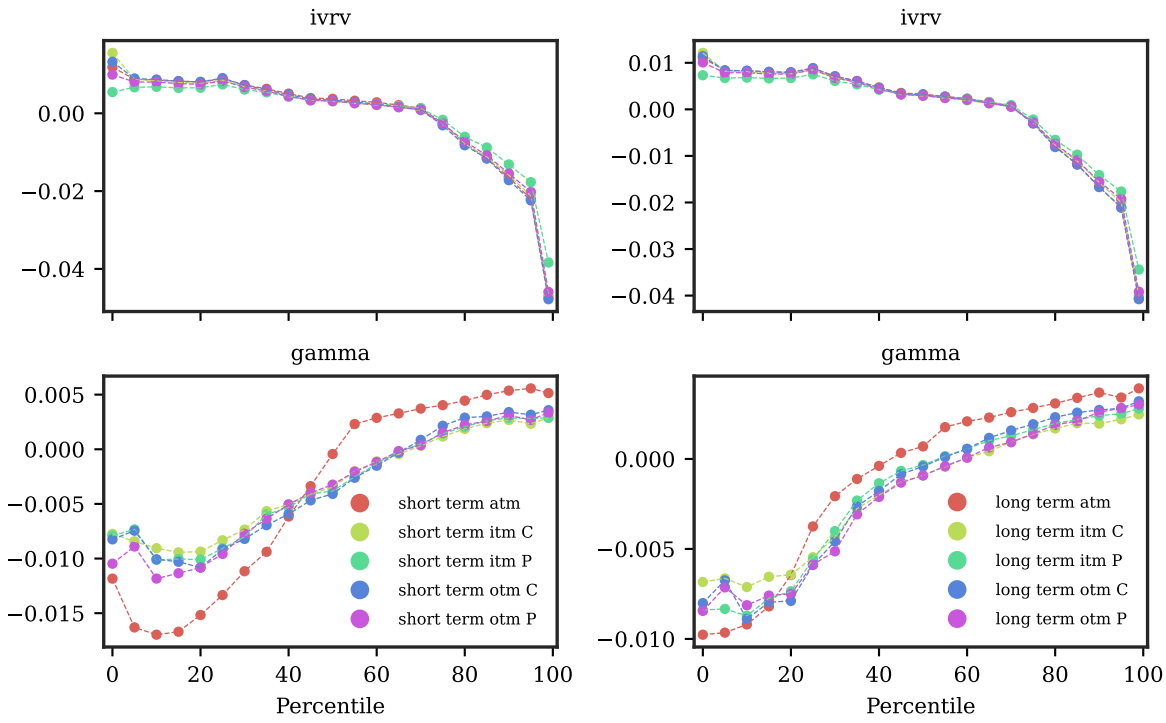


Fig. 11. Relative Impact of Volatility and Jump Risk Premia Per Bucket

The figure shows the impact of features proxying for volatility and jump risk, respectively, on the predictions of the nonlinear ensemble (N-En) per bucket. The impact of the feature is measured using SHAP values following Lundberg and Lee (2017). ivrv is the volatility risk premium following Goyal and Saretto (2009) and gamma is the option's gamma.

lead to predictions of lower returns, whereas we find a positive impact on the predictions for large gamma values. We can furthermore show that short-term at-the-money options show the largest and most diverse sensitivity to changes in gamma. The functional form for long-term options differs little for different moneyness regions. Figure I.6 in the Online Appendix replicates these graphs for the second-most important volatility (vega) and jump risk proxy (tlm30). We find a highly nonlinear dependence of N-En's predictions on both proxies, with striking differences across option buckets.

5.4. Impact of the Information Set

Instead of analyzing the importance of single characteristics, we investigate how delta-hedged option return predictability changes when using only a subset of characteristics. We re-estimate each model using only stock-based characteristics (S), option-based characteristics (O), or only those characteristics operating on the bucket- or contract-level

(B+C) and contrast the resulting predictability with that of the full information set.

Figure 12 provides the resulting full-sample R^2_{OS} values. Using all available information produces the highest R^2_{OS} consistent with the idea that more information leads to better forecasts if the models used are sufficiently able to capture this information. Less weight is put on uninformative characteristics and important nonlinear interactions between them are taken advantage of. Restricting to only option-based information (O) comes in as the second place. The benefit of adding stock-based to the 77 option-based characteristics is substantial, given that the R^2_{OS} drops from 2.65% to 1.77% (both significant at the 1% level) for the whole sample if we exclude the whole stock-based characteristics block. Only considering the subset stock-based characteristics (S), however, is detrimental to uncovering option return predictability. The out-of-sample R^2 drops to 0.26%. The benefit of option-derived characteristics is huge when making informed forecasts of future delta-hedged option returns. As an additional check, we also consider whether option-contract and option-bucket information is sufficiently informative, which would render the addition of option-based characteristics for the underlying pointless. We strongly reject this idea, given that the inclusion of option-based characteristics for the underlying boosts out-of-sample predictability R^2_{OS} from 0.93% (B+C) to 1.77% (O). Consistent with the feature group importance shown in Figure 8, contract-based characteristics are highly important, but alone do not suffice to forecast future single-equity option returns.[24]

We use Diebold and Mariano (1995) forecast comparisons to assess whether the forecasts of N-En using more information are statistically more informative. For completeness, we also consider the linear ensemble (L-En) once more to understand in how far allowing for nonlinearities helps when using only a restricted set of characteristics.

Panel A of Table 6 reports results. For the linear and nonlinear ensemble, we find that adding more information is always worthwhile. At the same time, we find a clear hierarchy, which puts the informational content of option-based characteristics above that of stock-based characteristics. Models using both bucket and contract information (B+C) in their respective ensemble class beat models relying solely on stock-based information (S), but are outperformed by models also leveraging the information inherent in option-based information about the underlying (O). The full model for L-En and N-En performs significantly better than all three alternative model specifications. Comparing the fore-

[24]Our findings are robust to using the cross-sectional out-of-sample $R^2_{OS;XS}$ (Figure J.1 in the Online Appendix) and focusing on option buckets (Table J.1 in the Online Appendix). In case of the latter, the resulting best model uses all information for nearly all option buckets. Notable exceptions are in the case of in-the-money short-term puts for both N-En and L-En.

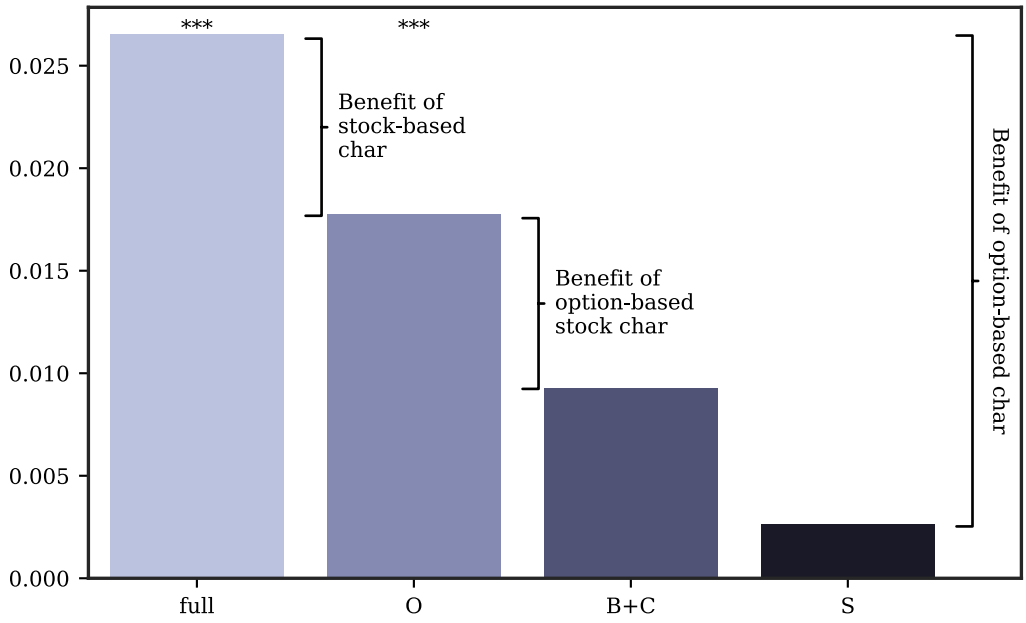


Fig. 12. Restricting the Information Set for N-En

The figure shows the out-of-sample R^2 defined in Equation (4) for N-En with restricted access to the full set of characteristics. The full model is shown in the left bar for reference, and is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Section D. ***, **, * below the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7) for the sample of "all" options. The testing sample spans the years 2003 through 2020.

casts of the linear and nonlinear ensemble, we find that N-En restricted to option-based information (O) manages to surpass all but the full L-En model. In contrast, the full L-En model only manages to provide marginally more accurate forecasts than N-En restricted to stock-based characteristics (S). In line with our intuition that a) more information is always better and b) that nonlinear interactions are most valuable when making informed investment decisions in the options space, we find that the full nonlinear ensemble manages to outperform all four L-En specifications.

The impression from the Diebold and Mariano (1995) comparison carries over to evidence from forecast correlations in Panel B of Table 6. We find the largest similarity of the resulting predictions for linear or nonlinear ensembles estimated on option-based information (both B+C and O) and the full information set ($\rho_{t+1} > 0.7$). The forecast correlation of models using only option- and only stock-based characteristics is particularly low.[25]

[25]For the sake of completeness, Table J.1 in the Online Appendix replicates this analysis for different option buckets. For most buckets we find that the full model leads to the best predictions.

Panel A: Diebold and Mariano (1995) Forecast Comparison

	L-En			N-En			
	O	S	full	B+C	O	S	full
L-En: B+C	**2.79**	−0.81	**3.82**	1.70	**2.60**	0.71	**3.64**
L-En: O		**-2.43**	**3.40**	1.08	**2.22**	0.05	**3.28**
L-En: S			**6.68**	**2.56**	**3.53**	1.50	**4.90**
L-En: full				−0.56	1.23	−1.92	**2.86**
N-En: B+C					**4.53**	**-1.98**	**4.83**
N-En: O						**-5.09**	**3.30**
N-En: S							**8.34**

Panel B: Forecast Correlation

	L-En			N-En			
	O	S	full	B+C	O	S	full
L-En: B+C	**0.89**	0.47	**0.77**	0.67	0.60	0.32	0.58
L-En: O		0.44	**0.84**	0.63	0.66	0.31	0.63
L-En: S			0.53	0.34	0.28	0.61	0.33
L-En: full				0.55	0.61	0.37	0.70
N-En: B+C					**0.86**	0.38	**0.76**
N-En: O						0.35	**0.84**
N-En: S							0.50

Table 6: Forecast Comparison for Characteristic Sets

Panel A of the table shows Diebold and Mariano (1995) test statistics defined in Equation (8) to compare the forecasts for models with restricted access to the full set of characteristics. The full model is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Section D. Significance at the 1% (5%) level is highlighted in light blue (blue). Panel B shows forecast correlations defined in Equation (10). Here, highlighting in light blue (blue) denotes large values with a cutoff at 90% (70%).

6. Sources of Option Return Predictability

Hong and Stein (1999) show theoretically the impact of gradual information diffusion on asset return predictability. In their seminal model, investors are equipped with different subsets of publicly available information. There are various reasons why investors might not infer all public information. First, investors might have limited capabilities to process all information. Second, the same information can be presented in various different forms resulting in a costly search process (Hirshleifer and Teoh, 2003). Third, limits to arbitrage (Shleifer and Vishny, 1997; Pontiff, 2006) can pose restrictions on the elimination of pricing inefficiencies even in the presence of perfect information acquisi-

tion. Given limited attention and informational frictions, some investors are not able to infer the information sets of other market participants from observable asset prices. Consequently, asset prices are predictable as new information is only partially incorporated (Bali and Murray, 2020). In this section, we investigate potential economic mechanisms underlying the sources of option return predictability. In particular, we test whether investor attention, informational frictions, and option mispricing provide an explanation to the observed return predictability in the options market.

6.1. Presence of Informed Investors

We analyze if the return predictability is concentrated in options with lower presence of informed investors and higher trading by non-institutional traders. We hypothesize that option return predictability originates partly by informational frictions, such that the information implied from stock- and option-based characteristics is not directly incorporated into option prices. Informed or sophisticated traders are able to detect and process information sooner, and are therefore able to exploit mispricing. Hence, in the presence and trading of many informed traders, information is impounded into prices faster, or put differently, information diffuses faster. Consequently, slow information diffusion leads to additional option return predictability that is more pronounced for options with a lower share of informed traders.

We use two proxies for the presence of informed investors. First, we consider the case of institutional investors. Institutional investors are deemed to be more sophisticated and to exhibit superior ability in processing and reacting to information (see, among others, Chang, Hsieh, and Wang, 2015). Boehmer and Kelley (2009) find that stocks with a higher degree of institutional ownership are priced more efficiently. Following Eisdorfer et al. (2020), we take institutional ownership in the underlying stock as a proxy for the fraction of informed traders in the options market. Hence, option return predictability should be weaker for options with higher share of institutional investors. Additionally, we use analyst coverage as a second proxy for the presence of informed investors. The previous literature has provided empirical evidence that analyst coverage is an important channel through which information is incorporated into asset prices (Womack, 1996; Barber, Lehavy, McNichols, and Trueman, 2001; Gleason and Lee, 2003; Jegadeesh, Kim, Krische, and Lee, 2004; Kelly and Ljungqvist, 2012). Ellul and Panayides (2018) show that the termination of analysts covering a specific stock impairs price efficiency, causes more informed trading and leads to higher profitability of insider traders. Therefore,

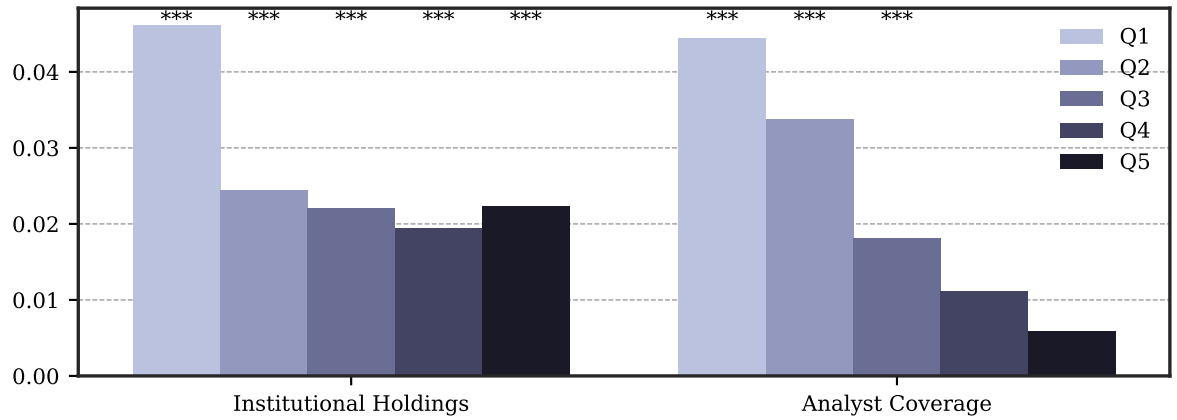


Fig. 13. Predictability and Profitability Conditional on Informed Investor Presence

The figure shows out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En for different quintiles of institutional ownership and analyst coverage of the underlying stocks, respectively. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7).

we hypothesize that higher analyst coverage of the underlying stocks translates into less option return predictability.

In order to investigate the resulting differences in return predictability, we form quintile splits at time t of the stocks in our sample, either by institutional ownership, or number of analysts covering the firm (Q1–Q5). Subsequently, we contrast the predictability of options written on stocks with different levels of institutional ownership (analyst coverage). Figure 13 depicts the R^2_{OS} values for the sub-samples. Higher institutional ownership directly translates to lower predictability of option contracts by N-En, confirming our hypothesis. We find the highest predictability for stocks in the lowest institutional ownership quintile at $R^2_{OS} = 4.6\%$. For stocks with higher institutional ownership, we find levels of predictability at around 1.9% to 2.5%. The predictability uncovered is statistically significant for all levels of institutional ownership. Turning to analyst coverage, option return predictability is strictly monotonically decreasing in the number of analysts covering a firm. In the lowest quintile, $R^2_{OS} = 4.4\%$, which drops to an insignificant 0.6% in Q5. Figure K.1 in the Online Appendix shows similar findings in the case of cross-sectional out-of-sample predictability.

The above findings motivate us to study the effect of high and low institutional ownership (analyst coverage) on the economic significance of return predictability in more detail. Precisely, we perform bivariate portfolio level analyses to study how option return predictability varies across different levels of institutional ownership and analyst coverage. We sort options into quintiles based on the level of institutional ownership (or

	Low Pred.	2	3	4	High Pred.	H-L
			Institutional Holdings			
Low	-1.985***	-0.505***	-0.222	-0.067	0.508**	2.492***
2	-1.138***	-0.342**	-0.177	-0.059	0.304	1.441***
3	-1.091***	-0.338**	-0.149	-0.024	0.311	1.402***
4	-1.027***	-0.308**	-0.129	0.031	0.445**	1.473***
High	-1.270***	-0.379***	-0.131	0.055	0.512**	1.782***
H-L	0.715***	0.126**	0.091**	0.122***	0.004	-0.710***
			Analyst Coverage			
Low	-2.056***	-0.483***	-0.076	0.213	0.870***	2.926***
2	-1.439***	-0.348**	-0.068	0.143	0.646***	2.085***
3	-0.967***	-0.229*	-0.063	0.097	0.458**	1.425***
4	-0.799***	-0.209	-0.077	0.039	0.406*	1.205***
High	-0.736***	-0.272**	-0.094	-0.003	0.279	1.015***
H-L	1.321***	0.211***	-0.017	-0.216***	-0.591***	-1.911***

Table 7: Bivariate Portfolios of Institutional Attention and Expected Returns

The table shows realized returns for quintile portfolios following the predictions by the nonlinear ensemble N-En within quintiles sorted by analyst coverage in the upper panel and institutional ownership share in the lower panel. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level as defined in Equation (7). We also show the realized returns and significance for the resulting high-minus-low portfolios.

analyst coverage) of the underlying stock at the end of month t. Subsequently, within each quintile, options are further sorted into quintiles by the one-month-ahead expected return forecast of N-En.

Panel A of Table 7 presents the results for the 25 (5x5) portfolios sorted by expected returns conditional on institutional ownership. It is evident from the last column that H-L return spreads are significantly positive regardless of the level of institutional ownership. Monthly returns range from 2.49% to 1.40%. The spreads between the lowest and highest return forecast quintiles are nearly monotonically decreasing such that the difference-in-difference return spread is economically (-0.71% per month) and statistically significant at the 1% level.

Panel B of Table 7 shows even stronger results for analyst coverage. The return spreads in the last column confirm pervasive option return predictability for all analyst coverage quintiles by the N-En model. The return spreads are strictly monotonically decreasing in the degree of analyst coverage, confirming our hypotheses that option return predictability is directly related to institutional attention. Furthermore, this leads to a negative and highly significant difference between the high-minus-low spreads of the highest and lowest

analyst coverage quintile of -1.91% per month. Taken together, the results for analyst coverage and institutional ownership are in line with the notion that the presence of informed traders and the resulting faster diffusion of information reduces option return predictability.

6.2. Option Demand by Types of Market Participants

We now investigate in how far option demand by professional traders and public customers translates into return predictability. For this, we obtain open/close option transaction data for various types of market participants from four NASDAQ exchanges and the CBOE C1 exchange from January 2011 through December 2020.[26] The open/close data set unfortunately does not provide a clear distinction between institutional and retail trading. As an approximation, we separately consider "public customers", which we denote as Customers (C), and the joint trading activity of "proprietary traders" and "professional customers", which we summarize as Professionals (P). While this classification does not guarantee that trades of institutions do not end up in the group of C, the relative concentration of informed trades is expected to be much higher in group P compared to group C.[27] If anything, under our assumption that institutional trading leads to more efficient prices, our results should be even stronger if we were able to adequately differentiate between retail and institutional options trading. Eisdorfer et al. (2020) construct investor groups in the options market in a similar fashion to understand whether retail-driven trades exhibit more behavioral biases.

An investor may express pessimistic views about the underlying stock by purchasing puts or alternatively short-selling call options. The opposite holds true for betting on optimistic expectations. We therefore consider signed open interest in individual option contracts over the last month, and analyze how buying or selling pressure by C and P impacts the resulting return predictability during month $t+1$. Define the signed volume over the last month for contract i as:

$$Volume_{i,t}^{x;Buy} = Volume_{i,t}^{x;OpenBuy} + Volume_{i,t}^{x;CloseBuy} \tag{13}$$

$$Volume_{i,t}^{x;Sell} = Volume_{i,t}^{x;OpenSell} + Volume_{i,t}^{x;CloseSell}, \tag{14}$$

[26]The four NASDAQ exchanges are NASDAQ PHOTO, NASDAQ Options Market, NASDAQ ISE, and NASDAQ GEMX.

[27]We also note that the classification into "professional customer" differs between the CBOE and NASDAQ exchanges.

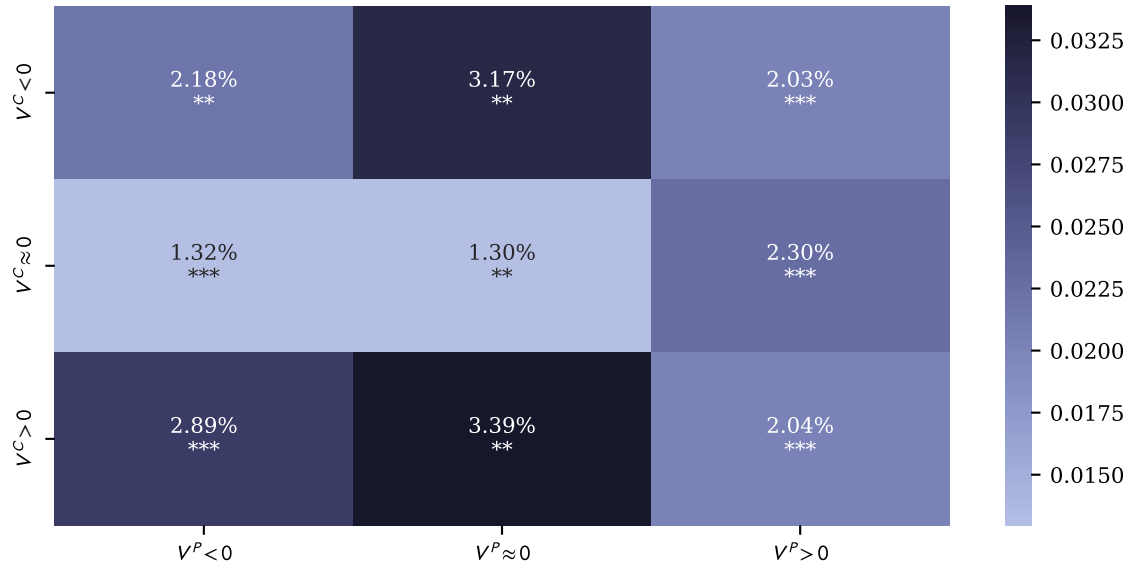


Fig. 14. Predictability Conditional on Option Demand

The figure shows the out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En conditional on the net open interest of retail and institutional investors, I. We independently form terciles based on the cross-sectional distribution of I^R and I^I. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

for option j at time t and $x \in$ [C, P]. Net demand for each investor group is consequently defined as the difference between the daily buy and sell volume, which we denote as V^x. We calculate monthly net demand of option i by summing up the respective daily signed volume. We scale the resulting signed volume measures V^x by the total *non-signed* volume of public customers, professional traders, and proprietary trading desks to make it comparable across stocks. Furthermore, we aggregate individual option volume $V^x_{i,t}$ at the level of the underlying to arrive at $V^x_{S,t}$. Following Garleanu et al. (2009), demand in one option has price effects also on neighboring options of the same underlying. While any demand potentially increases the informativeness of market prices, we argue that the impact of institutional demand is amplified. This in turn should correspond to decreasing return predictability.

Figure 14 shows the resulting R^2_{OS} for public customers (C) vs. professional traders (P). We define terciles based on the cross-sectional distribution of V^x during month t and analyze the resulting predictability for $t+1$. We find multiple interesting results: First, option return predictability is highest for options with little to no demand by professionals ($V^P \approx 0$), while it is lowest if demand by both investor groups is missing ($R^2_{OS} = 1.30\%$). Second, compared to these cases, a lack of customer demand if demand by

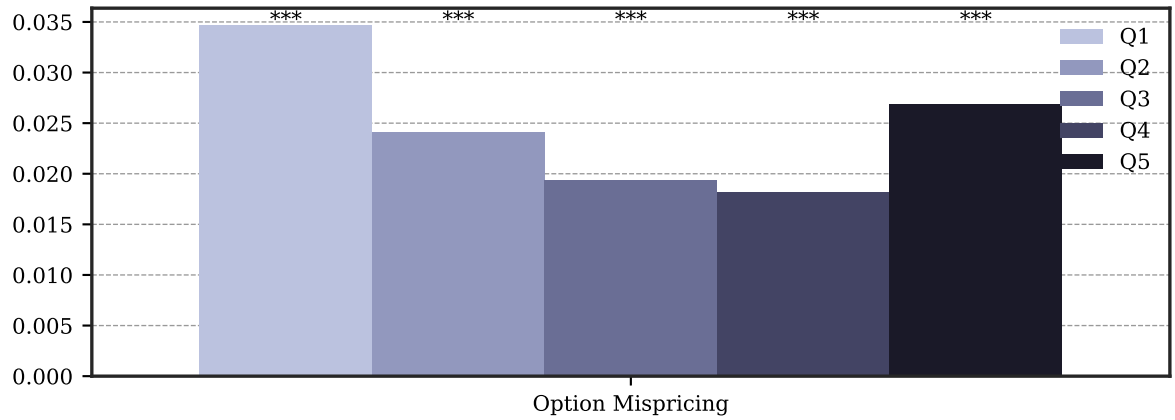


Fig. 15. Predictability and Profitability Conditional on Option Mispricing – R^2_{OS}

The left panel of the figure shows out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En for different quintiles of option mispricing. We calculate option mispricing using the stocks' realized volatility over the last quarter and calculate "fair prices" for short-term at-the-money options using it as input to the Black and Scholes (1973) model. Mispricing is then defined as $\log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the obtained "fair price". ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7).

professionals is present leads to much lower predictability, consistent with our hypothesis that professionals are quicker to impound information into market prices. Our model (N-En) manages to identify these situations of mispricing and allows an investor to profit from them. The model also uncovers high predictability in cases where public customers are buying options of an underlying and professionals facilitate their buying intent by shorting, which is in line with Garleanu et al. (2009) who argue that professional traders tend to act as market markers. We add to this by showing that they tend to act this way in cases where N-En uncovers high predictability of the respective option from which professionals can profit.

6.3. Option Mispricing

We expect to find higher levels of predictability for options that are mispriced as the nonlinear ensemble model should manage to identify these opportunities and correctly propose shorting over- and investing into undervalued options.

We follow Eisdorfer et al. (2020) to quantify option mispricing as the ratio between theoretical and observed option prices. The theoretical option price is given by the Black and Scholes (1973) pricing model, where we use each underlying stock's realized volatility over the past quarter, estimated using high-frequency price data from the NYSE TAQ database, as our estimate for the expected volatility. For all short-term at-the money

options, we compare the log of the theoretical price with the log of the price observed in the market, i.e., Mispricing $= \log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the theoretical and O the observed mid price. A high (low) level of mispricing thus denotes overvaluation (undervaluation). Averaging over all short-term at-the-money options, we obtain one level of mispricing per underlying stock at each point in time.

The predictability for different mispricing quintiles is provided in Figure 15. Consistent with our hypothesis that predictability clusters within over- and under-valued options, we find the highest level of predictability in the most undervalued, followed by the most overvalued options. This create a U-shaped predictability pattern in the option's mispricing.

This predictability pattern also translates into additional profitability, as depicted in Table 8. The high-minus-low spread within each option mispricing portfolio is highly significant, generating monthly profits upwards of 1.52% per month. Interestingly, the difference between these spreads for under- and overvalued stocks is insignificant at -0.06% per month, suggesting that significant trading opportunities exist within over- *and* undervalued options. Instead, when we compare the high-minus-low portfolio returns of mispriced options with the portfolios using only fairly priced options (so $O \approx \widetilde{O}$), we find that they are significantly different from each other, with higher returns both for undervalued (0.44%) and overvalued options (0.38% per month). To summarize, the employed machine learning methods manage to pick up on mispricing opportunities in the options market and allow investors to profit from them.

7. Conclusion

An extensive literature examines cross-sectional determinants of stocks, bonds, currencies, mutual funds, and hedge funds. However, research on cross-sectional predictors of option returns is relatively scarce and not very well understood. In this paper, we close this gap in the literature and identify variables that predict the cross-sectional differences in delta-hedged option returns. Predicting option returns is of foremost relevance for retail and institutional investors as the importance of option markets for hedging and speculation purposes has strongly increased in the past years.

In this paper, we apply machine learning to predict individual U.S. option returns using a set of 77 option-based and 193 stock-based characteristics in the period from 1996 to 2020. Empirically, we derive several results that enhance our knowledge on empirical asset pricing. First, we show that the complexity of the machine learning models matters

	Low Pred.	2	3	4	High Pred.	H-L	vs. Fair
				Option Mispricing			
Under	-1.278***	-0.304**	-0.054	0.156	0.678***	1.956***	0.441***
2	-1.101***	-0.255*	-0.045	0.132	0.593***	1.694***	0.179**
Fair	-0.995***	-0.234	-0.087	0.069	0.520**	1.515***	---
4	-0.994***	-0.232*	-0.031	0.134	0.545**	1.539***	0.024
Over	-1.467***	-0.381***	-0.103	0.059	0.427*	1.894***	0.379***
Over-Under	-0.190*	-0.077	-0.050	-0.097*	-0.251**	-0.062	---

Table 8: Bivariate Portfolios of Option Mispricing and Expected Returns

The table shows realized returns for quintiles portfolios following the predictions by the nonlinear ensemble N-En within quintiles sorted by option mispricing. We calculate option mispricing using the stocks' realized volatility over the last quarter and calculate "fair prices" for short-term at-the-money options using it as input to the Black and Scholes (1973) model. Mispricing is then defined as $\log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the obtained "fair price". We also show the realized returns and significance for the resulting high-minus-low portfolios. Column "vs. Fair" tests statistical differences between the H-L portfolios for under- and overvalued options, and fairly valued options. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level as defined in Equation (7).

for prediction and observe that nonlinear models outperform linear models in terms of out-of-sample R-squared. Second, our results reveal that a trading strategy based on nonlinear machine learning forecasts is highly profitable and remains statistically and economically significant even after accounting for high levels of transaction costs. Third, we find that characteristics describing the option's location on the underlying's implied volatility surface and nonlinear option risk measures such as jump and volatility risk are important determinants of option returns. Finally, we document that option return predictability is largely driven by informational frictions and mispricing in the options market. In line with this notion, we find that option return predictability is higher for options on stocks characterized by (i) low institutional ownership and low analyst coverage, (ii) with little demand by professional traders, but substantial interest by public customers, such as retail investors, and (iii) mispriced options.

Bibliography

Agarwal, V., Naik, N. Y., 2004. Risks and portfolio decisions involving hedge funds. Review of Financial Studies 17, 63–98.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

An, B.-J., Ang, A., Bali, T. G., Cakici, N., 2014. The joint cross section of stocks and options. Journal of Finance 69, 2279–2337.

Arik, S. O., Pfister, T., 2019. Tabnet: Attentive interpretable tabular learning (2019). arXiv preprint arXiv:1908.07442 .

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. Review of Financial Studies 16, 527–566.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2021. Different strokes: Return predictability across stocks and bonds with machine learning and big data.

Bali, T. G., Hovakimian, A., 2009. Volatility spreads and expected stock returns. Management Science 55, 1797–1812.

Bali, T. G., Murray, S., 2020. In search of a factor model for optionable stocks.

Baltussen, G., van Bekkum, S., van der Grient, B., 2018. Unknown Unknowns: Uncertainty About Risk and Stock Returns. Journal of Financial and Quantitative Analysis 53, 1615–1651.

Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. Journal of Finance 66, 911–946.

Barber, B., Lehavy, R., McNichols, M., Trueman, B., 2001. Can investors profit from the prophets? security analyst recommendations and stock returns. Journal of Finance 56, 531–563.

Bates, J. M., Granger, C. W., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

Bergsma, K., Fodor, A., Tedford, E., 2020. A closer look at the disposition effect in u.s. equity option markets. Journal of Behavioral Finance 21, 66–77.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Review of Financial Studies 34, 1046–1089.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Blau, B. M., Nguyen, N., Whitby, R. J., 2014. The information content of option ratios. Journal of Banking & Finance 43, 179 – 187.

Boehmer, E., Kelley, E. K., 2009. Institutional investors and the informational efficiency of prices. Review of Financial Studies 22, 3563–3594.

Bollen, N. P., Whaley, R. E., 2004. Does net buying pressure affect the shape of implied volatility functions? Journal of Finance 59, 711–753.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. Journal of Financial Economics 118, 113–134.

Borochin, P., Chang, H., Wu, Y., 2020. The information content of the term structure of risk-neutral skewness. Journal of Empirical Finance 58, 247 – 274.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Buchner, M., Kelly, B. T., 2020. A Factor Model for Option Returns.

Buraschi, A., Jackwerth, J., 2001. The price of a smile: Hedging and spanning in option markets. Review of Financial Studies 14, 495–527.

Byun, S.-J., Kim, D.-H., 2016. Gambling preference and individual equity option returns. Journal of Financial Economics 122, 155–174.

Cao, J., Han, B., 2013. Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics 108, 231–249.

Cao, J., Han, B., Tong, Q., Zhan, X., 2021. Option return predictability. Review of Financial Studies, forthcoming.

Cao, J., Vasquez, A., Xiao, X., Zhan, X., 2019. Volatility Uncertainty and the Cross-Section of Option Returns.

Cao, M., Wei, J., 2010. Option market liquidity Commonality and other characteristics. Journal of Financial Markets p. 29.

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

Chang, C.-C., Hsieh, P.-F., Wang, Y.-H., 2015. Sophistication, sentiment, and misreaction. Journal of Financial and Quantitative Analysis 50, 903–928.

Chen, L., Pelger, M., Zhu, J., 2021. Deep learning in asset pricing.

Christoffersen, P., Goyenko, R., Jacobs, K., Karoui, M., 2018. Illiquidity premia in the equity options market. Review of Financial Studies 31, 811–851.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cremers, M., Halling, M., Weinbaum, D., 2015. Aggregate jump and volatility risk in the cross-section of stock returns. Journal of Finance 70, 577–614.

Cremers, M., Weinbaum, D., 2010. Deviations from put-call parity and stock return predictability. Journal of Financial and Quantitative Analysis 45, 335–367.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J., Santos, A. A. P., 2021. Can machine learning help to select portfolios of mutual funds?

Dennis, P., Mayhew, S., 2002. Risk-neutral skewness: Evidence from stock options. Journal of Financial and Quantitative Analysis 37, 471–493.

Dew-Becker, I., Giglio, S., 2020. Cross-sectional uncertainty and the business cycle: evidence from 40 years of options data. National Bureau of Economic Research.

Dew-Becker, I., Giglio, S., Kelly, B., 2021. Hedging macroeconomic and financial uncertainty and volatility. Journal of Financial Economics 142, 23–45.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Eisdorfer, A., Goyal, A., Zhdanov, A., 2020. Cheap options are expensive.

Ellul, A., Panayides, M., 2018. Do financial analysts restrain insiders' informational advantage? Journal of Financial and Quantitative Analysis 53, 203–241.

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

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fama, E. F., French, K. R., 2018. Choosing factors. Journal of Financial Economics 128, 234–252.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. Journal of Finance 75, 1327–1370.

Filippou, I., Rapach, D., Taylor, M. P., Zhou, G., 2020. Exchange rate prediction with machine learning and a smart carry portfolio.

Fong, K. Y., Holden, C. W., Trzcinka, C. A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Review of Financial Studies 33, 2326–2377.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics pp. 1189–1232.

Garleanu, N., Pedersen, L. H., Poteshman, A. M., 2009. Demand-based option pricing. Review of Financial Studies 22, 4259–4299.

Giglio, S., Liao, Y., Xiu, D., 2021. Thousands of alpha tests. Review of Financial Studies 34, 3456–3496.

Gilad-Bachrach, R., Rashmi, K., 2015. Dart: Dropouts meet multiple additive regression trees. Cornell University: Cornell, Ithaca, NY, USA .

Gleason, C. A., Lee, C. M. C., 2003. Analyst forecast revisions and market price discovery. Accounting Review 78, 193–225.

Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. Journal of Financial Economics 94, 310–326.

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

Goyenko, R., Zhang, C., 2021. The joint cross section of option and stock returns predictability with big data and machine learning.

Goyenko, R. Y., Holden, C. W., Trzcinka, C. A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153–181.

Grammig, J., Hanenberg, C., Schlag, C., Sönksen, J., 2020. Diverging roads: Theory-based vs. machine learning-implied stock risk premia.

Green, J., Hand, J. R. M., Zhang, X. F., 2017. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. Review of Financial Studies 30, 4389–4436.

Grünthaler, T., Lorenz, F., Meyerhof, P., 2020. The leverage bearing capacity: A new tool for intermediary asset pricing.

Gu, S., Kelly, B. T., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Han, Y., He, A., Rapach, D. E., Zhou, G., 2021. Cross-sectional out-of-sample stock return prediction with many characteristics.

Heston, S. L., Sadka, R., 2008. Seasonality in the cross-section of stock returns. Journal of Financial Economics 87, 418–445.

Hiraki, K., Skiadopoulos, G., 2020. The Contribution of Frictions to Expected Returns.

Hirshleifer, D., Teoh, S. H., 2003. Limited attention, information disclosure, and financial reporting. Journal of Accounting and Economics 36, 337–386.

Hoerl, A. E., Kennard, R. W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

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

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural networks 2, 359–366.

Huang, T., Li, J., 2019. Option-implied variance asymmetry and the cross-section of stock returns. Journal of Banking & Finance 101, 21–36.

Jegadeesh, N., Kim, J., Krische, S. D., Lee, C. M. C., 2004. Analyzing the analysts: When do recommendations add value? Journal of Finance 59, 1083–1124.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2021. Is there a replication crisis in finance? National Bureau of Economic Research.

Johnson, T. L., So, E. C., 2012. The option to stock volume ratio and future returns. Journal of Financial Economics 106, 262 – 286.

Kanne, S., Korn, O., Uhrig-Homburg, M., 2020. Stock illiquidity and option returns.

Karakaya, M. M., 2014. Characteristics and expected returns in individual equity options.

Kelly, B. T., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. Review of Financial Studies 25, 1366–1413.

Kelly, B. T., Palhares, D., Pruitt, S., 2020a. Modeling corporate bond returns.

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kelly, B. T., Pruitt, S., Su, Y., 2020b. Instrumented principal component analysis.

Keloharju, M., Linnainmaa, J. T., Nyberg, P., 2016. Return seasonalities. Journal of Finance 71, 1557–1590.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25, 1097–1105.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2016. Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474 .

Leippold, M., Wang, Q., Zhou, W., 2021. Machine-learning in the chinese factor zoo. Journal of Financial Economics, forthcoming.

Lesmond, D. A., Ogden, J. P., Trzcinka, C. A., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113–1141.

Lettau, M., Pelger, M., 2020. Factors that fit the time series and cross-section of stock returns. Review of Financial Studies 33, 2274–2325.

Li, B., Rossi, A. G., 2020. Selecting mutual funds from the stocks they hold: A machine learning approach.

Lu, Z., Murray, S., 2019. Bear beta. Journal of Financial Economics 131, 736–760.

Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777.

Martin, I., Nagel, S., 2020. Market efficiency in the age of big data. National Bureau of Economic Research.

Moritz, B., Zimmermann, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns.

Muravyev, D., Pearson, N. D., 2020. Options trading costs are lower than you think. Review of Financial Studies 33, 4973–5014.

Muravyev, D., Pearson, N. D., Pollet, J. M., 2021. Is there a risk premium in the stock lending market? evidence from equity options. Journal of Finance, forthcoming.

Murray, S., Xiao, H., Xia, Y., 2021. Charting by machines.

Nagel, S., 2021. Machine Learning in Asset Pricing. Princeton University Press.

Neuhierl, A., Tang, X., Varneskov, R. T., Zhou, G., 2021. Option characteristics as cross-sectional predictors.

Newey, W. K., West, K. D., 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica 55, 703–708.

Ofek, E., Richardson, M., Whitelaw, R. F., 2004. Limited arbitrage and short sales restrictions: Evidence from the options markets. Journal of Financial Economics 74, 305–342.

Pástor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. Journal of Accounting and Economics 42, 35–52.

Ramachandran, L. S., Tayal, J., 2021. Mispricing, short-sale constraints, and the cross-section of option returns. Journal of Financial Economics 141, 297–321.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Rapach, D. E., Strauss, J. K., Zhou, G., 2013. International stock return predictability: what is the role of the united states? Journal of Finance 68, 1633–1662.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of finance 39, 1127–1139.

Roll, R., Schwartz, E., Subrahmanyam, A., 2010. O/S: The relative trading activity in options and stock. Journal of Financial Economics 96, 1–17.

Schlag, C., Thimme, J., Weber, R., 2020. Implied volatility duration: A measure for the timing of uncertainty resolution. Journal of Financial Economics, forthcoming.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.

Steel, M. F., 2020. Model averaging and its use in economics. Journal of Economic Literature 58, 644–719.

Tian, M., Wu, L., 2021. Limits of arbitrage and option risk premium.

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

Vasquez, A., 2017. Equity volatility term structures and the cross section of option returns. Journal of Financial and Quantitative Analysis 52, 2727–2754.

Vasquez, A., Xiao, X., 2021. Default risk and option returns.

Vilkov, G., Xiao, Y., 2012. Option-implied information and predictability of extreme returns.

Womack, K. L., 1996. Do brokerage analysts' recommendations have investment value? Journal of Finance 51, 137–167.

Wu, W., Chen, J., Yang, Z., Tindall, M. L., 2021. A cross-sectional machine learning approach for hedge fund return prediction and selection. Management Science 67, 4577–4601.

Xing, Y., Zhang, X., Zhao, R., 2010. What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? Journal of Financial and Quantitative Analysis 45, 641–662.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology) 67, 301–320.

Online Appendix

Online Appendix A Methods Used

Following Gu et al. (2020) we compare a variety of simpler and complex methods in our empirical analysis. Within the subgroup of linear models we include simple penalized regressions in the form of an elastic net (ENet), Ridge and Lasso, as well as a combination of dimension reduction techniques and linear regression, partial least squares (PLS) and principal component regressions (PCR).

For nonlinear estimators we differentiate between tree-based methods and neural networks. Explicitly, we compare the performance of random forests (RF), gradient-boosted regression trees (GBR), and gradient-boosted regression trees with dropout (DART), proposed in Gilad-Bachrach and Rashmi (2015). Here, leaves are randomly "dropped" during training, which regularizes the process and helps avoid overfitting. We use Microsoft's LightGBM implementation for our tree-based methods Ke, Meng, Finley, Wang, Chen, Ma, Ye, and Liu (2017), which grows trees leaf-wise, aiding in faster convergence.

Feed-forward neural networks are implemented in PyTorch Paszke, Gross, Massa, Lerer, Bradbury, Chanan, Killeen, Lin, Gimelshein, Antiga, Desmaison, Kopf, Yang, DeVito, Raison, Tejani, Chilamkurthy, Steiner, Fang, Bai, and Chintala (2019). In contrast to Gu et al. (2020) we vary the number of hidden layers and nodes during hyperparameter optimization. This way, we combine the predictions of shallow and deep neural nets in one ensemble, having the benefit of probing different parts of the data and combining the results. For the neural network implementations we rely on the optimizer AdamW (Loshchilov and Hutter, 2017)to tune the weights, which adapts the learning rates during training and correctly implements weight-decay of individual training weights as an improvement upon the well-known Adam optimizer (Kingma and Ba, 2014). We also follow the idea of Reddi, Kale, and Kumar (2019) which promises better theoretical convergence of our optimization procedure.

To come up with candidate solutions of our models, we optimize over the mean squared error for a given set of hyperparameters θ, which are unique to the respective model class (more on this below):

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{i,s,t+1} - g(z_{i,s,t}))^2 \tag{A.1}$$

Online Appendix B Estimation Details

Machine learning algorithms crucially depend on hyperparameters that govern the amount of regularization of the model in question, which ultimately determines the generalizability of the resulting representation of $g(\star)$ in Equation (2). Hyperparameters have to be set by the researcher before the actual training of the model begins. Following Gu et al. (2020) we optimize over the model's hyperparameter in a validation sample. More specifically, we estimate model parameters on the first five years of data, validate the hyperparameters in the next two, and test the resulting model's predictions in the following year. We repeat this procedure for each year in the testing sample from 2003 through 2020, increasing the number of training years by one at each iteration.

Within each training sample, we optimize the mean-squared error (Equation (A.1)) of the in-sample prediction, for a given set of randomly-chosen hyperparameters θ (Bergstra and Bengio, 2012). The different sets are compared by their mean-squared error in the validation sample. To decrease the computational burden, and allocate more time to the most promising θs, we use the *asynchronous successive halving algorithm* put forth by Li, Jamieson, Rostamizadeh, Gonina, Hardt, Recht, and Talwalkar (2018).[28] This is an extension of the popular *Hyperband* scheme for hyperparameter optimization, which allocates more iterations to the most promising θs Li, Jamieson, DeSalvo, Rostamizadeh, and Talwalkar (2017). This search exercise has the added benefit of providing close-to-best solutions on the go. We thus use an equally-weighted ensemble of the eight best models within each model class. This ensemble generalizes better to unseen data. While estimating, we do not apply a weighting scheme to the return observations, but note that one benefit of our option sample is that the total information per underlying s used in the estimation procedure scales linearly in the number of outstanding option contracts available for it. Thereby we automatically shift estimation towards larger and more liquid stocks. To assure that we do not overfit on the training data, we employ early stopping if a trial's validation error $\mathcal{L}$ has not decreased for eight iterations (32 for tree-based methods).

Table B.1 shows the hyperparameter ranges used for each model type as well as additional information on how we use stochastic gradient descent to estimate model parameters when applicable.

[28] We use the implementation in Ray Tune (Liaw, Liang, Nishihara, Moritz, Gonzalez, and Stoica (2018)) and carry out model estimation on Palma II, the high-performance computing cluster of the University of Muenster: `https://www.uni-muenster.de/IT/services/unterstuetzungsleistung/hpc/`.

	ENet, Lasso, & Ridge	
	Max Epochs	64
	Random search trials	512
	Batch size	$\in [2^{12}, 2^{14}, 2^{16}]$
	Learning rate	$\in [0.001, 0.01, 0.1]$
	α	$\mathcal{LU}(1e^{-6}, 1e^{-2})$
ENet	λ	$\mathcal{U}(0, 1)$
	PCR & PLS	
	Number of Components	$\in [1, 2, 3, 4, 5, 6]$
	FFN	
	Max Epochs	64
	Random search trials	512
	Batch size	$\in [2^{12}, 2^{14}, 2^{16}]$
	Learning rate	$\in [0.001, 0.01, 0.1]$
	Weight decay	$\mathcal{U}(0, 0.1)$
	Amsgrad (Reddi et al., 2019)	True
	First layer size	$\in [32, 64, 128]$
	Number of hidden layers	$\in [1, 2, 3, 4, 5]$
	Dropout probability	$\mathcal{U}(0, 0.5)$
	RF, GBR, & Dart	
	Max trees	1024
	Random search trials	512
	Learning rate	$\in [0.01, 0.1, 1]$
	Max depth per tree	$\mathcal{U}^{\text{int}}(2, 10)$
	Max number of leaves per tree	$\mathcal{U}^{\text{int}}(2, 512)$
	l1 regularization	$\mathcal{U}(0, 0.1)$
	l2 regularization	$\mathcal{U}(0, 0.1)$
	Fraction of features per run	$\mathcal{U}(0.25, 1)$
	Bagging fraction	$\mathcal{U}(0.25, 1)$
	Bagging frequency	$\in [1, 10, 50]$
Dart	Dropout probability	$\in [0.05, 0.1, 0.15]$
Dart	Probability of skipping dropout	$\in [0.25, 0.5]$

Table B.1: Hyperparameters for the Models Considered.

The table shows the hyperparameters and the boundaries from which they are randomly drawn to optimize them for each model considered. $\mathcal{U}$ ($\mathcal{LU}$, $\mathcal{U}^{\text{int}}$) refers to drawing from a uniform (log-uniform, integer-wise uniform) distribution within the respective boundaries.

Online Appendix C Option-Based Characteristics

This section describes a broad set of the 77 option-based characteristics, motivated by earlier studies on the cross-section of option and/or stock returns. Out of the 77 we compute, 42 characteristics operate on the level of the underlying stock, 19 on the level of option buckets (that is, we differentiate between different parts of the time-to-maturity and moneyness domain of options, described in Section 4.3), and 16 on the level of individual option contracts.

C.1 Stock-Level

1. **Implied volatility slope (*ivslope*)**. Following Vasquez (2017), the slope of the implied volatility term structure is defined as

$$ivslope = IV_{LT} - IV_{1M},$$

   where IV_{1M} is the average of short-term atm put and call implied volatilities and IV_{LT} denotes the average volatility of atm put and call options that have the longest time to maturity available and the same strikes as the short-term options.

2. **Risk-neutral skewness (*rnsτ*)**. Risk-neutral skewness for different times to maturity τ. We include $\tau \in [30, 91, 182, 273, 365]$ days as Borochin, Chang, and Wu (2020) has stressed the importance of short term and long term risk-neutral skewness for the cross-section of equity returns.

3. **Risk-neutral kurtosis (*rnkτ*)**. Risk-neutral kurtosis for different times to maturity τ. We include $\tau \in [30, 91, 182, 273, 365]$ days.

4. **Option-implied variance asymmetry (*ivarud30*)**. The difference between upside and downside risk-neutral semivariances according to Huang et al. (2019).

5. **Option implied tail loss (*tlm30*)**. A forward-looking tail loss measure according to Vilkov and Xiao (2012). It is computed as

$$tlm30 = \frac{\beta(K)}{1-\xi},$$

   where $\beta(K)$ and ξ are the scaling parameter and tail shape parameter of a generalized Pareto distribution $G_{\xi,\beta(K)}$. The scaling parameter β depends on a cutoff

value K.

6. **Stock vs. option volume (*so*)**. Following Roll et al. (2010), the ratio of the number of the underlying's traded shares and the trading volume for all options on the underlying.

7. **Log of stock vs. option volume (*lso*)**. Following Roll et al. (2010), the natural logarithm of ***so***.

8. **Stock vs. option volume (*dso*)**. Following Roll et al. (2010), the ratio of the transacted dollar amount in the underyling's shares and the transacted dollar amount of all options on the underlying.

9. **Log of stock vs. option volume (*ldso*)**. Following Roll et al. (2010), the natural logarithm of ***dso***.

10. **Modified stock vs. option volume (*modso*).** Following Johnson and So (2012), the ratio of the number of the underlying's traded shares and the trading volume for all options on the underlying. The difference to ***so*** is that Johnson and So (2012) apply stricter data filters than Roll et al. (2010).

11. **Put-call ratio (*pcratio*)**. Following Blau et al. (2014), the total put volume divided by the total options volume over the last month for a given underlying.

12. **Contribution of market frictions to expected returns (*fric*)**. Hiraki and Skiadopoulos (2020) show that scaled deviations of put-call-parity measure the contribution of market frictions to expected returns. Consequently, *fric* is defined as

$$fric = R^0_{t,T} \frac{\tilde{S}_t(K,T) - S_t}{S_t},$$

where $\tilde{S}_t(K,T) = C_t(K,T) - P_t(K,T) + \frac{K+D_t}{R^0_{t,T}}$. S_t denotes the stock price at time t, its dividend payment is given by D_t. The time t price of a call option and put option with strike price K and maturity date T are given by $C_t(K,T)$ and $P_t(K,T)$, respectively. $R^0_{t,T}$ denotes the gross risk-free rate over the period from t to T.

13. **Proportional bid-ask spread (*pba*).** Following Cao and Wei (2010), we use the proportional bid-ask spread as a measure of illiquidity

$$pba = \frac{\sum_j VOL_j \times \frac{ask_j - bid_j}{0.5 \times (ask_j + bid_j)}}{\sum_j VOL_j},$$

where VOL_j denotes the trading volume in option j, ask_j and bid_j the bid and ask spread of option j, respectively.

14. **Dollar trading volume (*dvol*).** Following Cao and Wei (2010), we include the dollar trading volume across all options,

$$\sum_j VOL_j \times (ask_j + bid_j)/2.$$

15. **Absolute illiquidity (*ailliq*).** Following Cao and Wei (2010), we introduce the absolute illiquidity as

$$ailliq = \frac{\sum_j \times \frac{|\Pi_t|}{DVOL_j}}{\sum_j VOL_j},$$

where $DVOL_j$ denotes the dollar trading volume in option j.

16. **Percentage illiquidity (*pilliq*).** Following Cao and Wei (2010), we introduce the absolute illiquidity as

$$ailliq = \frac{\sum_j \times \frac{|\Pi_t|}{DVOL_j \times O_j}}{\sum_j VOL_j},$$

where $DVOL_j$ denotes the dollar trading volume in option j, and O_j the price of option j.

17. **Trading volume (*vol*).** Following Cao and Wei (2010), we include the trading volume across all options as defined as $\sum_j VOL_j$, with VOL_j being the volume in option j.

18. **Number of traded options (*nopt*).** The average number of options per underlying stock per month.

19. **Total open interest (*toi*).** Open interest across all options on an underlying.

20. **Volatility uncertainty (*volunc*).** Following Cao et al. (2019), we calculate monthly volatility-of-volatility based on different measures of daily volatility estimates. As a first measure, we take implied volatilities of call options that have a delta of 0.5 and 30 days to maturity. As a second measure, the estimate an EGARCH(1,1) model with daily stock returns over a rolling window of the past twelve months. For both measures of volatility, we calculate the return of volatility as $\frac{\Delta\sigma}{\sigma} = \frac{\sigma_t - \sigma_{t-1}}{\sigma_{t-1}}$, where σ_t is the volatility on day t. Subsequently, we calculate for

each measure a monthly volatility-of-volatility estimate as the standard deviatio of the daily percentage in volatility. Next, we rank stocks based on the two measures. Finally, we compute ***volunc*** as the average of the ranking percentile of the two individual volatility-of-volatility measures. Note that Cao et al. (2019) includes a third volatility-of-volatility measure based on realized variance based on intraday data.

21. **Atm iv volatility (*ivvol*).** Following Baltussen et al. (2018), volatility of atm implied volatility scaled by average implied volatility, that is

$$VOV_t^{1M} = \frac{\sqrt{\frac{1}{20}\sum_{j=t-19}^{t}\left(\sigma_j^{IV} - \bar{\sigma}_t^{IV}\right)^2}}{\bar{\sigma}_t^{IV}},$$

where $\bar{\sigma}_t^{IV} = (1/20)\sum_{j=t-19}^{t}\sigma_j^{IV}$, and σ_j^{IV} is implied volatility.

22. **Variance spread (*ivrv*).** Following Bali and Hovakimian (2009), the realized-implied volatility spread, defined as

$$ivrv = RVol - IVol,$$

where $RVol$ is the realized volatility over month t and $IVol$ is the average volatility implied by atm call and atm put options observed at the end of month t.

23. **Variance spread (*ivrv_ratio*).** The realized-implied volatility ratio, defined as

$$ivrv = \frac{IVol}{RVol},$$

where $RVol$ is the realized volatility over month t and $IVol$ is the average volatility implied by atm call and atm put options observed at the end of month t.

24. **Near-the-money call minus put implied volatility (*civpiv*).** Following Bali and Hovakimian (2009), the implied volatility spread of call and put options, defined as

$$civpiv = CVol - PVol,$$

where $CVol$ and $PVol$ denote call and put near-the-money implied volatility, respectively.

25. **Atm call minus put implied volatility based on implied volatility surface**

data (*atm_civpiv*). The implied volatility spread of call and put options, defined as

$$atm_civpiv = CVol - PVol,$$

where $CVol$ and $PVol$ denote call and put atm implied volatility based on implied volatility surface data, respectively.

26. **Change in atm call IV (*dciv*)**. Following An et al. (2014), the change in the implied volatility at-the-money call options, defined as

$$dciv = CVol_t - CVol_{t-1},$$

where $CVol_t$ denotes month-t call implied volatility based on implied volatility surface data.

27. **Change in atm put IV (*dpiv*)**. Following An et al. (2014), the change in the implied volatility at-the-money put options, defined as

$$dpiv = PVol_t - PVol_{t-1},$$

where $PVol_t$ denotes month-t put implied volatility based on implied volatility surface data.

28. **Change in atm put minus call IV (*atm_dcivpiv*)**. Following An et al. (2014), the change in the implied volatility spread of call and put options, defined as

$$atm_dcivpiv = (CVol_t - CVol_{t-1}) - (PVol_t - PVol_{t-1}),$$

where $CVol_t$ and $PVol_t$ denote month-t call and put atm implied volatility based on implied volatility surface data, respectively.

29. **IV skew (*skewiv*)**. Following Xing et al. (2010), an implied volatility smirk measure as the difference between the implied volatilities of otm puts and atm calls, denoted by VOL^{OTMP} and VOL^{ATMC}, respectively, that is

$$skewiv = VOL^{OTMP} - VOL^{ATMC}.$$

We compute monthly ***skewiv*** by averaging over daily ***skewiv***.

30. **Weighted put-call spread (*vs_level*)**. Following Cremers and Weinbaum (2010), the call-put spread is

$$VS_t = IV_t^{calls} - IV_t^{puts} = \sum_{j=1}^{N_t} w_{j,t} \left(IV_{j,t}^{call} - IV_{j,t}^{put} \right),$$

where j denotes pairs of put and call options with the same strike and time to maturity, $w_{j,t}$ are weights, N_t denotes the number of valid pairs of options on day t, and $IV_{j,t}$ denotes Black and Scholes (1973) implied volatility. Average open interest in the call and puts are used as weights.

31. **Change in weighted put-call spread (*vs_change*)**. Following Cremers and Weinbaum (2010), we compute changes in ***vs_level***.

32. **Put-call parity violations (*pcpv*)**. We follow Ofek et al. (2004) and record violations of put-call-parity via the midpoints of option quotes and the closing price of the stock. Precisely, ***pcpv*** is given as

$$pcpv = 100 \log \left(\frac{S}{S^*} \right),$$

where S denotes the stock price, and $S^* = PV(K) + C - P$, where $PV(K)$ is the present value of the strike price K, and C and P denote the prices of a call and put option, respectively. Ofek et al. (2004) focus on ATM and intermediate maturities (i.e. between 91 and 182 days). As the authors filter for dividends and we want to exclude as little stocks as possible, we focus on ATM and short-term maturities, which are also studied by Ofek et al. (2004), but in a sub-analysis. We compute an average over the previous month.

33. **Implied shorting fees in options (*shrtfee*).** Muravyev, Pearson, and Pollet (2021) propose an option-based shorting fee measure as

$$h_Q^{imp} = \frac{1}{\delta} \left(1 - \left(1 - \frac{S_t - C_t + P_t - PV(D) - PV(K)}{S_t} \right)^{1/k} \right),$$

where C_t and P_t are the midpoints of quoted call and put quote prices, $PV(D)$ is the present value of dividends with ex-dividend dates before the expiration date, k denotes the time to expiration, S_t is the current stock price and K is the strike

price, and δ is the one-day discount factor. We take the median of the implied borrowing fees from put-call pairs.

34. **Implied volatility duration (*ivd*).** Measure for the expected timeliness of the resolution of uncertainty, following Schlag, Thimme, and Weber (2020). It is defined as

$$ivd = \sum_{j=1}^{J} \frac{\Delta IV_j^2}{\sum_j \Delta IV_j^2} \times \tau_j,$$

where $\Delta IV_j^2 = IV_{j,\tau_j}^2 - IV_{j,\tau_{j-1}}^2$ is the difference between the non-annualized squared IVs for all options at τ_j and those at τ_{j-1}, and $(\tau_1, \ldots, \tau_8) = (30, 60, 91, 122, 152, 182, 273, 365)$ (days).

C.2 Bucket-Level

For each bucket-underlying stock combination, we first calculate open-interest weighted average returns, mid prices and implied volatilities at the daily frequency. Open-interest, mid prices, and implied volatilities are obtained from OptionMetrics.

35. **Illiquidity (*illiq*)** Following Bao, Pan, and Wang (2011) in case of corporate bond markets, we construct an illiquidity measure, which aims to extract the transitory component from option prices. Precisely, let $\Delta p_{itd} = p_{itd} - p_{itd-1}$ be the log price change for option i on day d of month t. Then,

$$\text{illiq} = -COV_t(\Delta p_{itd}, \Delta p_{itd+1}).$$

36. **Roll's daily measure of illiquidity *roll*)** As an alternative measure of option-level illiquidity using daily option returns, the Roll (1984) measure is defined as,

$$\text{roll} = \begin{cases} 2\sqrt{-cov(r_d, r_{d-1})}, & \text{if } cov(r_d, r_{d-1}) < 0 \\ 0, & \text{otherwise,} \end{cases}$$

where r_d is the option return on day d.

37. **Illiquidity measure based on zero returns (*pzeros*)** As in Lesmond, Ogden, and Trzcinka (1999), we take the proportion of zero return days as a measure of

liquidity. We compute their measure on a monthly basis as

$$\text{pzeroes} = \frac{\text{\# of zero return days}}{T},$$

where T denotes the number of days in a month.

38. **Modified illiquidity measure based on zero returns (*pfht*)** Fong, Holden, and Trzcinka (2017) propose a modified version of Lesmond et al. (1999), given as

$$\text{pfht} = 2 \times \sigma \times \Phi^{-1}\left(\frac{1 + \text{pzeros}}{2}\right),$$

where σ denotes the volatility of an option contract and Φ is the cumulative standard normal distribution.

39. **Amihud measure of illiquidity (*amihud*)** Following Amihud (2002), the measure aims at capturing the price impact and is defined as

$$\text{amihud} = \frac{1}{N}\sum_{d=1}^{N}\frac{|r_d|}{Q_d},$$

where N is the number of positive-volume days in a given month, r_d the daily return, and Q_d the trading volume on day d.

40. **An extended Roll's measure (*piroll*)** Goyenko, Holden, and Trzcinka (2009) motivate an extended transaction cost proxy measure, which is defined for every transaction cost proxy *tcp* and average daily dollar volume $\bar{Q}$ in the period under observation as

$$\text{piroll} = \frac{\text{tcp}}{\bar{Q}}$$

where we subsitute *tcp* by *roll*.

41. **An extended FHT measure based on zero returns (*pifht*)**

$$\text{pifht} = \frac{\text{pfht}}{\bar{Q}},$$

with pfht being the modified illiquidity measure based on zero returns (Fong et al., 2017), and $\bar{Q}$ is the average daily dollar volume in the period under observation.

42. **Std.dev of the Amihud measure (*stdamihud*)** The standard deviation of the daily Amihud (2002) measure within a month.

43. **Pastor and Stambaugh's liquidity measure (*gammaps*)** Pástor and Stambaugh (2003) introduce a measure for the price impac based on price reversals for the equity market. It is given by γ in the following regressions:

$$r^e_{t+1} = \theta + \psi \times r_t + \gamma \times sign(r^e_t) \times Q_t + \epsilon_t,$$

where r^e_t denotes the asset's e excess return over a market index, r_t is the asset's return, Q_t the trading volume at day t. We choose the risk-free as the market index and set $gammaps = -\gamma$.

44. **Volatility (*hvol*).** Historical volatility is estimated using daily data over the last month.

45. **Skewness (*hskew*).** Historical skewness is estimated using daily data over the last month.

46. **Kurtosis (*hkurt*).** Historical kurtosis is estimated using daily data over the last month.

47. **Disposition effect (*ocgo*).** We follow Bergsma, Fodor, and Tedford (2020) in our definition of the disposition effect in options markets. Let O_t denote the price of an options contract, V_t the option turnover as the daily volume divided by open interest. We calculate the return R_t as

$$R_t = \sum_{n=1}^{20} \left(V_{t-n} \prod_{\tau=1}^{n-1} [1 - V_{t-n+\tau}] \right) O_{t-n},$$

then our measure is defined as

$$\text{ocgo} = -\frac{O_{t-2} - R_{t-1}}{O_{t-2}}.$$

48. **Open interest vs. stock volume (*oistock*).** As a measure of demand, we compute the ratio of open interest to underlying stock volume.

49. **Volume (*bucket_vol*).** The options volume as the sum of the volume of all options contracts within the bucket.

50. **Dollar volume (*bucket_dvol*).** The options dollar volume as the sum of the dollar volume of all options contracts within the bucket.

51. **Relative volume (*bucket_vol_share*).** ***bucket_vol*** divided by the options volume of all options contracts for the same underlying.

52. **Turnover (*turnover*).** The ratio of options volume to options open interest.

53. **Implied volatility rank vs. last year (*iv_rank*).** Heston and Li (2020) and Jones, Khorram, and Mo (2020) document momentum and reversal in option returns. Though Heston and Li (2020) and Jones et al. (2020) use options returns in their analyses, both consider positions in options with exactly 28 days to expiration, held from one expiration day to the subsequent expiration day in the next month. Moreover, both analyses yield one observation for each stock-month combination. As our sample allows for more than one observation for each stock-month combination, we aim at measuring momentum or reversal on the bucket level by means of implied volatilities. Precisely, we use the rank of time t implied volatility with respect to implied volatility over the last year at the daily frequency, normalized by the maximum rank over the last year.

C.3 Contract-Level

Data on contract level are obtained from OptionMetrics.

54. **Call indicator (*C*).** Indicator equalling 1 if option is a call option, 0 otherwise.

55. **Put indicator (*P*).** Indicator equalling 1 if option is a put option, 0 otherwise.

56. **Expiration flag (expiration_month).** Indicator equalling 1 if the option expires within the observation month, 0 otherwise.

57. **Time-to-maturity (*ttm*).** The number of calendar years to maturity.

58. **Moneyness (*moneyness*).** The moneyness of the option contract, measured as

$$m = \frac{K}{S},$$

where K denotes the strike price of the option contract and S the spot price of the underlying stock.

59. **Implied volatility (*iv*).** Following Büchner and Kelly (2022), the Black and Scholes (1973) implied volatility of the option contract.

60. **Delta (*delta*).** Following Büchner and Kelly (2022), the Black and Scholes (1973) delta of the option contract, i.e., the sensitivity of the option with respect to point-changes in the underlying.

61. **Gamma (*gamma*).** Following Büchner and Kelly (2022), the Black and Scholes (1973) gamma of the option contract, i.e., the sensitivity of Δ with respect to changes in the underlying. We multiply *gamma* by the price of the underlying stock divided by 100 to make it comparable in the cross-section.

62. **Theta (*theta*).** Following Büchner and Kelly (2022), the Black and Scholes (1973) theta, i.e., the time-decay of the option value. We scale *theta* by the price of the underlying stock to make it comparable in the cross-section.

63. **Vega (*vega*).** Following Büchner and Kelly (2022), the Black and Scholes (1973) vega, i.e., the sensitivity of the option with respect to changes in the implied volatility. We scale *vega* by the price of the underlying stock to make it comparable in the cross-section.

64. **Volga (*volga*).** Following Büchner and Kelly (2022), the Black and Scholes (1973) volga, i.e, the sensitivity of vega with respect to changes in the implied volatility. We scale *volga* by the price of the underlying stock to make it comparable in the cross-section.

65. **Embedded leverage (*embedlev*).** Following Karakaya (2014), the embedded leverage of the option contract, defined as

$$\Omega = \frac{S}{O} \times |\Delta|,$$

where S denotes the stock price, O the options price and Δ the delta of the option.

66. **Open interest (*oi*).** The open interest of the option contract.

67. **Dollar open interest (*doi*).** The dollar open interest of the option contract.

68. **Mid price (*mid*).** The mid price of the option, defined as

$$\frac{O_{bid} + O_{ask}}{2},$$

where O_{ask} and O_{bid} denote the ask and bid price of the option, respectively.

69. **Bid-ask spread (*optspread*)**. The bid-ask spread of the option contract, measured as

$$\frac{2 \times (O_{bid} - O_{ask})}{O_{bid} + O_{ask}},$$

where O_{ask} and O_{bid} denote the ask and bid price of the option, respectively.

Online Appendix D Classification of Characteristics

In this section we provide a detailed summary of the characteristics used in our analyses. We consider a total of 270 characteristics, of which 77 are derived from option-based information and the remainder from stock-based information. This information source is provided in the table below. We further provide the instrument level the respective characteristic relates to. Here, we consider three different levels. "Contract" information relate to a single option contract. Examples for this are the open interest or the option's delta. Since flow-based measures cannot be estimated for individual option contracts due to migrating moneyness and fleeting time-to-maturity, we construct various option buckets, outlined in Section 4.3. The characteristics on this level are denoted by instrument level "Bucket". The final group is that of characteristics operating on the level of the "Underlying" stock.

As a last set of information we also group the characteristics into 12 groups. Group "Accruals" contains five characteristics, "Contract" six, "Frictions" contains four characteristics, "Illiquidity" 28, "Industry" 90, "Informed Trading" 17, "Investment" 11, "Past Prices" 13, "Profitability" 16, "Quality" 29, "Risk" 40, and "Value" 10. The grouping for stock-based characteristics follows the intuition formed by Green et al. (2017) and Jensen et al. (2021). We group the remaining option-based characteristics accordingly.

Table D.1: Classification of Characteristics

Feature	Description	Source	Information Source	Instrument Level	Group
absacc	Absolute accruals	Green et al. (2017)	Underlying	Underlying	Accruals
acc	Working capital accruals	Green et al. (2017)	Underlying	Underlying	Accruals
aeavol	Abnormal earnings announcement volume	Green et al. (2017)	Underlying	Underlying	Profitability
age	# years since first Compustat coverage	Green et al. (2017)	Underlying	Underlying	Quality
agr	Asset growth	Green et al. (2017)	Underlying	Underlying	Investment
ailliq	Absolute illiquidity	Cao and Wei (2010)	Options	Underlying	Illiquidity
amihud	Amihud illiquidity per bucket	Amihud (2002)	Options	Bucket	Illiquidity
atm_civpiv	At-the-money put vs. call implied volatility		Options	Underlying	Informed Trading
atm_dcivpiv	Change in atm put vs. call implied volatility	An et al. (2014)	Options	Underlying	Informed Trading
baspread	Bid-ask spread	Green et al. (2017)	Underlying	Underlying	Illiquidity
bear_beta	Bear beta	Lu and Murray (2019)	Underlying	Underlying	Risk
beta	Beta	Green et al. (2017)	Underlying	Underlying	Risk
betasq	Beta squared	Green et al. (2017)	Underlying	Underlying	Risk
bm	Book-to-market	Green et al. (2017)	Underlying	Underlying	Value
bm_ia	Industry-adjusted book-to-market	Green et al. (2017)	Underlying	Underlying	Value
bucket_dvol	Option bucket dollar volume		Options	Bucket	Illiquidity
bucket_vol	Option bucket volume		Options	Bucket	Illiquidity
bucket_vol_share	Relative option bucket volume		Options	Bucket	Illiquidity
C	Call indicator		Options	Contract	Contract
cash	Cash holdings	Green et al. (2017)	Underlying	Underlying	Quality
cashdebt	Cash flow to debt	Green et al. (2017)	Underlying	Underlying	Value
cashpr	Cash productivity	Green et al. (2017)	Underlying	Underlying	Profitability
cfp	Cash-flow-to-price ratio	Green et al. (2017)	Underlying	Underlying	Risk
cfp_ia	Industry-adjusted cash-flow-to-price ratio	Green et al. (2017)	Underlying	Underlying	Risk
chatoia	Industry-adjusted change in asset turnover	Green et al. (2017)	Underlying	Underlying	Quality
chcsho	Change in shares outstanding	Green et al. (2017)	Underlying	Underlying	Investment
chempia	Industry-adjusted change in employees	Green et al. (2017)	Underlying	Underlying	Investment
chinv	Change in inventory	Green et al. (2017)	Underlying	Underlying	Investment
chmom	Change in 6-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
chpmia	Industry-adjusted change in profit margin	Green et al. (2017)	Underlying	Underlying	Profitability
chtx	Change in tax expense	Green et al. (2017)	Underlying	Underlying	Quality

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
cinvest	Corporate investment	Green et al. (2017)	Underlying	Underlying	Investment
civpiv	Near-the-money put vs. call implied volatility	Bali and Hovakimian (2009)	Options	Underlying	Informed Trading
close	Close price	Eisdorfer et al. (2020)	Underlying	Underlying	Informed Trading
convind	Convertible debt indicator	Green et al. (2017)	Underlying	Underlying	Risk
currat	Current ratio	Green et al. (2017)	Underlying	Underlying	Accruals
defrisk	Default risk	Vasquez and Xiao (2021)	Underlying	Underlying	Risk
dciv	Change in atm call implied volatility	An et al. (2014)	Options	Underlying	Informed Trading
delta	Delta	Büchner and Kelly (2022)	Options	Contract	Risk
depr	Depreciation / PP&E	Green et al. (2017)	Underlying	Underlying	Investment
divi	Dividend initiation	Green et al. (2017)	Underlying	Underlying	Value
divo	Dividend omission	Green et al. (2017)	Underlying	Underlying	Value
dolvol	Dollar trading volume	Green et al. (2017)	Underlying	Underlying	Profitability
doi	Dollar open interest		Options	Contract	Illiquidity
dpiv	Change in atm put implied volatility	An et al. (2014)	Options	Underlying	Informed Trading
dso	Stock vs. option volume in USD	Roll et al. (2010)	Options	Underlying	Informed Trading
dvol	Dollar trading volume	Cao and Wei (2010)	Options	Underlying	Illiquidity
dy	Dividend to price	Green et al. (2017)	Underlying	Underlying	Value
ear	Earnings announcement return	Green et al. (2017)	Underlying	Underlying	Profitability
egr	Growth in common shareholder equity	Green et al. (2017)	Underlying	Underlying	Investment
embedlev	Embedded Leverage	Karakaya (2014)	Options	Contract	Risk
ep	Earnings to price	Green et al. (2017)	Underlying	Underlying	Value
expiration_month	Expiration month indicator		Options	Contract	Informed Trading
fric	Contribution of market frictions to expected returns	Hiraki and Skiadopoulos (2020)	Options	Underlying	Frictions
gamma	Gamma	Büchner and Kelly (2022)	Options	Contract	Risk
gammaps	Pastor and Stambaugh liquidity measure	Pástor and Stambaugh (2003)	Options	Bucket	Illiquidity
gma	Gross profitability	Green et al. (2017)	Underlying	Underlying	Quality
grcapx	Growth in capital expenditures	Green et al. (2017)	Underlying	Underlying	Profitability
grltnoa	Growth in long-term net operating assets	Green et al. (2017)	Underlying	Underlying	Profitability
herf	Industry sales concentration	Green et al. (2017)	Underlying	Underlying	Quality
hire	Employee growth rate	Green et al. (2017)	Underlying	Underlying	Profitability
hkurt	Historic kurtosis		Options	Bucket	Risk
hskew	Historic skewness		Options	Bucket	Risk
hvol	Historic Volatility		Options	Bucket	Risk

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
idiovol	Idiosyncratic return volatility	Green et al. (2017)	Underlying	Underlying	Risk
ill	Amihud Illiquidity	Green et al. (2017)	Underlying	Underlying	Illiquidity
illiq	Illiquidity	Bao et al. (2011)	Options	Bucket	Illiquidity
ind_10	Industry code		Underlying	Underlying	Industry
ind_11	Industry code		Underlying	Underlying	Industry
ind_12	Industry code		Underlying	Underlying	Industry
ind_13	Industry code		Underlying	Underlying	Industry
ind_14	Industry code		Underlying	Underlying	Industry
ind_15	Industry code		Underlying	Underlying	Industry
ind_16	Industry code		Underlying	Underlying	Industry
ind_17	Industry code		Underlying	Underlying	Industry
ind_18	Industry code		Underlying	Underlying	Industry
ind_19	Industry code		Underlying	Underlying	Industry
ind_20	Industry code		Underlying	Underlying	Industry
ind_21	Industry code		Underlying	Underlying	Industry
ind_22	Industry code		Underlying	Underlying	Industry
ind_23	Industry code		Underlying	Underlying	Industry
ind_24	Industry code		Underlying	Underlying	Industry
ind_25	Industry code		Underlying	Underlying	Industry
ind_26	Industry code		Underlying	Underlying	Industry
ind_27	Industry code		Underlying	Underlying	Industry
ind_28	Industry code		Underlying	Underlying	Industry
ind_29	Industry code		Underlying	Underlying	Industry
ind_30	Industry code		Underlying	Underlying	Industry
ind_31	Industry code		Underlying	Underlying	Industry
ind_32	Industry code		Underlying	Underlying	Industry
ind_33	Industry code		Underlying	Underlying	Industry
ind_34	Industry code		Underlying	Underlying	Industry
ind_35	Industry code		Underlying	Underlying	Industry
ind_36	Industry code		Underlying	Underlying	Industry
ind_37	Industry code		Underlying	Underlying	Industry
ind_38	Industry code		Underlying	Underlying	Industry
ind_39	Industry code		Underlying	Underlying	Industry

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ind_40	Industry code		Underlying	Underlying	Industry
ind_41	Industry code		Underlying	Underlying	Industry
ind_42	Industry code		Underlying	Underlying	Industry
ind_43	Industry code		Underlying	Underlying	Industry
ind_44	Industry code		Underlying	Underlying	Industry
ind_45	Industry code		Underlying	Underlying	Industry
ind_46	Industry code		Underlying	Underlying	Industry
ind_47	Industry code		Underlying	Underlying	Industry
ind_48	Industry code		Underlying	Underlying	Industry
ind_49	Industry code		Underlying	Underlying	Industry
ind_50	Industry code		Underlying	Underlying	Industry
ind_51	Industry code		Underlying	Underlying	Industry
ind_52	Industry code		Underlying	Underlying	Industry
ind_53	Industry code		Underlying	Underlying	Industry
ind_54	Industry code		Underlying	Underlying	Industry
ind_55	Industry code		Underlying	Underlying	Industry
ind_56	Industry code		Underlying	Underlying	Industry
ind_57	Industry code		Underlying	Underlying	Industry
ind_58	Industry code		Underlying	Underlying	Industry
ind_59	Industry code		Underlying	Underlying	Industry
ind_60	Industry code		Underlying	Underlying	Industry
ind_61	Industry code		Underlying	Underlying	Industry
ind_62	Industry code		Underlying	Underlying	Industry
ind_63	Industry code		Underlying	Underlying	Industry
ind_64	Industry code		Underlying	Underlying	Industry
ind_65	Industry code		Underlying	Underlying	Industry
ind_66	Industry code		Underlying	Underlying	Industry
ind_67	Industry code		Underlying	Underlying	Industry
ind_68	Industry code		Underlying	Underlying	Industry
ind_69	Industry code		Underlying	Underlying	Industry
ind_70	Industry code		Underlying	Underlying	Industry
ind_71	Industry code		Underlying	Underlying	Industry
ind_72	Industry code		Underlying	Underlying	Industry

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ind_73	Industry code		Underlying	Underlying	Industry
ind_74	Industry code		Underlying	Underlying	Industry
ind_75	Industry code		Underlying	Underlying	Industry
ind_76	Industry code		Underlying	Underlying	Industry
ind_77	Industry code		Underlying	Underlying	Industry
ind_78	Industry code		Underlying	Underlying	Industry
ind_79	Industry code		Underlying	Underlying	Industry
ind_80	Industry code		Underlying	Underlying	Industry
ind_81	Industry code		Underlying	Underlying	Industry
ind_82	Industry code		Underlying	Underlying	Industry
ind_83	Industry code		Underlying	Underlying	Industry
ind_84	Industry code		Underlying	Underlying	Industry
ind_85	Industry code		Underlying	Underlying	Industry
ind_86	Industry code		Underlying	Underlying	Industry
ind_87	Industry code		Underlying	Underlying	Industry
ind_88	Industry code		Underlying	Underlying	Industry
ind_89	Industry code		Underlying	Underlying	Industry
ind_90	Industry code		Underlying	Underlying	Industry
ind_91	Industry code		Underlying	Underlying	Industry
ind_92	Industry code		Underlying	Underlying	Industry
ind_93	Industry code		Underlying	Underlying	Industry
ind_94	Industry code		Underlying	Underlying	Industry
ind_95	Industry code		Underlying	Underlying	Industry
ind_96	Industry code		Underlying	Underlying	Industry
ind_97	Industry code		Underlying	Underlying	Industry
ind_98	Industry code		Underlying	Underlying	Industry
ind_99	Industry code		Underlying	Underlying	Industry
indmom	Industry momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
invest	Capital expenditures and inventory	Green et al. (2017)	Underlying	Underlying	Investment
iv	Implied volatility	Büchner and Kelly (2022)	Options	Contract	Contract
iv_rank	Implied volatility rank vs. last year		Options	Bucket	Past Prices
ivarud30	Option implied variance asymmetry	Huang et al. (2019)	Options	Underlying	Risk
ivd	Implied volatility duration	Schlag et al. (2020)	Options	Underlying	Risk

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ivrv	Implied volatility minus realized volatility	Bali and Hovakimian (2009)	Options	Underlying	Risk
ivrv_ratio	Implied volatility minus realized volatility ratio		Options	Underlying	Risk
ivslope	Implied volatility slope	Vasquez (2017)	Options	Underlying	Risk
ivvol	Volatility of atm volatility	Baltussen et al. (2018)	Options	Underlying	Risk
ldso	Log changes in the stock to option volume	Roll et al. (2010)	Options	Underlying	Informed Trading
lev	Leverage	Green et al. (2017)	Underlying	Underlying	Quality
lgr	Growth in long-term debt	Green et al. (2017)	Underlying	Underlying	Quality
lso	Log of stock vs. option volume	Roll et al. (2010)	Options	Underlying	Informed Trading
maxret	Maximum daily return	Green et al. (2017)	Underlying	Underlying	Risk
mid	Option mid price		Options	Contract	Contract
modos	Modified stock vs. option volume	Johnson and So (2012)	Options	Underlying	Informed Trading
mom12m	12-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
mom1m	1-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
mom36m	36-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
mom6m	6-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
moneyness	Moneyness		Options	Contract	Contract
ms	Financial statement score	Green et al. (2017)	Underlying	Underlying	Quality
mve	Size	Green et al. (2017)	Underlying	Underlying	Quality
mve_ia	Industry-adjusted size	Green et al. (2017)	Underlying	Underlying	Quality
nincr	Number of earnings increases	Green et al. (2017)	Underlying	Underlying	Quality
nopt	Number of options trading		Options	Underlying	Illiquidity
ocgo	Disposition Effect	Bergsma et al. (2020)	Options	Bucket	Past Prices
oi	Open interest		Options	Contract	Illiquidity
oistock	Open interest vs. stock volume		Options	Bucket	Informed Trading
operprof	Operating profitability	Green et al. (2017)	Underlying	Underlying	Quality
optspread	Option bid-ask spread		Options	Contract	Illiquidity
orgcap	Organization capital	Green et al. (2017)	Underlying	Underlying	Quality
P	Put-flag		Options	Contract	Contract
pba	Proportional bid-ask spread	Cao and Wei (2010)	Options	Underlying	Illiquidity
pchcapx_ia	Industry-adjusted % change in capital expenditures	Green et al. (2017)	Underlying	Underlying	Investment
pchcurrat	% change in current ratio	Green et al. (2017)	Underlying	Underlying	Quality
pchdepr	% change in depreciation	Green et al. (2017)	Underlying	Underlying	Quality
pchgm_pchsale	% change in gross margin - % change in sales	Green et al. (2017)	Underlying	Underlying	Quality

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
pchquick	% change in quick ratio	Green et al. (2017)	Underlying	Underlying	Quality
pchsale_pchinvt	% change in sales - % change in inventory	Green et al. (2017)	Underlying	Underlying	Profitability
pchsale_pchrect	% change in sales - % change in A/R	Green et al. (2017)	Underlying	Underlying	Profitability
pchsale_pchxsga	% change in sales - % change in SG&A	Green et al. (2017)	Underlying	Underlying	Profitability
pchsaleinv	% change in sales-to-inventory	Green et al. (2017)	Underlying	Underlying	Profitability
pcpv	Put-call parity deviations	Ofek et al. (2004)	Options	Underlying	Frictions
pcratio	Put-call ratio	Blau et al. (2014)	Options	Underlying	Informed Trading
pctacc	Percent accruals	Green et al. (2017)	Underlying	Underlying	Accruals
pfht	Modified illiquidity measure based on zero returns	Fong et al. (2017)	Options	Bucket	Illiquidity
pifht	An extended FHT measured based on zero returns		Options	Bucket	Illiquidity
pilliq	Percentage illiquidity	Cao and Wei (2010)	Options	Underlying	Illiquidity
piroll	Extended Roll's measure	Goyenko et al. (2009)	Options	Bucket	Illiquidity
pricedelay	Price delay	Green et al. (2017)	Underlying	Underlying	Illiquidity
ps	Financial statements score (Piotroski)	Green et al. (2017)	Underlying	Underlying	Quality
pzeros	Illiquidity measure based on zero returns	Lesmond et al. (1999)	Options	Bucket	Illiquidity
quick	Quick ratio	Green et al. (2017)	Underlying	Underlying	Quality
rd	R&D increase	Green et al. (2017)	Underlying	Underlying	Investment
rd_mve	R&D to market capitalization	Green et al. (2017)	Underlying	Underlying	Quality
rd_sale	R&D to sales	Green et al. (2017)	Underlying	Underlying	Quality
realestate	Real estate holdings	Green et al. (2017)	Underlying	Underlying	Quality
retvol	Return volatility	Green et al. (2017)	Underlying	Underlying	Risk
rnk182	182-day risk-neutral kurtosis		Options	Underlying	Risk
rnk273	273-day risk-neutral kurtosis		Options	Underlying	Risk
rnk30	30-day risk-neutral kurtosis		Options	Underlying	Risk
rnk365	365-day risk-neutral kurtosis		Options	Underlying	Risk
rnk91	91-day risk-neutral kurtosis		Options	Underlying	Risk
rns182	182-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns273	273-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns30	30-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns365	365-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns91	91-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
roaq	Return on assets	Green et al. (2017)	Underlying	Underlying	Profitability
roavol	Earnings volatility	Green et al. (2017)	Underlying	Underlying	Quality

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
roeq	Return on equity	Green et al. (2017)	Underlying	Underlying	Profitability
roic	Return on invested capital	Green et al. (2017)	Underlying	Underlying	Profitability
roll	Roll's measure of illiquidity	Roll (1984)	Options	Bucket	Illiquidity
rsup	Revenue surprise	Green et al. (2017)	Underlying	Underlying	Profitability
rv	Realized variance	Cao et al. (2019)	Underlying	Underlying	Risk
salecash	Sales to cash	Green et al. (2017)	Underlying	Underlying	Value
saleinv	Sales to inventory	Green et al. (2017)	Underlying	Underlying	Value
salerec	Sales to receivables	Green et al. (2017)	Underlying	Underlying	Value
season1	Seasonal return - 1 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
season2	Seasonal return - 2 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
season3	Seasonal return - 3 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
season4	Seasonal return - 4 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
secured	Secured debt	Green et al. (2017)	Underlying	Underlying	Quality
securedind	Secured debt indicator	Green et al. (2017)	Underlying	Underlying	Quality
sgr	Sales growth	Green et al. (2017)	Underlying	Underlying	Investment
shrtfee	Implied shorting fees	Muravyev et al. (2021)	Options	Underlying	Frictions
sin	Sin stocks	Green et al. (2017)	Underlying	Underlying	Quality
skewiv	IV skew	Xing et al. (2010)	Options	Underlying	Informed Trading
so	Stock vs. option volume	Roll et al. (2010)	Options	Underlying	Informed Trading
sp	Sales to price	Green et al. (2017)	Underlying	Underlying	Quality
std_dolvol	Volatility of liquidity (dollar trading volume)	Green et al. (2017)	Underlying	Underlying	Illiquidity
std_turn	Volatility of liquidity (share turnover)	Green et al. (2017)	Underlying	Underlying	Illiquidity
stdacc	Accrual volatility	Green et al. (2017)	Underlying	Underlying	Accruals
stdamihud	Standard deviation of Amihud's illiquidity measure		Options	Bucket	Illiquidity
stdcf	Cash flow volatility	Green et al. (2017)	Underlying	Underlying	Risk
tang	Dept capacity/firm tangibility	Green et al. (2017)	Underlying	Underlying	Quality
tb	Tax income to book income	Green et al. (2017)	Underlying	Underlying	Past Prices
theta	Theta	Büchner and Kelly (2022)	Options	Contract	Risk
tlm30	Tail loss measure	Vilkov and Xiao (2012)	Options	Underlying	Risk
toi	Total option open interest		Options	Underlying	Illiquidity
ttm	Time-to-maturity		Options	Contract	Contract
turn	Share turnover	Green et al. (2017)	Underlying	Underlying	Risk
turnover	Option turnover		Options	Bucket	Illiquidity

Continued on Next Page

Table D.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
underlying_return	Return of the underlying		Underlying	Underlying	Frictions
vega	Vega	Büchner and Kelly (2022)	Options	Contract	Risk
vol	Trading volume in options		Options	Underlying	Illiquidity
volga	Volga	Büchner and Kelly (2022)	Options	Contract	Risk
volunc	Volatility uncertainty	Cao et al. (2019)	Options	Underlying	Risk
vs_change	Change in weighted put-call spread	Cremers and Weinbaum (2010)	Options	Underlying	Informed Trading
vs_level	Weighted put-call spread	Cremers and Weinbaum (2010)	Options	Underlying	Informed Trading
zerotrade	Zero trading days	Green et al. (2017)	Underlying	Underlying	Illiquidity

Not Continued on Next Page

The table provides a detailed summary of the characteristics. For each characteristic, the table shows the name (Feature), a description of the characteristic (Description) and the source, if applicable (Source). Furthermore, it displays if the characteristic is derived form option-based or stock-based information (Information Source), the instrument level the characteristic relates to (Instrument Level), and to which characteristic group (Group) it belongs to.

Online Appendix E Additional Summary Statistics

E.1 Summary Statistics for the Underlying Stocks

	Mean	Std	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: Time-Series Distribution							
Number of stocks in the sample each month	1747.31	142.84	1560.1	1664.0	1755.5	1850.75	1922.2
Stock coverage of stock universe (EW)	38.05	10.09	24.88	28.07	40.04	46.64	50.98
Stock coverage of stock universe (VW)	83.81	5.0	77.26	81.25	84.63	86.98	89.63
Stock traded at NYSE or AMEX	51.86	2.36	48.74	50.95	52.01	53.32	54.35
Stock already included in previous month	86.11	5.87	82.36	84.31	86.56	88.74	90.25
Panel B: Time-Series Average of Cross-Sectional Distributions							
Firm size in million	7945	27163	240	550	1516	4793	15532
Firm size CSRP percentile	72	18	45	59	75	87	94
Firm volatility CSRP percentile	45	25	11	23	44	66	81

Panel C: Time-Series Average of Industry Distribution					
FF-12 Industry	Optionable Stocks	CRSP sample	FF-12 Industry	Optionable Stocks	CRSP sample
Consumer nondurables	4.48%	4.92%	Telecom	3.47%	2.56%
Consumer durables	0.52%	0.47%	Utilities	2.74%	2.21%
Manufacturing	9.34%	8.15%	Wholesale	11.11%	8.49%
Energy	4.92%	2.68%	Healthcare	11.4%	9.24%
Chemicals	2.45%	1.59%	Finance	10.08%	24.08%
Business Equipment	20.22%	14.44%	Other	19.28%	21.16%

Table E.1: Summary Statistics of Underlying Stocks

The table reports summary statistics for the sample of underlying stocks. We compare our sample of underlying stocks with all stocks in CRSP, which have share codes 10 or 11 and exchange codes 1, 2, 3, 31, 32, 33. Panel A reports the time-series summary statistics and Panel B reports the time-series averages of the cross-sectional distribution. Percent coverage of the stock universe (EW) is the number of stocks in the sample, divided by the total number of CRSP stocks. Percent coverage of the stock universe (VW) is the total market capitalization of sample stocks divided by the total CRSP market capitalization. Percent coverage of stocks traded at NYSE or AMEX is the number of stocks in the sample trading at NYSE or AMEX, divided by the total number of stocks. The firm size percentiles are computed using the CRSP sample. Panel C reports time-series averages of industry distributions of the Fama-French 12-industry classification. The industry distributions are reported for the sample of optionable stocks as well as for the CRSP universe. The sample period is from January 1996 to December 2020.

E.2 Delta-Hedged Option Return per Bucket

Table E.2: Delta-Hedged Option Return per Bucket

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: Long Term (N=6,903,957)							
Delta-Hedged Return	0.15	5.03	-4.15	-1.84	-0.27	1.47	4.73
Days to Maturity	269.01	185.56	112.0	141.0	200.0	324.0	570.0
Moneyness	1.04	0.38	0.71	0.86	1.0	1.15	1.36
Implied volatility	45.85	23.08	23.55	29.88	40.13	55.64	75.54
Absolute Delta	0.44	0.24	0.13	0.25	0.43	0.63	0.79
Panel B: Long Term Atm (N=2,495,362)							
Delta-Hedged Return	0.2	3.63	-2.85	-1.41	-0.22	1.24	3.6
Days to Maturity	241.32	167.94	112.0	141.0	176.0	234.0	506.0
Moneyness	1.0	0.06	0.93	0.96	1.0	1.05	1.08
Implied volatility	38.96	18.75	20.94	26.02	34.42	46.64	62.97
Absolute Delta	0.5	0.12	0.33	0.4	0.5	0.6	0.66
Panel C: Long Term Itm Call (N=865,963)							
Delta-Hedged Return	0.14	3.91	-2.99	-1.25	-0.14	1.09	3.46
Days to Maturity	288.93	198.5	113.0	141.0	203.0	386.0	628.0
Moneyness	0.77	0.12	0.59	0.71	0.8	0.86	0.89
Implied volatility	52.16	22.33	29.0	36.08	47.1	63.07	82.1
Absolute Delta	0.81	0.07	0.72	0.75	0.81	0.87	0.92

Continued on Next Page

Table E.2 from previous page

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel D: Long Term Itm Put (N=639,054)							
Delta-Hedged Return	0.11	2.45	-1.98	-0.85	-0.09	0.8	2.38
Days to Maturity	280.43	192.15	113.0	141.0	203.0	355.0	598.0
Moneyness	1.42	0.73	1.13	1.16	1.25	1.44	1.79
Implied volatility	54.63	31.31	25.24	33.16	46.53	67.18	93.66
Absolute Delta	0.69	0.15	0.5	0.59	0.69	0.8	0.89
Panel E: Long Term Otm Call (N=1,528,873)							
Delta-Hedged Return	0.25	7.3	-6.89	-3.29	-0.44	2.96	8.14
Days to Maturity	285.64	193.94	113.0	143.0	204.0	381.0	603.0
Moneyness	1.36	0.35	1.13	1.17	1.26	1.42	1.67
Implied volatility	47.24	23.26	23.87	30.61	41.46	58.08	78.48
Absolute Delta	0.3	0.14	0.12	0.19	0.29	0.4	0.48
Panel F: Long Term Otm Put (N=1,374,705)							
Delta-Hedged Return	-0.03	5.58	-5.05	-2.71	-0.7	1.65	5.36
Days to Maturity	282.94	188.97	113.0	142.0	204.0	358.0	598.0
Moneyness	0.74	0.13	0.55	0.67	0.78	0.85	0.88
Implied volatility	48.75	22.56	27.2	33.19	43.05	58.12	77.45
Absolute Delta	0.16	0.08	0.05	0.1	0.16	0.23	0.27
Panel G: Short Term (N=5,225,887)							
Delta-Hedged Return	-0.21	5.36	-4.65	-2.07	-0.43	1.04	4.1
Days to Maturity	44.39	23.02	17.0	21.0	49.0	52.0	80.0
Moneyness	1.01	0.18	0.84	0.92	1.0	1.07	1.17
Implied volatility	50.08	28.44	23.33	30.7	42.74	61.37	86.07
Absolute Delta	0.48	0.26	0.14	0.26	0.47	0.69	0.84
Panel H: Short Term Atm (N=3,242,014)							
Delta-Hedged Return	-0.11	4.08	-3.67	-1.79	-0.4	0.99	3.5
Days to Maturity	40.22	22.35	17.0	19.0	46.0	51.0	79.0
Moneyness	1.0	0.05	0.93	0.96	1.0	1.04	1.07
Implied volatility	41.61	21.33	21.09	26.94	36.37	50.33	68.81
Absolute Delta	0.49	0.19	0.24	0.35	0.49	0.63	0.74

Continued on Next Page

Table E.2 from previous page

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel I: Short Term Itm Call (N=471,757)							
Delta-Hedged Return	0.19	5.14	-2.72	-1.15	-0.23	0.61	2.71
Days to Maturity	49.39	22.92	18.0	22.0	50.0	77.0	81.0
Moneyness	0.82	0.08	0.72	0.79	0.85	0.88	0.9
Implied volatility	65.11	30.15	34.89	43.82	58.16	79.04	103.74
Absolute Delta	0.85	0.07	0.75	0.8	0.85	0.9	0.94
Panel J: Short Term Itm Put (N=382,619)							
Delta-Hedged Return	-0.41	3.87	-2.59	-1.14	-0.31	0.46	1.96
Days to Maturity	47.64	23.03	18.0	22.0	50.0	77.0	80.0
Moneyness	1.29	0.34	1.12	1.14	1.19	1.31	1.53
Implied volatility	72.13	41.5	33.13	43.58	61.24	88.33	124.21
Absolute Delta	0.8	0.11	0.64	0.72	0.81	0.89	0.94
Panel K: Short Term Otm Call (N=537,456)							
Delta-Hedged Return	-0.32	9.09	-9.78	-5.0	-0.96	3.36	9.64
Days to Maturity	54.05	21.64	21.0	47.0	50.0	78.0	81.0
Moneyness	1.23	0.17	1.12	1.14	1.18	1.26	1.4
Implied volatility	61.01	30.5	30.14	38.92	53.55	76.05	101.68
Absolute Delta	0.21	0.11	0.08	0.13	0.2	0.29	0.36
Panel L: Short Term Otm Put (N=592,041)							
Delta-Hedged Return	-0.85	7.47	-7.99	-4.29	-1.43	1.42	6.3
Days to Maturity	52.39	22.06	19.0	46.0	50.0	78.0	81.0
Moneyness	0.82	0.09	0.7	0.78	0.84	0.88	0.89
Implied volatility	60.34	29.45	31.8	39.81	52.79	73.11	98.45
Absolute Delta	0.14	0.07	0.05	0.08	0.13	0.19	0.24

Not Continued on Next Page

The table reports the descriptive statistics of delta-hedged option returns for the period 1996 to 2020. Delta-hedged option returns are measured over a period of one calendar month, or until option maturity. Delta-hedging is performed daily. Days to maturity is the number of calendar days until option expiration. Moneyness is the ratio between the underlying's stock price and the option's strike price. Option implied volatility is provided by OptionMetrics. Absolute delta is the absolute value of the Black-Scholes delta. We differentiate between different parts of the time-to-maturity and moneyness domain of a single option, which we refer to as "buckets", as defined in Section 4.3. Specifically, we separately consider predictability for short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (itm: $K/S > 1.1$ for puts, $K/S < 0.9$ for calls), out-of-the-money (otm: $K/S < 0.9$ for puts, $K/S < 1.1$ for calls) calls and puts, and at-the-money options (atm: $0.9 \leq K/S \leq 1.1$), as well as time-to-maturity and moneyness combinations. Each panel shows statistics for pooled options belonging to one bucket.

E.3 Number of Options per Bucket

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Long Term	14.98	26.69	1.05	2.38	5.86	16.48	39.51
Long Term Atm	6.47	8.98	1.0	1.69	3.31	7.78	15.89
Long Term Itm Call	3.58	5.29	1.0	1.0	1.9	3.99	7.91
Long Term Itm Put	3.33	4.49	1.0	1.0	1.83	3.68	7.51
Long Term Otm Call	4.87	7.09	1.0	1.08	2.45	5.75	11.41
Long Term Otm Put	5.34	8.57	1.0	1.03	2.4	6.16	12.76
Short Term	10.93	15.89	1.0	2.4	5.83	13.35	26.48
Short Term Atm	7.38	8.84	1.0	2.28	4.44	9.36	17.38
Short Term Itm Call	2.4	2.77	1.0	1.0	1.42	2.7	4.74
Short Term Itm Put	2.43	2.79	1.0	1.0	1.37	2.68	4.9
Short Term Otm Call	2.69	3.34	1.0	1.0	1.54	3.03	5.47
Short Term Otm Put	3.08	4.17	1.0	1.0	1.74	3.48	6.42

Table E.3: Number of Options for Option Buckets

The table reports summary statistics on the number of options within certain regions of the time-to-maturity and moneyness domain, denoted by "buckets", as defined in Section 4.3. The time-to-maturity and moneyness domain is divided into short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (itm: $K/S > 1.1$ for puts, $K/S < 0.9$ for calls) out-of-the-money (otm: $K/S < 0.9$ for puts, $K/S < 1.1$ for calls) calls and puts, and at-the-money options (atm: $0.9 \leq K/S \leq 1.1$), as well as time-to-maturity and moneyness combinations. We first compute descriptive statistics for each underlying stock and subsequently take the average across all stocks in the sample. Values here correspond to the number of options *per underlying stock*. The sample period is from January 1996 to December 2020.

Online Appendix F Model Comparison – Additional Analyses

F.1 Cross-sectional DM Tests

Panel A: Diebold and Mariano (1995) Cross-Sectional Forecast Comparison

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	1.47	1.38	−0.93	**-4.86**	**3.25**	**8.94**	**5.63**	**7.45**	**10.02**	**10.23**
Lasso		−0.27	**-3.11**	**-5.21**	**3.88**	**4.53**	**3.91**	**3.55**	**5.09**	**6.10**
ENet			**-3.27**	**-4.95**	**3.77**	**4.68**	**4.14**	**3.59**	**5.19**	**6.19**
PCR				**-2.56**	**6.06**	**6.22**	**6.49**	**5.14**	**6.08**	**8.04**
PLS					**6.28**	**10.74**	**8.06**	**9.49**	**11.21**	**11.22**
L-En						**3.64**	**2.86**	**2.92**	**3.47**	**5.97**
GBR							**-3.47**	0.52	**-1.98**	**4.56**
RF								**2.18**	1.13	**8.22**
Dart									−1.36	**2.05**
FFN										**4.56**

The table shows Diebold and Mariano (1995) test statistics following Equation (8), using cross-sectional errors as inputs, for the nine models and two ensembles considered in the paper. A positive number indicates that the model in the column outperforms the row model. If it is highlighted in light blue (blue), this outperformance is statistically significant at the 1% level (5% level).

F.2 Cross-sectional Comparison between N-En and L-En

N-En beats L-En in 86.5% of the months when tasked with predicting future cross-sectional return spreads for delta-hedged single-equity options:

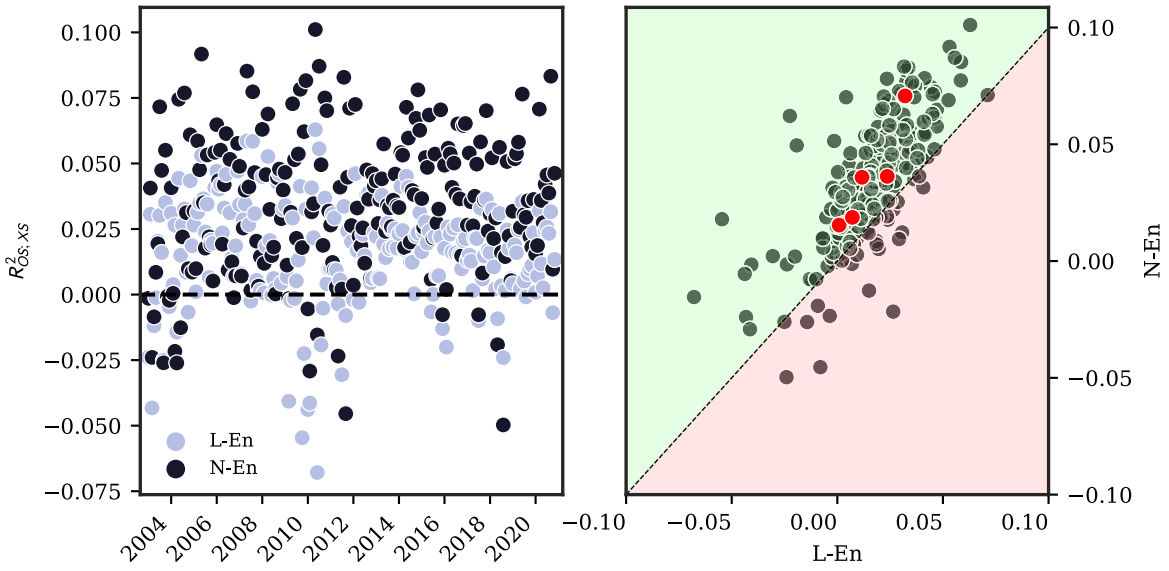


Fig. F.1. Comparing Linear and Nonlinear Ensembles – $R^2_{OS;XS}$

The left panel of the figure shows monthly cross-sectional $R^2_{OS;XS}$ for the testing sample from 2003 through 2020 for the linear (L-En) and nonlinear (N-En) ensembles. The right panel compares the two by showing the resulting $R^2_{OS;XS}$ for L-En on the x-axis and for N-En and the y axis. The green-shaded area represents a relative outperformance in terms of predictability for N-En, while the red-shaded area represents the opposite. The five red circles represent the Coronavirus selloff from December 2019 through April 2020.

F.3 Cross-sectional Comparison between N-En and L-En for Different Buckets

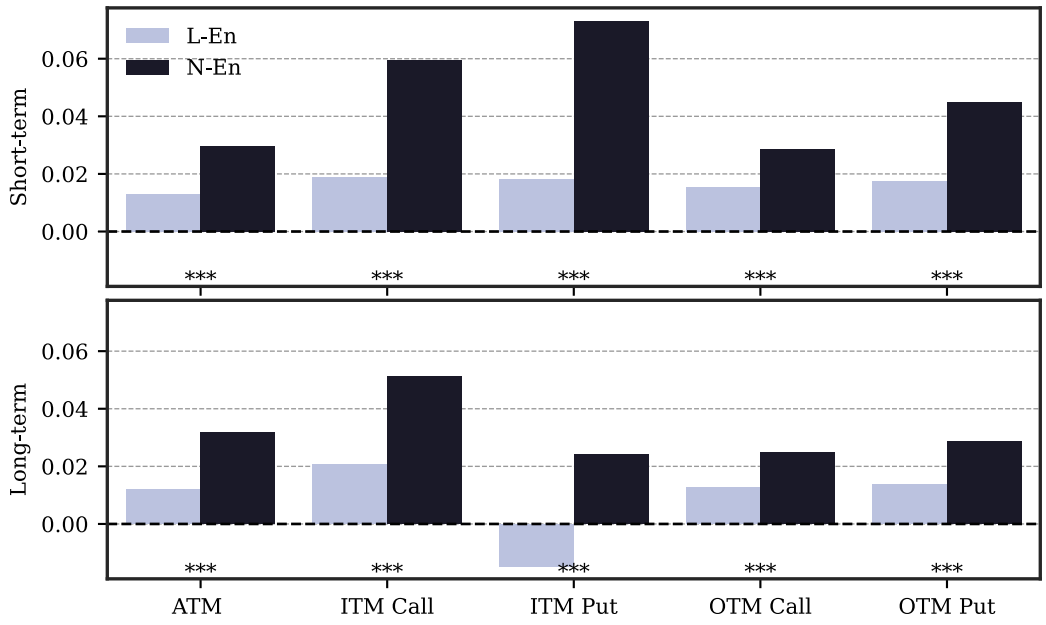


Fig. F.2. Comparing Linear and Nonlinear Ensembles – $R^2_{OS;XS}$

The figure shows monthly $R^2_{OS;XS}$ for the testing sample from 2003 through 2020 for the linear (L-En) and nonlinear (N-En) ensembles using options that fall in different moneyness and maturity buckets, as defined in Section 4.

F.4 Predicting Individual Contracts vs. Option Portfolios

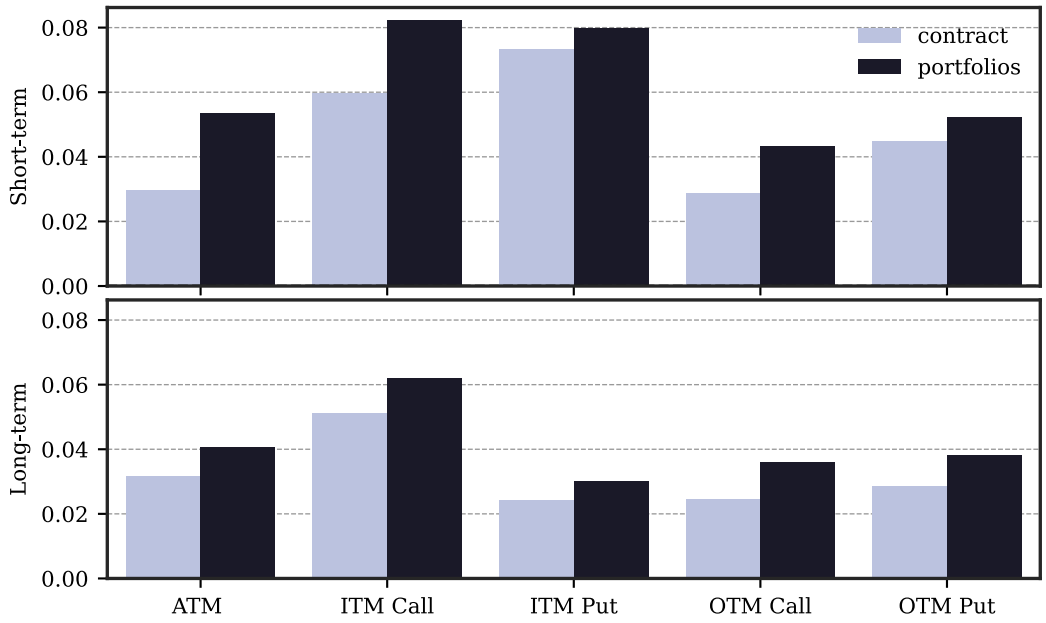


Fig. F.3. Predictive Power and Impact of Nonlinear Models for Option Portfolios

The figure shows monthly $R^2_{OS;XS}$ for the testing sample from 2003 through 2020 for the nonlinear (N-En) ensembles using option portfolios. Specifically, we form dollar open interest-weighted portfolios using all options of a given underlying that fall in a respective bucket (defined in Section 4.3) and predict future returns using N-En. This approach closely mimics recent studies on option return predictability (Cao and Han, 2013; Cao et al., 2021; Goyenko and Zhang, 2021). We compare the resulting predictability levels with those obtained when predicting the individual contracts included in each bucket.

Online Appendix G Predicting All vs. Only Short-Term At-the-Money Options

In this section of the internet appendix, we compare all nine model as well as the linear and nonlinear ensemble fit on all options, regardless of the maturity and moneyness, denoted as "Full", and fit on only short-term at-the-money options, following the definition in Section 4.3. The models are consequently denoted as "ATM". We assess in how far it pays off to focus on similar options when estimating the models, whether there are learning and regularization effects from also considering options on other parts of the implied volatility surface, and in how far nonlinearities continue to play a role when using only short-term at-the-money options to fit the models.

We find comparable levels of predictability for either model applied to their respective sample, but better performance of the specialized model "ATM" when applied only to short-term at-the-money options. This is expected, as it performs a much simpler task of focusing on a small subset of all traded options (more than 70% of the dollar open interest outstanding is found in other options). Interestingly, however, and much in favor of our estimation procedure, we find that the resulting predictive power of the models is very similar for Dart and N-En.

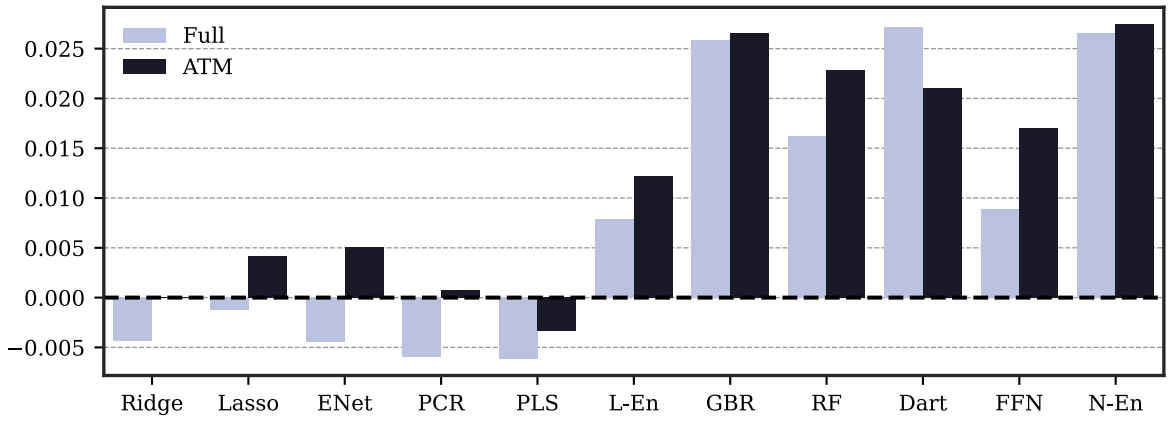


Fig. G.1. R^2_{OS} Model Comparison

The figure shows out-of-sample R^2_{OS} as defined in Equation (4) for the nine models considered, as well as the linear (L-En) and nonlinear (N-En) ensemble methods. We separately document the predictive power for a model fit on all options ("Full") and a model fit on only short-term at-the-money options ("ATM"). The predictive power is assessed on two testing samples using all or only short-term at-the-money options, respectively. The testing sample spans the years 2003 through 2020.

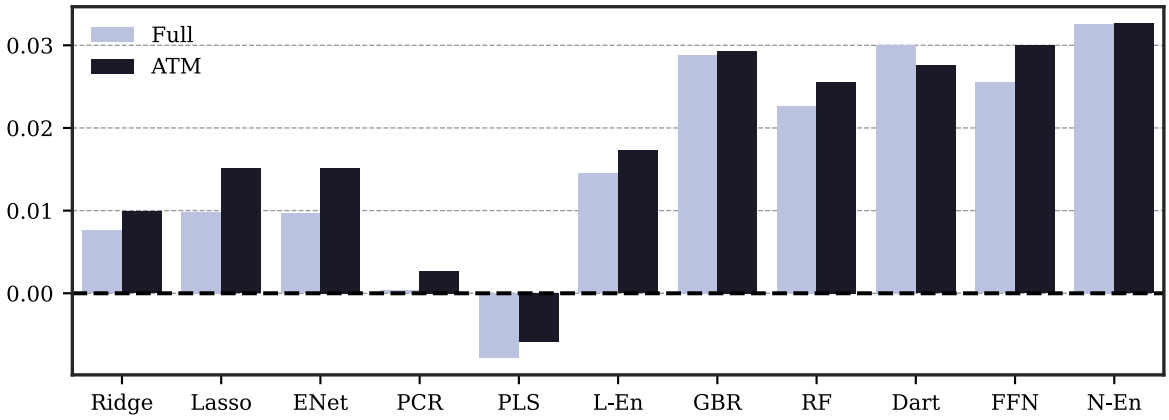


Fig. G.2. $R^2_{OS;XS}$ Model Comparison

The figure shows out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) for the nine models considered, as well as the linear (L-En) and nonlinear (N-En) ensemble methods. We separately document the predictive power for a model fit on all options ("Full") and a model fit on only short-term at-the-money options ("ATM"). The predictive power is assessed on two testing samples using all or only short-term at-the-money options, respectively. The testing sample spans the years 2003 through 2020.

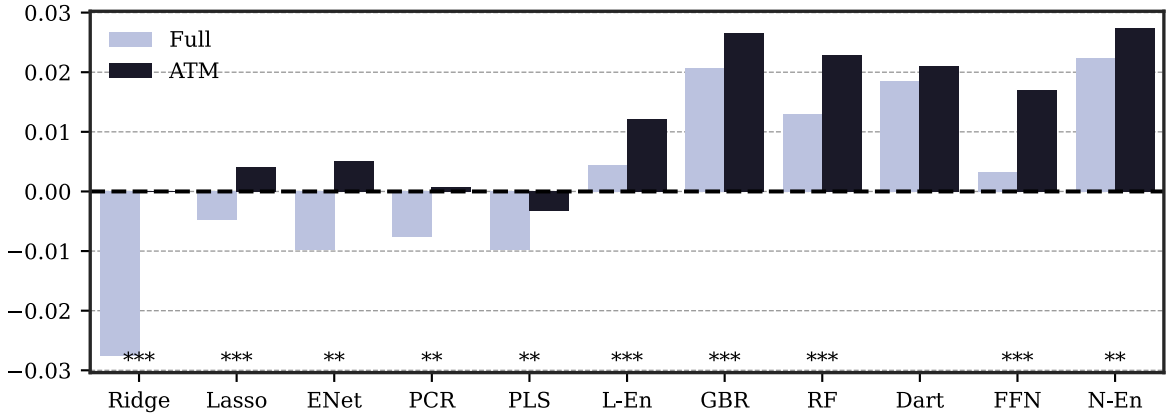


Fig. G.3. R^2_{OS} Model Comparison using only Short-Term at-the-money Options

The figure shows out-of-sample R^2_{OS} as defined in Equation (4) for the nine models considered, as well as the linear (L-En) and nonlinear (N-En) ensemble methods. We separately document the predictive power for all options and for calls and puts. ***, **, * below the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7) for the sample of "all" options. The testing sample spans the years 2003 through 2020.

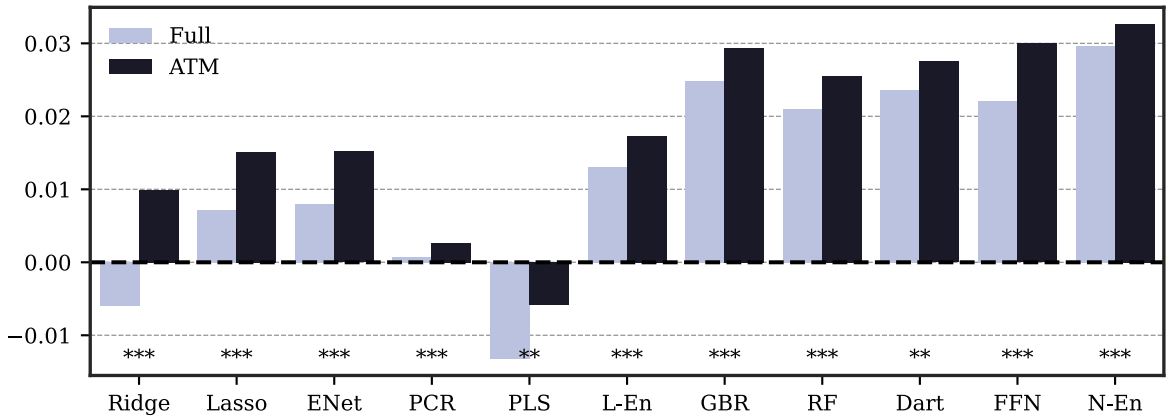


Fig. G.4. $R^2_{OS;XS}$ Model Comparison using only Short-Term at-the-money Options

The figure shows out-of-sample R^2_{OS} as defined in Equation (4) for the nine models considered, as well as the linear (L-En) and nonlinear (N-En) ensemble methods. We separately document the predictive power for all options and for calls and puts. ***, **, * below the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7) for the sample of "all" options. The testing sample spans the years 2003 through 2020.

Online Appendix H Machine Learning Portfolios – Additional Analyses

H.1 Performance for ML Portfolios using Value-weighted Returns

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.308	−0.967	1.335	−0.725	−1.612	−1.556	1.845	−0.843	***
2	−0.722	−0.482	1.450	−0.332	−0.709	−0.637	1.564	−0.408	
3	−0.491	−0.319	1.434	−0.222	−0.397	−0.393	1.393	−0.282	
4	−0.321	−0.220	1.432	−0.153	−0.223	−0.220	1.272	−0.173	
5	−0.178	−0.154	1.520	−0.101	−0.098	−0.125	1.260	−0.099	
6	−0.047	−0.096	1.488	−0.065	0.006	−0.077	1.321	−0.059	
7	0.082	−0.026	1.485	−0.017	0.109	−0.038	1.444	−0.027	
8	0.219	0.027	1.475	0.018	0.228	0.042	1.494	0.028	
9	0.385	0.084	1.450	0.058	0.393	0.135	1.607	0.084	
Hi	0.695	0.314	1.580	0.199	0.854	0.460	1.852	0.248	
H-L	2.004	1.281	1.240	1.033	2.466	2.016	1.466	1.375	***
		(11.52)		(8.64)		(14.28)		(9.57)	

Table H.1: Trading on Equally-weighted Machine Learning Predictions

The table shows returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods and weighted equally across contracts. Pred denotes the average predicted return within the respective portfolio, Avg the average realized return, SD the standard deviation of realized returns and finally SR the realized Sharpe ratio. All values are given per month. The last column (N vs. L) gives the significance of comparing the mean realized returns for N-En and L-En. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

H.2 Summary Statistics for ML Portfolios – Put and Call Composition

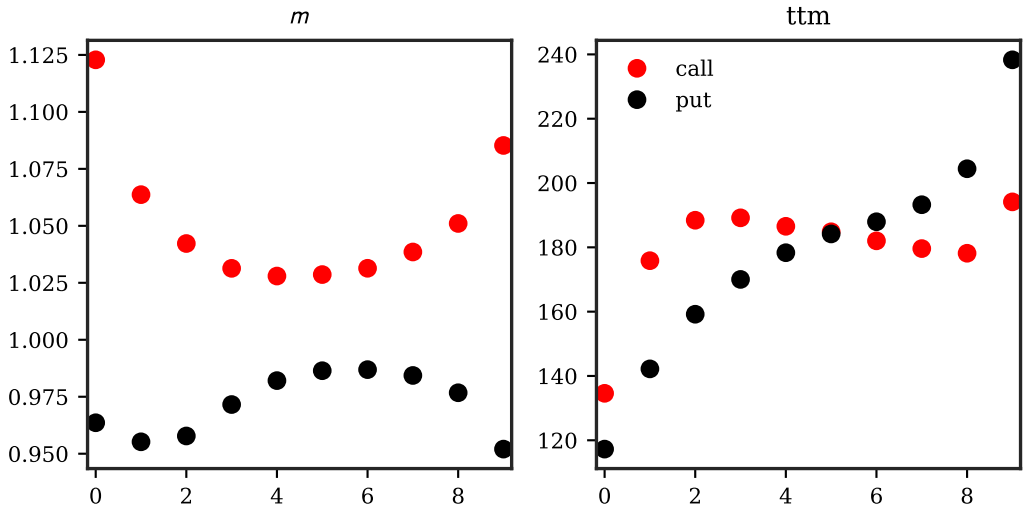


Fig. H.1. Machine Learning Portfolios – Moneyness and Days-to-maturity

This figure shows the average moneyness and time-to-maturity for the decile portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We split the information by put and call options included.

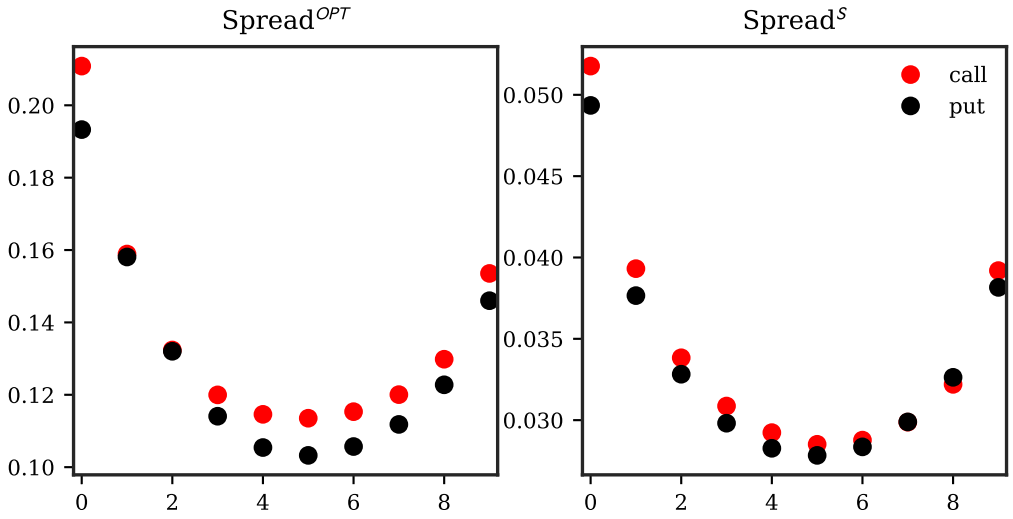


Fig. H.2. Machine Learning Portfolios – Spreads of the Options and the Underlyings

This figure shows the bid-ask spread of the options included in the left and of the underlying stocks in the right panel for the decile portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We split the information by put and call options included.

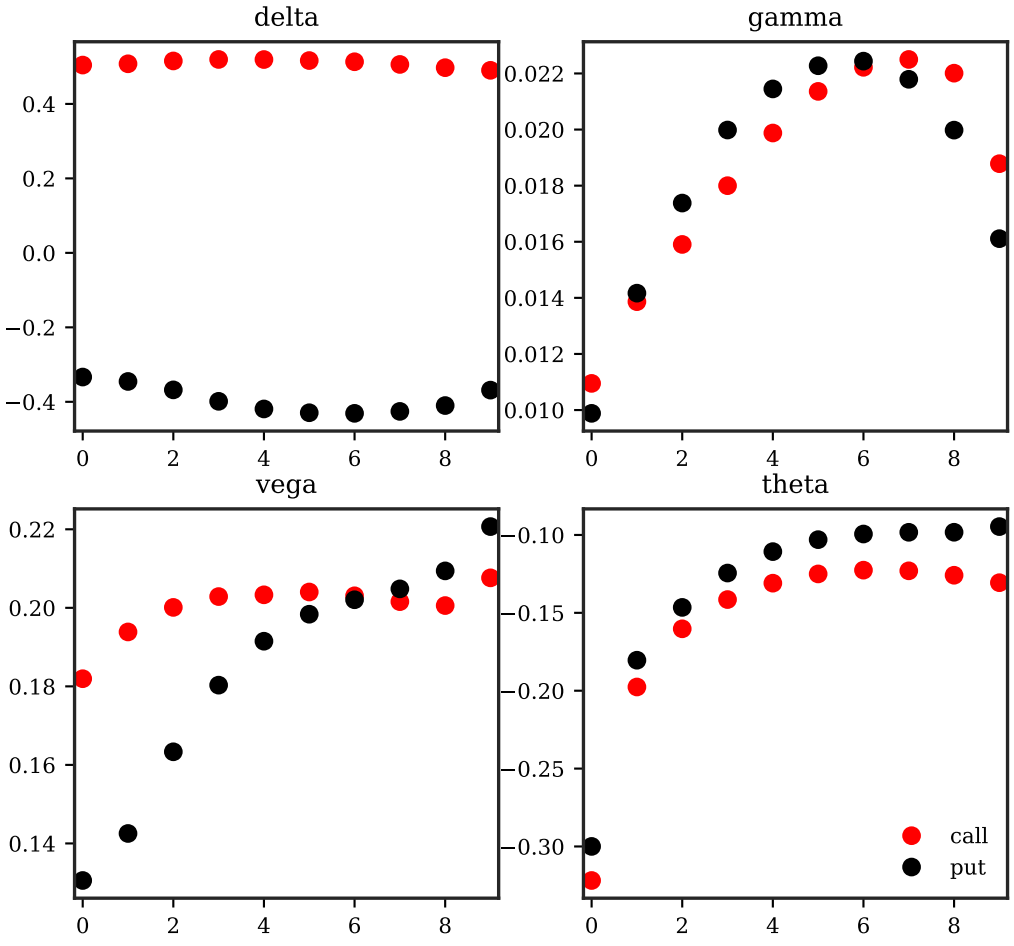


Fig. H.3. Machine Learning Portfolios – Greeks

This figure shows option Greeks for the decile portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We split the information by put and call options included. We show the (unhedged) delta of the option, the *gamma*, *vega*, and *theta*. *gamma* is expressed for a 1% move in the underlying stock ($gamma \times \frac{S}{100}$) and *vega* and *theta* in terms of the underlying stock price ($\frac{x}{S}$ for $x \in [vega, theta]$).



Fig. H.4. Machine Learning Portfolios – Call Share over Time

The figure shows the share of call options included in the portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We provide average numbers per year in our testing sample from 2003 through 2020.

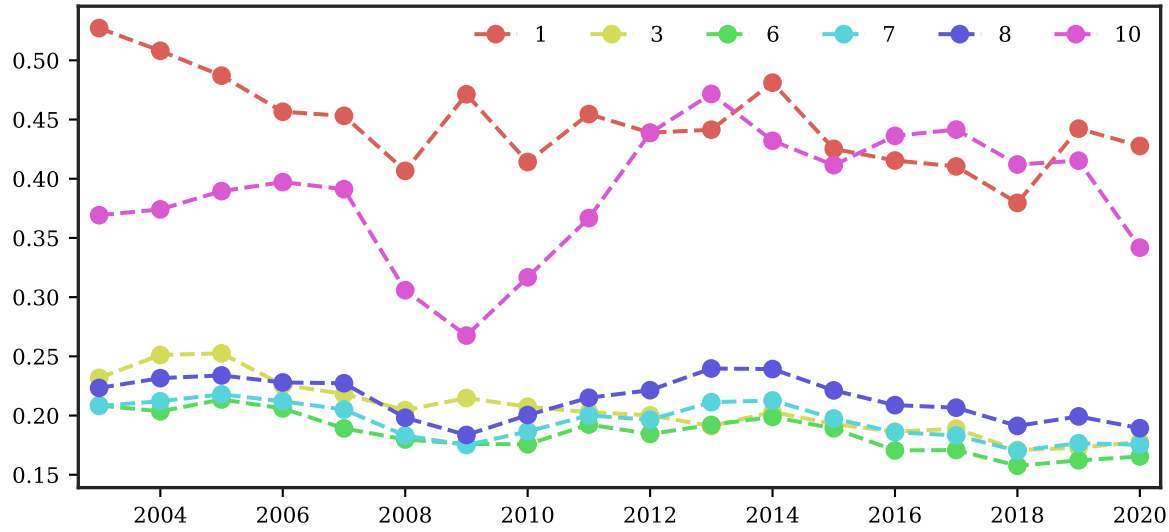


Fig. H.5. Machine Learning Portfolio – Underlying-Option Concentration

The figure shows the share of underlying stocks for which all options written on that stocks are classified into a single portfolio by the nonlinear ensemble N-En. We provide average numbers per year in our testing sample from 2003 through 2020.

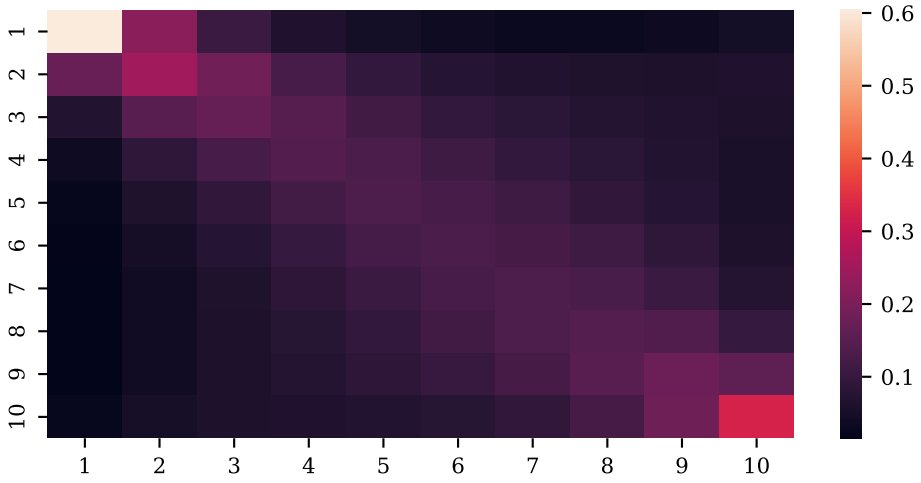


Fig. H.6. ML Portfolio Transition Matrix by Option Bucket

The figure shows the relative likelihood of options for a particular underlying transitioning from one portfolio to another in the next month. Since we cannot estimate this transition for single options due to their fleeting moneyness and time-to-maturity, we use changes in the portfolio mode for a given Permno-bucket combination as an approximation. Buckets are defined as in Section 4.3.

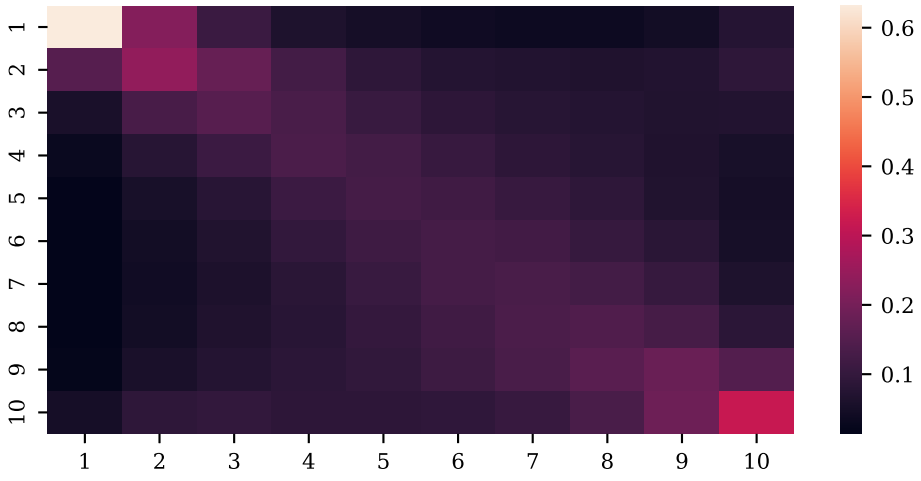


Fig. H.7. ML Portfolio Transition Matrix by Underlying

The figure shows the relative likelihood of options for a particular underlying transitioning from one portfolio to another in the next month. Since we cannot estimate this transition for single options due to their fleeting moneyness and time-to-maturity, we use changes in the portfolio mode for a given Permno as an approximation.

H.3 Performance for ML Portfolios in Different Market Phases

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Low VIX	1.741	1.130	0.923	1.224	2.369	1.782	1.154	1.544	***
		(10.76)		(5.36)		(17.62)		(5.35)	
High VIX	2.269	1.434	1.483	0.967	2.564	2.251	1.698	1.326	***
		(7.75)		(6.66)		(9.85)		(9.02)	
Low EPU	2.081	1.288	1.065	1.209	2.615	1.917	1.256	1.526	***
		(12.15)		(6.41)		(16.64)		(6.50)	
High EPU	1.926	1.274	1.400	0.910	2.316	2.116	1.651	1.281	***
		(7.98)		(7.14)		(9.16)		(8.26)	
Neg. CFNAI	1.983	1.267	1.229	1.031	2.374	2.005	1.373	1.460	***
		(8.19)		(7.07)		(11.84)		(8.70)	
Pos. CFNAI	2.026	1.297	1.259	1.030	2.570	2.027	1.571	1.290	***
		(10.06)		(5.65)		(10.53)		(6.18)	
Low FED Stress	1.883	1.174	1.109	1.058	2.450	1.892	1.391	1.360	***
		(12.72)		(6.02)		(13.58)		(6.58)	
High FED Stress	2.270	1.519	1.471	1.032	2.501	2.290	1.596	1.435	***
		(6.36)		(6.92)		(8.50)		(6.38)	
Low SENT	2.136	1.133	1.057	1.072	2.514	1.876	1.341	1.398	***
		(10.36)		(6.03)		(14.55)		(7.10)	
High SENT	1.923	1.611	1.258	1.280	2.549	1.967	1.180	1.667	
		(13.93)		(5.02)		(17.02)		(5.89)	

Table H.2: Trading on Machine Learning Predictions – Market Phases

The table shows returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods for different sample splits capturing economic states. We split the sample by the median VIX from 2003 through 2020, the median EPU index by Baker, Bloom, and Davis (2016), by the sign of the Chicago Fed National Activity Index, the median of the St. Louis Fed Stress Index, and the median of the sentiment index proposed by Baker and Wurgler (2006). Pred denotes the average predicted return within the respective portfolio, Avg the average realized return, SD the standard deviation of realized returns and finally SR the realized Sharpe ratio. All values are given per month. The last column (N vs. L) gives the significance of comparing the mean realized returns for N-En and L-En. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

H.4 Profitability of Machine Learning Portfolios Per Bucket

			L-En				N-En				
TTM	Mon.	Type	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
	atm		1.788	1.363	1.353	1.008	2.216	1.800	1.469	1.226	***
	itm	C	1.994	0.942	1.304	0.722	3.070	1.775	2.289	0.776	***
$\tau \leq 90$	itm	P	2.333	1.149	1.281	0.898	2.442	1.944	1.875	1.037	***
	otm	C	2.109	3.221	3.636	0.886	3.744	3.523	4.307	0.818	
	otm	P	1.801	2.813	3.235	0.870	3.346	3.874	3.735	1.037	***
	atm		1.668	1.180	1.639	0.720	1.812	1.686	1.882	0.896	***
	itm	C	1.713	1.238	1.442	0.858	2.173	2.207	2.063	1.070	***
$\tau > 90$	itm	P	2.124	0.862	1.472	0.585	1.922	1.333	1.644	0.811	***
	otm	C	1.886	2.461	3.174	0.775	2.722	3.513	3.319	1.058	***
	otm	P	1.666	1.783	2.694	0.662	2.362	2.579	2.878	0.896	***

Table H.3: Trading on Machine Learning Predictions – Buckets

The table shows the returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods for the option buckets defined in Section 4.3 We show the returns to the resulting high-minus-low portfolios. Pred denotes the average predicted return within the respective portfolio, Avg the average realized return, SD the standard deviation of realized returns and finally SR the realized Sharpe ratio. The last column (N vs. L) gives the significance of comparing the mean realized returns for N-En and L-En. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

H.5 Risk Attribution (N-En)

			CAPM		FF6		FF6+PS	
Full Sample			2.058	(13.64)	2.021	(17.46)	2.018	(16.37)
					Buckets			
$\tau \leq 90$	atm		1.792	(14.87)	1.757	(17.38)	1.754	(17.86)
	itm	C	1.740	(8.95)	1.692	(9.96)	1.690	(9.95)
	itm	P	1.898	(8.45)	1.824	(8.37)	1.825	(8.50)
	otm	C	3.617	(11.28)	3.463	(12.22)	3.461	(12.06)
	otm	P	4.112	(13.75)	4.047	(13.66)	4.042	(13.05)
$\tau > 90$	atm		1.699	(11.51)	1.659	(13.10)	1.654	(13.24)
	itm	C	2.270	(11.21)	2.213	(12.95)	2.215	(13.49)
	itm	P	1.327	(8.07)	1.324	(8.67)	1.323	(8.43)
	otm	C	3.579	(10.27)	3.542	(11.91)	3.537	(11.59)
	otm	P	2.693	(11.92)	2.718	(14.46)	2.712	(13.51)

			AN		BM		LBC	
Full Sample			1.975	(14.52)	1.885	(22.99)	1.941	(14.16)
					Buckets			
$\tau \leq 90$	atm		1.751	(15.06)	1.746	(17.24)	1.765	(15.66)
	itm	C	1.575	(9.51)	1.535	(7.96)	1.600	(9.71)
	itm	P	1.904	(9.43)	1.771	(6.78)	1.779	(8.04)
	otm	C	3.516	(12.74)	3.416	(10.91)	3.442	(11.49)
	otm	P	4.325	(15.77)	3.635	(12.31)	4.054	(12.61)
$\tau > 90$	atm		1.562	(11.48)	1.595	(14.22)	1.651	(12.02)
	itm	C	2.207	(9.76)	2.042	(13.46)	2.236	(9.60)
	itm	P	1.229	(9.43)	1.210	(7.11)	1.187	(8.32)
	otm	C	3.447	(9.68)	3.171	(11.26)	3.386	(9.52)
	otm	P	2.843	(12.21)	2.695	(16.96)	2.815	(11.61)

Table H.4: Common Factor Models and Machine Learning Predictions

The table shows the risk-adjusted returns of the high-minus-low portfolio following the predictions by N-En using risk factor models proposed in the literature. Risk-adjusted returns are provided for the CAPM, the Fama and French (2015) 5-factor model plus momentum (Carhart, 1997), FF6; the Fama and French (2015) 5-factor model plus momentum and the liquidity factor of Pástor and Stambaugh (2003), FF6+PS; the model following Agarwal and Naik (2004) including the returns of at-the-money and out-of-the-money index options plus the market factor, AN; the model for optionable stocks by Bali and Murray (2020), including the spread between implied and realized volatility by Bali and Hovakimian (2009), its difference through time by An et al. (2014), the call-minus-put implied volatility spread by Cremers and Weinbaum (2010), and the market factor, BM; and a model including the market factor and the leverage bearing capacity of financial intermediaries proposed in Grünthaler et al. (2020), LBC. We also provide the information for option buckets defined in Section 4.3.

Online Appendix I Which Characteristics Matter? – Additional Analyses

I.1 Feature Group Importance Per Bucket (N-En)

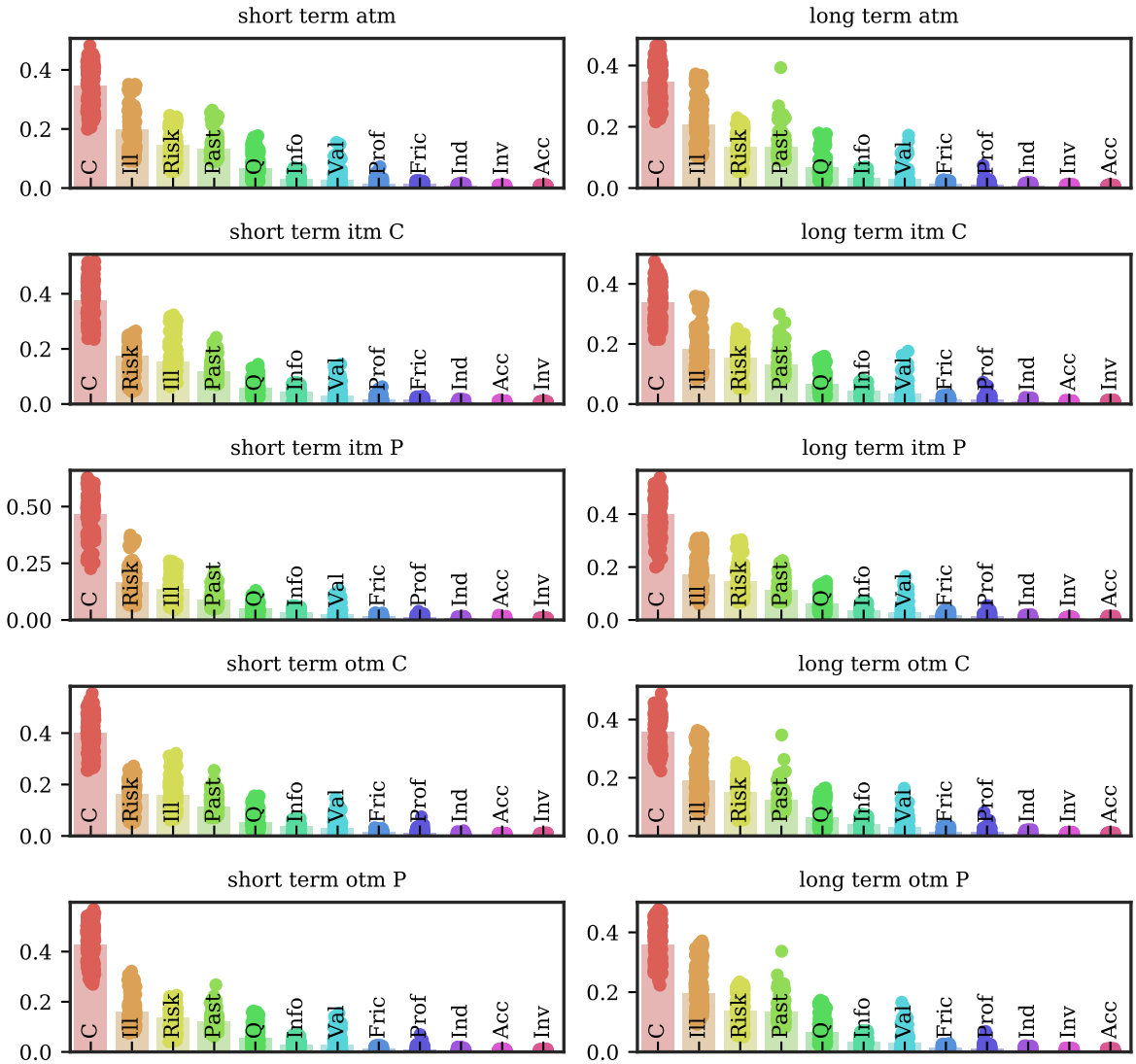


Fig. I.1. Feature Group Importance for N-En Per Bucket

The figure shows the feature group importance for the twelve feature groups defined in Section 4.3 for the nonlinear (N-En) ensemble per option bucket. We measure the importance using SHAP values following Lundberg and Lee (2017). The group importance is the sum of the resulting SHAP values for all features included in a given group. The values are scaled such that they sum to one. The bars represent the mean feature group importance for the entire testing sample, the dots the dispersion of the group importance for the months in the testing sample. The abbreviations used: Acc=Accruals, Prof=Profitability, Q=Quality, Inv=Investment, Ill=Illiquidity, Info=Informed Trading, Val=Value, C=Contract, Past=Past Prices, Fric=Frictions, Ind=Industry.

I.2 Feature Group Importance Over Time (N-En)

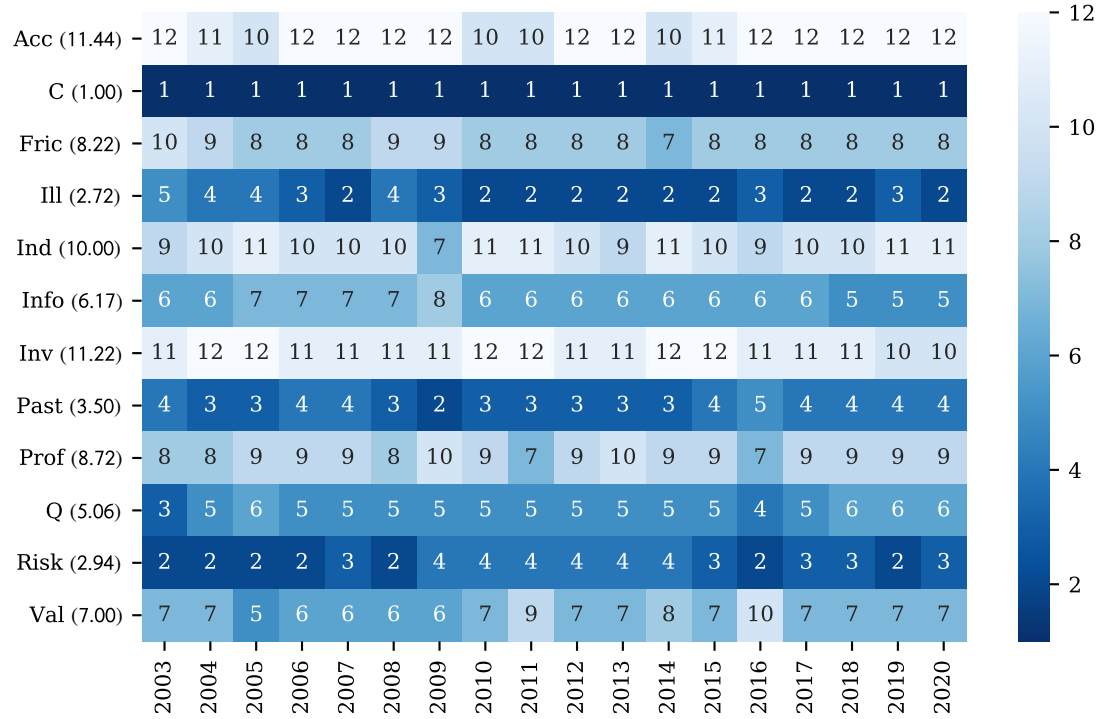


Fig. I.2. Feature Group Importance for N-En Over Time

The figure shows the time evolution of ranking the twelve feature groups defined in Section 4.3 by their importance for the nonlinear ensemble (N-En). 1 denotes the highest-ranking feature group, 12 the lowest-ranking. The average rank of each group is provided in parentheses. We measure the importance using SHAP values following Lundberg and Lee (2017). The group importance is the sum of the resulting SHAP values for all features included in a given group. The values are scaled such that they sum to one. The bars represent the mean feature group importance for the entire testing sample, the dots the dispersion of the group importance for the months in the testing sample. The abbreviations used: Acc=Accruals, Prof=Profitability, Q=Quality, Inv=Investment, Ill=Illiquidity, Info=Informed Trading, Val=Value, C=Contract, Past=Past Prices, Fric=Frictions, Ind=Industry.

I.3 Importance of Most Influential Features Per Bucket (N-En)

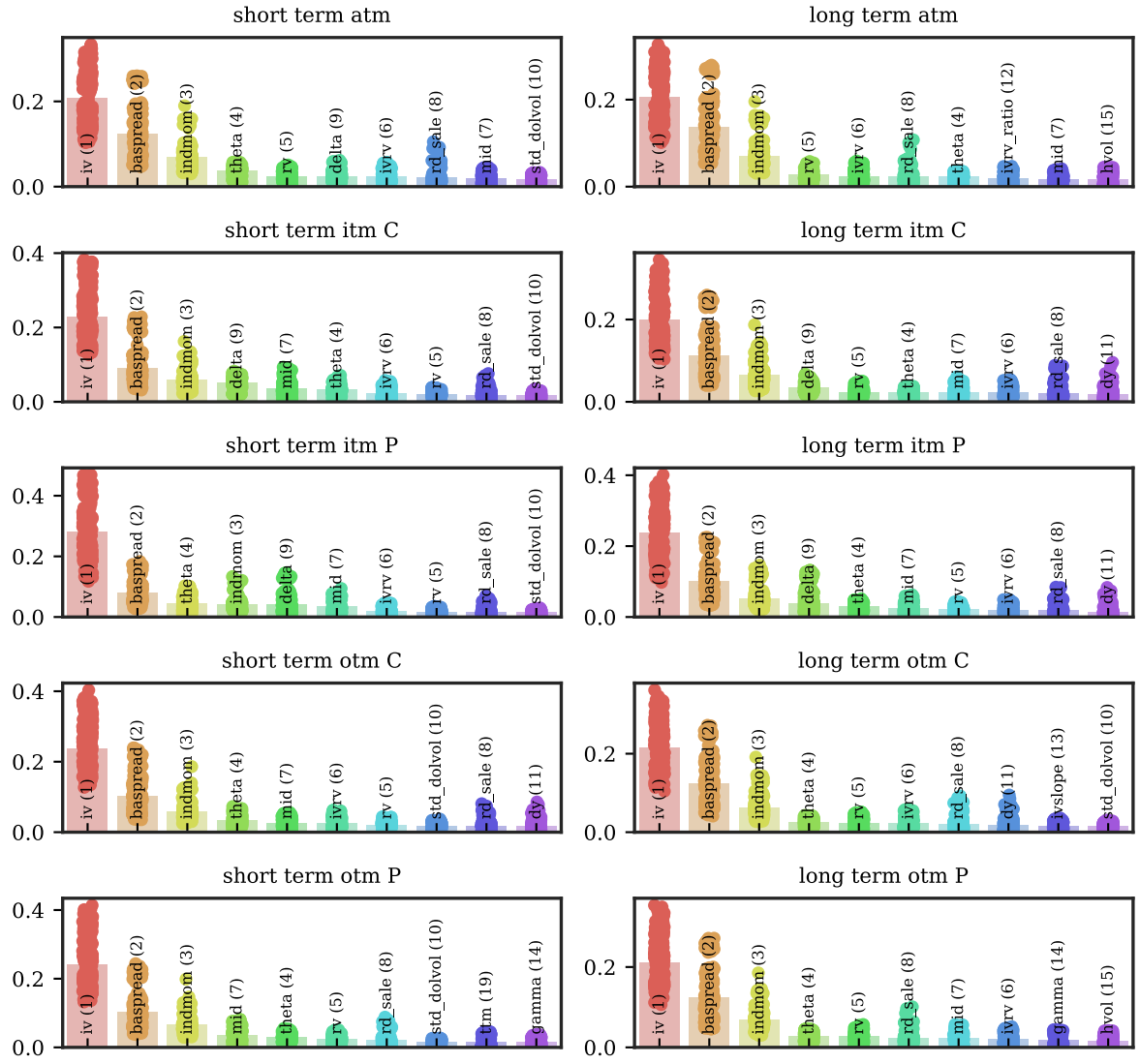


Fig. I.3. Importance of Most Influential Features for N-En Per Bucket

The figure shows the importance of the ten most influential features on the predictions of the nonlinear ensemble (N-En) per bucket. Importance is measured by SHAP values following Lundberg and Lee (2017). The values are scaled such that they sum to one across all 270 characteristics. The bars represent the mean feature importance for the entire testing sample, the dots the dispersion of the importance for the months in the testing sample.

I.4 Impact of Most Important Features Over Time (N-En)

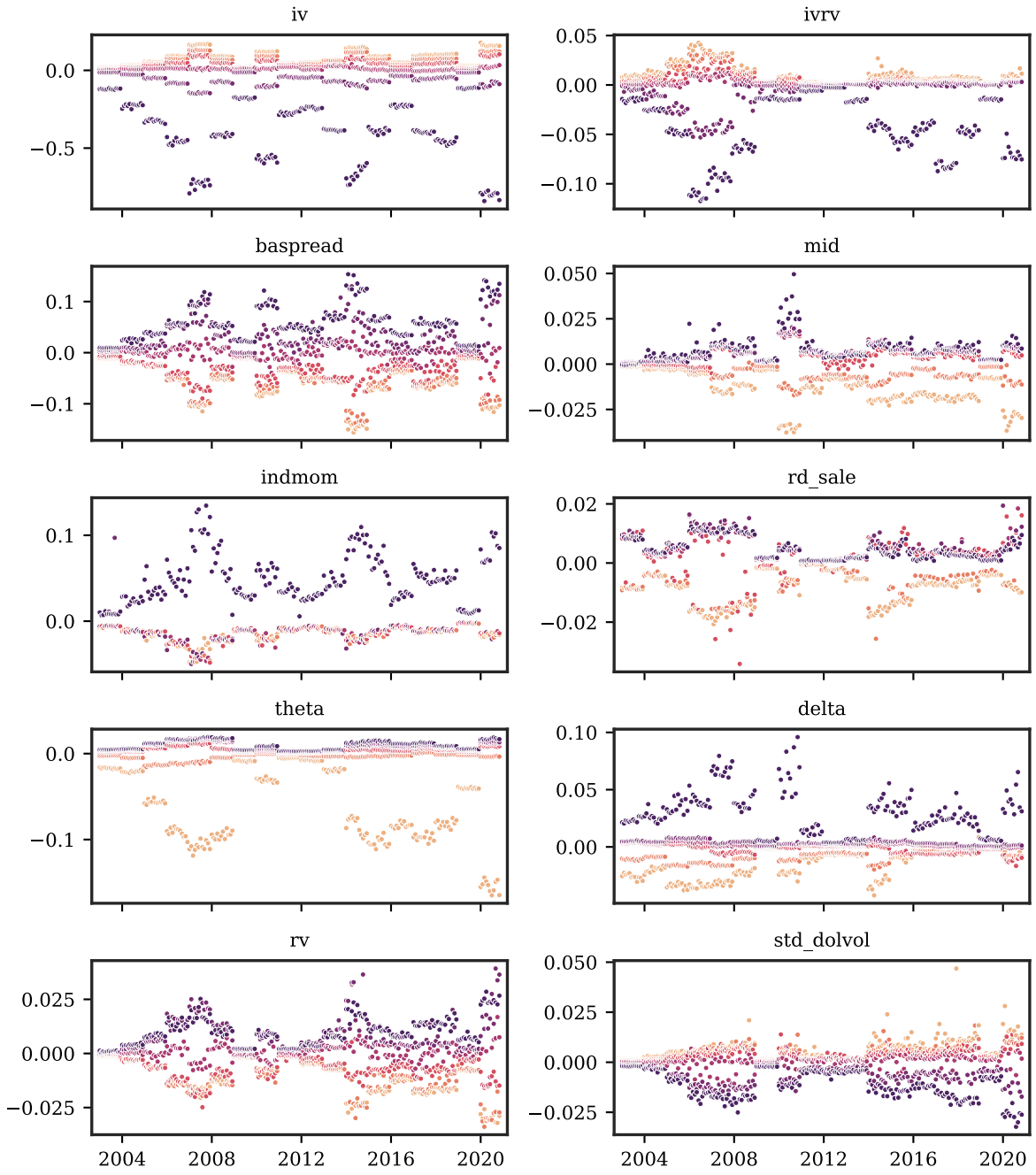


Fig. I.4. Impact of Most Important Features Over Time

The figure shows the impact of the ten most influential features on the predictions of the nonlinear ensemble (N-En) over time. We measure the impact using SHAP values following Lundberg and Lee (2017). Lighter (darker) colored dots denote low (high) feature values. We show the differential impact for quintiles of each feature, measured each month in the testing sample from 2003–2020.

I.5 Functional Form of Impact of Most Important Features (N-En)

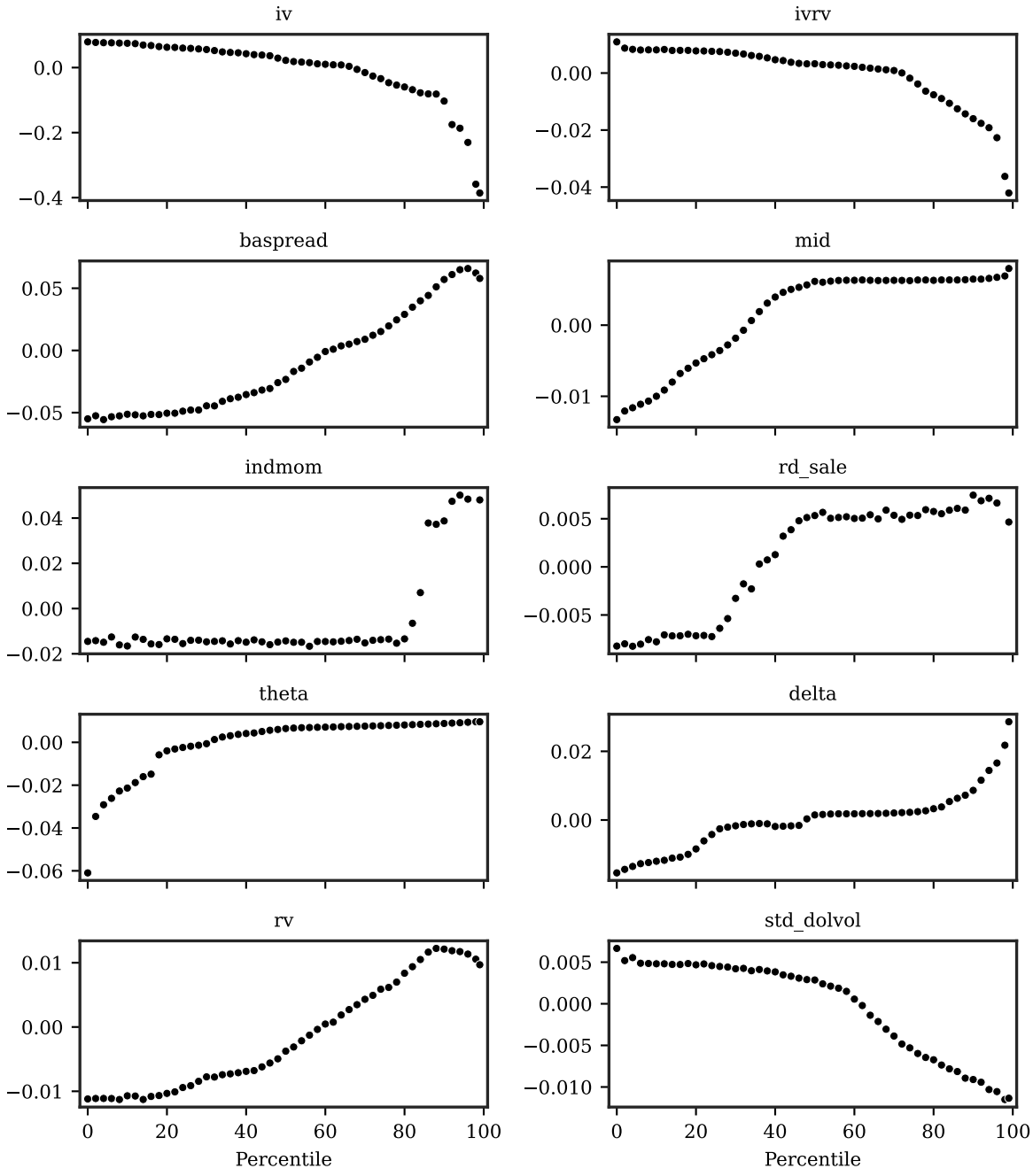


Fig. I.5. Functional Form of Impact of Most Important Features

The figure shows the functional form of the impact of the ten most influential features on the predictions of the nonlinear ensemble (N-En). We measure the impact using SHAP values following Lundberg and Lee (2017).

I.6 Importance of Volatility and Jump Risk (N-En)

			Volatility risk				Jump risk				
			ivvol	volunc	vega	ivrv	tlm30	skewiv	rns30	rnk30	gamma
	all Sample		67	141	48	6	17	122	95	92	14
						Buckets					
	atm		−7	−5	−1	0	−3	−3	4	−1	−3
	itm	C	12	1	−11	1	4	5	−3	2	2
$\tau \leq 90$	itm	P	10	0	−12	1	11	7	0	9	1
	otm	C	7	−1	0	1	−5	−11	−5	0	−1
	otm	P	6	−1	−22	5	−1	−8	−19	3	−4
	atm		−7	0	3	−1	0	−2	3	−3	8
	itm	C	11	3	2	1	2	3	−1	0	−1
$\tau > 90$	itm	P	10	3	−2	2	4	7	3	0	2
	otm	C	4	2	10	0	−1	1	−3	−1	3
	otm	P	3	2	−2	3	−2	−3	−5	−1	−6

Table I.1: Importance of Volatility and Jump Risk

The table shows the ranking of features proxying for volatility and jump risk, respectively. Ranks are formed by measuring importance of features using SHAP values following Lundberg and Lee (2017). Higher numbers for the full sample denote lower-ranking, i.e., less important features. Numbers for the buckets are expressed relative to the full sample, i.e., negative numbers for the buckets denote higher importance compared to the full sample. Proxies for volatility (jump) risk are: ivvol, volunc, vega, and ivrv (tlm30, skewiv, rns30, rnk30, and gamma).

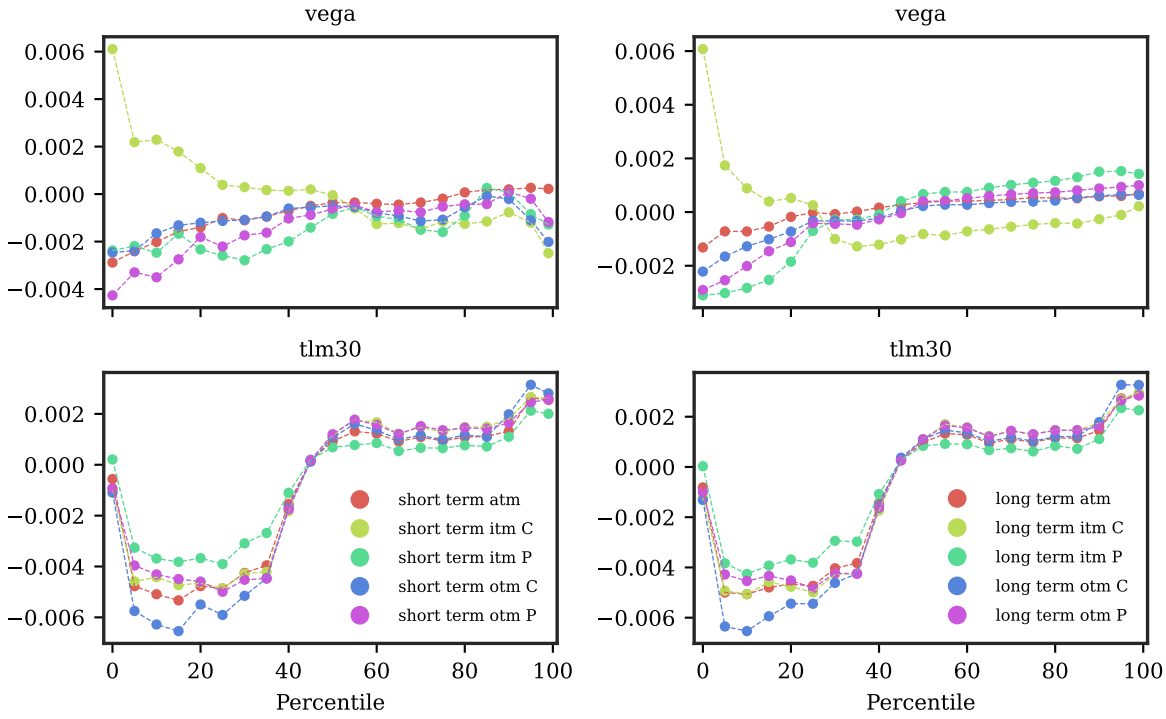


Fig. I.6. Relative Impact of Volatility and Jump Risk Premia Per Bucket

The figure shows the impact of features proxying for volatility (vega) and jump risk (tlm30), respectively, on the predictions of the nonlinear ensemble (N-En) per bucket. The impact of the feature is measured using SHAP values following Lundberg and Lee (2017). vega is the option's vega and tlm30 denotes the option implied tail risk following Vilkov and Xiao (2012).

Online Appendix J Impact of the Information Set – Additional Analyses

J.1 Restricting the Information Set (N-En)

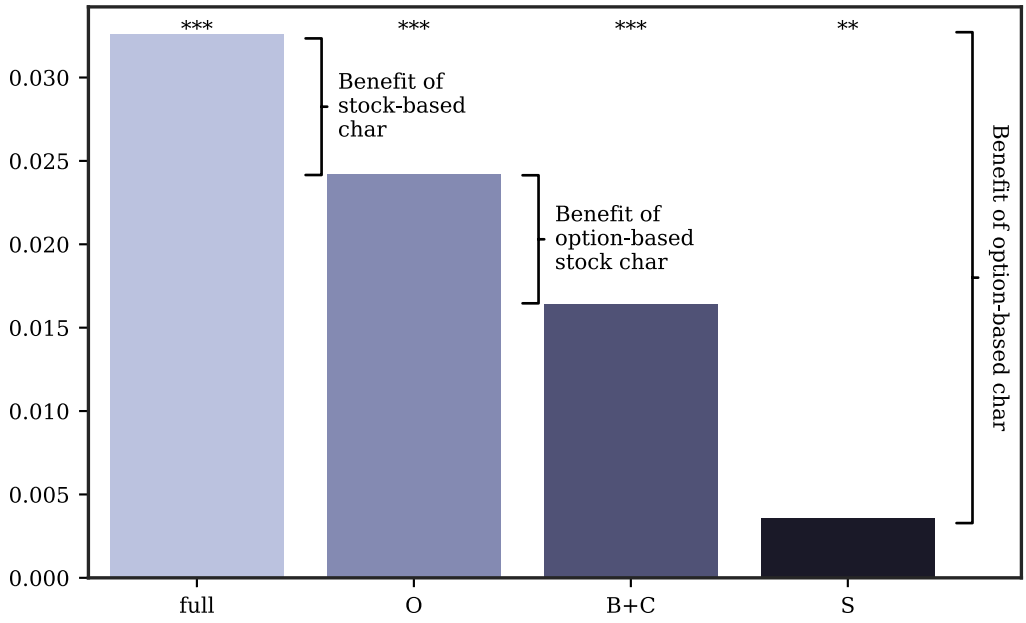


Fig. J.1. Restricting the Information Set for N-En – $R^2_{OS;XS}$

The figure shows the cross-sectional out-of-sample $R^2_{OS;XS}$ defined in Equation (5) for N-En with restricted access to the full set of characteristics. The full model is shown in the left bar for reference, and is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Section 4.3. ***, **, * below the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7) for the sample of "all" options. The testing sample spans the years 2003 through 2020.

J.2 Option-bucket Performance for Different Information Sets

			L-En				N-En			
TTM	Mon.		B+C	O	S	full	B+C	O	S	full
	atm		-0.003	-0.002	-0.000	0.004	0.011**	0.018***	0.006	0.022***
	itm	C	0.011	0.014	0.002	0.018*	0.007	0.029	0.016**	0.059***
$\tau \leq 90$	itm	P	0.026***	0.027**	0.010*	0.016	0.063***	0.060**	0.028***	0.061**
	otm	C	0.005	0.008*	0.002	0.012***	0.016***	0.023***	0.007***	0.027***
	otm	P	0.020***	0.021***	0.003	0.027***	0.033***	0.041***	0.005	0.048***
	atm		-0.012	-0.007	-0.013	0.005	0.003	0.013	-0.003	0.025***
	itm	C	-0.004	0.002	-0.011	0.014	0.013	0.033**	-0.002	0.047***
$\tau > 90$	itm	P	-0.033	-0.030	-0.029	-0.026	0.004	0.007	-0.013	0.015
	otm	C	-0.006	-0.003	-0.005	0.002	0.001	0.009	0.001	0.017***
	otm	P	-0.004	-0.001	-0.006	0.009*	0.002	0.012	0.001	0.023***

Table J.1: Option-bucket performance for different information sets

The table shows the out-of-sample R^2 defined in Equation (4) for models with restricted access to the full set of characteristics for options in a respective bucket, as defined in Section 4.3. The full model is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Section 4.3. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

Online Appendix K Sources of Option Return Predictability – Additional Analyses

K.1 Presence of Informed Investors

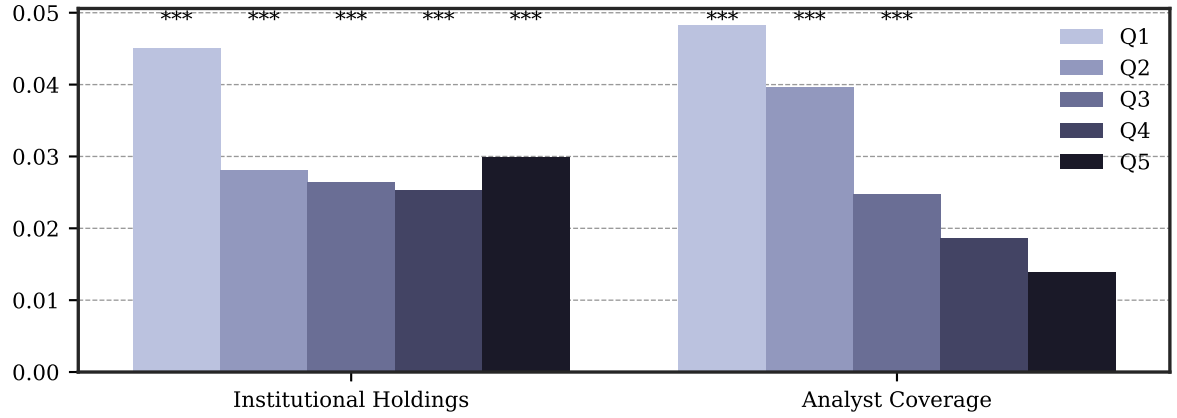


Fig. K.1. Predictability and Profitability Conditional on Informed Investor Presence – $R^2_{OS;XS}$

The left panel of the figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En for different quintiles of institutional ownership and analyst coverage of the underlying stocks. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7). The right panel shows the resulting Sharpe ratios of buying options in the highest predicted return decile and selling options in the lowest for each subsample.

K.2 Option Demand

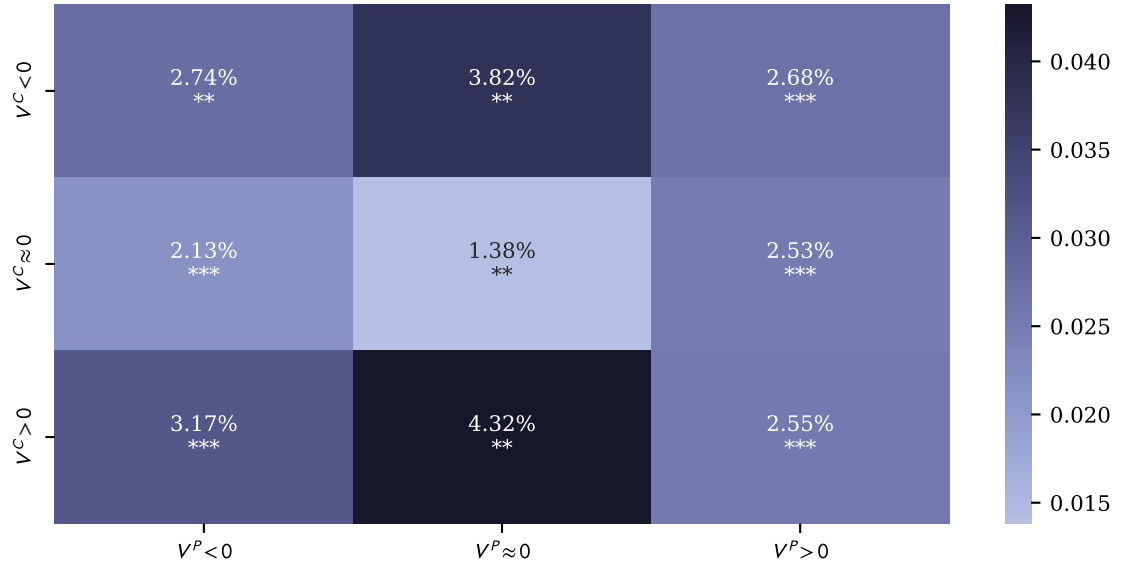


Fig. K.2. Predictability Conditional on Option Demand – $R^2_{OS;XS}$

The figure shows the cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En conditional on the net open interest of retail and institutional investors, I. We independently form terciles based on the cross-sectional distribution of I^R and I^I. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

K.3 Option Mispricing

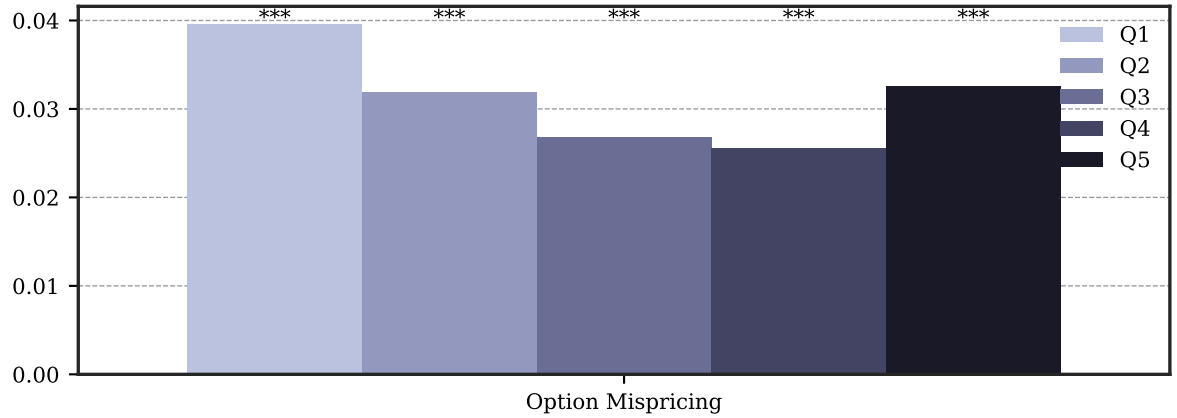


Fig. K.3. Predictability and Profitability Conditional on Option Mispricing – $R^2_{OS;XS}$

The left panel of the figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En for different quintiles of option mispricing. We calculate option mispricing using the stocks' realized volatility over the last quarter and calculate "fair prices" for short-term at-the-money options using it as input to the Black and Scholes (1973) model. Mispricing is then defined as $\log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the obtained "fair price". ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

Bibliography

Agarwal, V., Naik, N. Y., 2004. Risks and portfolio decisions involving hedge funds. Review of Financial Studies 17, 63–98.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

An, B.-J., Ang, A., Bali, T. G., Cakici, N., 2014. The joint cross section of stocks and options. Journal of Finance 69, 2279–2337.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The journal of Finance 61, 1645–1680.

Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring economic policy uncertainty. The Quarterly Journal of Economics 131, 1593–1636.

Bali, T. G., Hovakimian, A., 2009. Volatility spreads and expected stock returns. Management Science 55, 1797–1812.

Bali, T. G., Murray, S., 2020. In search of a factor model for optionable stocks.

Baltussen, G., van Bekkum, S., van der Grient, B., 2018. Unknown Unknowns: Uncertainty About Risk and Stock Returns. Journal of Financial and Quantitative Analysis 53, 1615–1651.

Bao, J., Pan, J., Wang, J., 2011. The illiquidity of corporate bonds. Journal of Finance 66, 911–946.

Bergsma, K., Fodor, A., Tedford, E., 2020. A closer look at the disposition effect in u.s. equity option markets. Journal of Behavioral Finance 21, 66–77.

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of Machine Learning Research 13.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Blau, B. M., Nguyen, N., Whitby, R. J., 2014. The information content of option ratios. Journal of Banking & Finance 43, 179 – 187.

Borochin, P., Chang, H., Wu, Y., 2020. The information content of the term structure of risk-neutral skewness. Journal of Empirical Finance 58, 247 – 274.

Buchner, M., Kelly, B. T., 2020. A Factor Model for Option Returns.

Cao, J., Vasquez, A., Xiao, X., Zhan, X., 2019. Volatility Uncertainty and the Cross-Section of Option Returns.

Cao, M., Wei, J., 2010. Option market liquidity Commonality and other characteristics. Journal of Financial Markets p. 29.

Cremers, M., Weinbaum, D., 2010. Deviations from put-call parity and stock return predictability. Journal of Financial and Quantitative Analysis 45, 335–367.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fong, K. Y., Holden, C. W., Trzcinka, C. A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Gilad-Bachrach, R., Rashmi, K., 2015. Dart: Dropouts meet multiple additive regression trees. Cornell University: Cornell, Ithaca, NY, USA .

Goyenko, R. Y., Holden, C. W., Trzcinka, C. A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153–181.

Green, J., Hand, J. R. M., Zhang, X. F., 2017. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. Review of Financial Studies 30, 4389–4436.

Grünthaler, T., Lorenz, F., Meyerhof, P., 2020. The leverage bearing capacity: A new tool for intermediary asset pricing.

Gu, S., Kelly, B. T., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Heston, S. L., Li, S., 2020. Option momentum.

Hiraki, K., Skiadopoulos, G., 2020. The Contribution of Frictions to Expected Returns.

Huang, T., Li, J., 2019. Option-implied variance asymmetry and the cross-section of stock returns. Journal of Banking & Finance 101, 21–36.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2021. Is there a replication crisis in finance? National Bureau of Economic Research.

Johnson, T. L., So, E. C., 2012. The option to stock volume ratio and future returns. Journal of Financial Economics 106, 262 – 286.

Jones, C. S., Khorram, M., Mo, H., 2020. Momentum, reversal, and seasonality in option returns.

Karakaya, M. M., 2014. Characteristics and expected returns in individual equity options.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 30, 3146–3154.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Lesmond, D. A., Ogden, J. P., Trzcinka, C. A., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113–1141.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., Talwalkar, A., 2017. Hyperband: A novel bandit-based approach to hyperparameter optimization. Journal of Machine Learning Research 18, 6765–6816.

Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Hardt, M., Recht, B., Talwalkar, A., 2018. Massively parallel hyperparameter tuning .

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J. E., Stoica, I., 2018. Tune: A research platform for distributed model selection and training. arXiv preprint arXiv:1807.05118 .

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 .

Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777.

Muravyev, D., Pearson, N. D., Pollet, J. M., 2021. Is there a risk premium in the stock lending market? evidence from equity options. Journal of Finance, forthcoming.

Ofek, E., Richardson, M., Whitelaw, R. F., 2004. Limited arbitrage and short sales restrictions: Evidence from the options markets. Journal of Financial Economics 74, 305–342.

Pástor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., pp. 8024–8035.

Reddi, S. J., Kale, S., Kumar, S., 2019. On the convergence of adam and beyond. arXiv preprint arXiv:1904.09237 .

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of finance 39, 1127–1139.

Roll, R., Schwartz, E., Subrahmanyam, A., 2010. O/S: The relative trading activity in options and stock. Journal of Financial Economics 96, 1–17.

Schlag, C., Thimme, J., Weber, R., 2020. Implied volatility duration: A measure for the timing of uncertainty resolution. Journal of Financial Economics, forthcoming.

Vasquez, A., 2017. Equity volatility term structures and the cross section of option returns. Journal of Financial and Quantitative Analysis 52, 2727–2754.

Vilkov, G., Xiao, Y., 2012. Option-implied information and predictability of extreme returns.

Xing, Y., Zhang, X., Zhao, R., 2010. What Does the Individual Option Volatility Smirk Tell Us About Future Equity Returns? Journal of Financial and Quantitative Analysis 45, 641–662.

Chapter 5

An Autoencoder Based Factor Model for Option Returns

Mathis Mörke

ABSTRACT

I adopt the recently proposed conditional autoencoder model of Gu, Kelly, and Xiu (2021) and study its performance in the S&P 500 index options market. The model consists of latent factors with time-variation in their loadings making it well suited for studying option returns. Its novelty is the incorporation of non-linearities which distinguishes itself from the state-of-the-art workhorse of factor modeling, instrumented principal component analysis. I find that a conditional autoencoder model with one hidden layer and three factors explains more than 75% of the out-of-sample variation in option returns described by the contemporaneous latent factors. Moreover, it excels in explaining variation in risk compensation and out-of-sample trading strategies.

Keywords: Machine learning, big data, option return predictability

JEL classification: G10, G12, G13, G14.

1. Introduction

Understanding the trade-off between risks and expected returns investors face in financial markets is a key element of asset pricing. The most common approach for evaluating the trade-off are factor models which are statistical models aimed at describing the cross-sectional dependence structure of returns utilizing a small set of factors. Whereas the risk-return trade-off is nowadays well understood in most asset classes, the literature on option returns is rather scarce and has been less of a focus. This is surprising given that the options market has grown tremendously over the last few years.

An important contribution in that regard has been made by Büchner and Kelly (2022). The authors use the recently suggested framework of instrumented principal component analysis (IPCA; Kelly et al., 2019) to propose a factor model for S&P 500 index option returns. IPCA is able to address subtleties of options markets compared to stock markets. Options exhibit very short lifespans and highly time-varying risk attributes. IPCA aims at solving the aforementioned problems by employing a latent conditional factor model. It does not require to pre-specify ad hoc factors as in the traditional approach of Fama and French (1993, 2015), in which observable factors are treated as true underlying factors, neglecting measurement and specification error (Kelly et al., 2020a). Additionally, IPCA incorporates time variation in factor loadings by assuming that loadings depend on observable characteristics.

However, IPCA rests on the simplifying assumption that factor loadings are linear functions of characteristics. As it is recognized in the literature, there is no theoretical or empirical basis for this. On the contrary, the prior literature has hinted toward non-linearities (Campbell and Cochrane, 1999; Bansal and Yaron, 2004; Pohl, Schmedders, and Wilms, 2018). Hence, Gu et al. (2021) propose the use of autoencoders, a machine learning technique for dimension reduction, in combination with the conditioning approach in factor loadings of IPCA. Whereas IPCA assumes a linear function of characteristics, Gu et al. (2021) use neural networks and term their approach conditional autoencoders. Allowing for nonlinear conditional exposure, the authors document lower out-of-sample pricing errors compared to IPCA for modeling single stock returns.

In this paper, I adopt the conditional autoencoder model of Gu et al. (2021) and apply it to S&P 500 index options returns. I benchmark its performance against IPCA, as Büchner and Kelly (2022) has shown that IPCA prevails over existing factor models proposed in the literature in the case of S&P 500 index options returns. The conditional autoencoder model might be especially suited for the use of options markets, as it inherits

the property of time-varying factor loadings, but allows for a flexible, non-linear way to map characteristics into betas.

Following Gu et al. (2021), I focus my comparison on conditional autoencoders and IPCA on a solely out-of-sample basis. First, I find that IPCA exhibits higher explanatory power of contemporaneous factor realizations, and hence, excels in describing the individual option contracts riskiness. For example, whereas the conditional autoencoder model with one hidden layer achieves a total R^2 of 76.8% in its three-factor specification, IPCA yields 87.1% for the same number of latent factors. I observe the same gap in statistical performance when focusing on characteristic-managed portfolios as test assets, instead of individual option contracts. The findings reserve as soon as the predictive R^2 is used as a measure for comparison. The predictive R^2, representing the accuracy of model-based predictions of future delta-hedged option returns, is more than twice and up to three times larger for conditional autoencoders. Moreover, it is higher the deeper the conditional autoencoder, i.e., the higher the number of hidden layer in the mapping between characteristics and factor loadings.

Despite the statistical performance, I compare conditional autoencoders to IPCA in economic terms. Precisely, I build long-short quintile spread portfolios based on sorted out-of-sample option return predictions for each model. I find that conditional autoencoders dwarf the performance of IPCA by means of average monthly returns, but also in risk-adjusted performance metrics. As an example, IPCA with three factors yields monthly excess returns of 0.25% whereas trading on the predictions of the conditional autoencoder with two hidden layers generates monthly excess returns of 0.46%. These findings are robust to either value-weighting or equally-weighting portfolios.

Conditional autoencoders and IPCA perform similarly well across different ranges in the time-to-maturity and moneyness dimensions. Specifically, I find that both types of models perform better for puts than for calls, for out-of-the money options than in-the-money options, and for medium-to-long-term options instead of short-term options. These options buckets mark also the performance drivers in the long-short portfolios.

Whereas IPCA can be specified to allow for arbitrage, the conditional autoencoder framework obeys the no-arbitrage principle. Hence, it enforces that characteristics can only predict option returns through systematic exposure. My results indicate that the assumption of no-arbitrage is economically supported, similar to the findings of Büchner and Kelly (2022) when using IPCA for modeling S&P 500 index option returns. However, comparing IPCA to conditional autoencoders, I find that the latter yields slightly higher out-of-sample pricing errors of characteristics-based portfolios.

Finally, I quantify the relative importance of different characteristics for the prediction of option returns. I adopt recent advances and resort to SHAP values (Lundberg and Lee, 2017) for identifying influential characteristics. SHAP is a game-theoretic approach that can be used to gauge the importance of characteristics without re-fitting. In line with findings using IPCA to model index option returns, the conditional autoencoder models agree on option implied volatility, vega, and gamma as the most important characteristics. However, contrary to IPCA, the first set of models puts also a high emphasis on time-to-maturity. Precisely, the importance of time-to-maturity is increasing in the number of hidden layers.

My paper is related to various strands of the literature. The traditional approach in options markets has used parametric no-arbitrage pricing models, typically featuring jump, volatility, and tail risks (Heston, 1993; Duffie, Pan, and Singleton, 2000; Carr and Wu, 2004). However, previous research has documented arbitrage opportunities in options markets. For example, Constantinides, Jackwerth, and Perrakis (2009) find evidence of stochastic dominance violations in one-month S&P 500 index options. The authors show that not only calls are priced differently from puts, but also document pricing differences in the moneyness dimension for index options. Ofek et al. (2004) point toward market segmentation due to investor demands for single name options. More recently, Israelov and Kelly (2017) show that even state-of-the-art affine option pricing models empirically fail in describing option returns.

The second strand has proposed various characteristics to explain option returns. These can be broadly classified into three themes. First, risk preferences might affect option prices. For example, Bakshi et al. (2003), similarly to Coval and Shumway (2001), show that volatility risk is priced in index option returns. Goyal and Saretto (2009) show that the difference between historical realized volatility and at-the-money implied volatility is priced in option returns. Second, option prices might be influenced by behavioral biases, as Eisdorfer et al. (2020) show that options on stocks with low nominal prices are overpriced. Moreover, Andreou, Bali, Kagkadis, and Lambertides (2021) document that options on stocks with high growth potential are overpriced. The authors trace the overpricing to retail customers exerting buying pressure on call options as they overestimate the return potential of the underlying stocks. Lastly, option prices are potentially affected by market frictions, such as embedded leverage (Frazzini and Pedersen, 2021) or the illiquidity of the underlying stock (Cao et al., 2021). Third, I contribute to the literature on factor models in asset pricing. Recently, the framework of IPCA has become the state-of-the-art workhorse for factor modeling (Kelly et al., 2019). Similar to

principal component analysis (Kozak, Nagel, and Santosh, 2018; Pukthuanthong, Roll, and Subrahmanyam, 2019; Giglio and Xiu, 2021), IPCA does not require to pre-specify ad hoc factors. However, different from principal component analysis, IPCA uses asset characteristics as conditioning information in specifying betas. Hence, it is able to provide fast updates of risk exposures and does not need to rely on rolling regressions for time-varying factor loadings. Previous research has documented large gains in describing asset returns when using IPCA. For example, Kelly et al. (2020a) documents superior performance in statistical and economic terms when describing the risk and returns of corporate bonds via IPCA. Similar conclusions have been made for equity markets (Kelly et al., 2019) and cryptocurrencies (Bianchi and Babiak, 2021). Lately, Gu et al. (2021) relax the linearity assumption in how asset characteristics influence factor loadings. The authors propose the conditional autoencoder model and show that their non-linear framework delivers smaller out-of-sample pricing errors than other leading factor models when applied to stock returns.

Finally, and more specifically, I add to the relatively scarce literature on factor models in options markets. As for U.S. single stock options, Karakaya (2014) identifies a three-factor model consisting of a level, slope, and value factor. Cao et al. (2021) suggest a two-factor model, idiosyncratic volatility and stock illiquidity, for at-the-money calls. Their approach resembles the traditional approach using long-short portfolios as factors. Horenstein, Vasquez, and Xiao (2020) use principal component analysis to study the factor structure in at-the-money call options on single stocks. Büchner and Kelly (2022) use IPCA for S&P 500 index options and document large gains in using a latent conditional factor model over alternative approaches. Jones (2006) estimates a non-linear factor model for short-term out-of-the-money S&P 500 index put options. Christoffersen, Fournier, and Jacobs (2018) use principal component analysis for single-name equity implied volatility, skew, and term structures. They document a strong correlation of single name principal components to the S&P 500 index option analogs and propose a one-factor option valuation model. More recently and closest to my study, Fournier, Jacobs, and Orłowski (2021) propose a non-linear factor model for S&P 500 index options across all maturities and moneyness ranges. However, there are important differences in my approach. Fournier et al. (2021) require pre-specified ad hoc factors. On the contrary, I employ the conditional autoencoder model which does not need to pre-specify ad hoc factors preventing potential measurement and specification error. Moreover, I am interested in non-linearities in the factor loadings, whereas Fournier et al. (2021) model the factors' impacts on returns in a non-linear way.

The remainder of the paper is structured as follows. Section 2 details my methodological framework. In Section 3, I present my data set. Section 4 describes and discusses my main empirical findings, whereas Section 5 concludes.

2. Methodolgy and Performance Evaluation

In this section, I describe the conditional autoencoder of Gu et al. (2021) and its estimation procedure. I start with outlining the approaches of instrumented principal component analysis (IPCA, thereafter) and autoencoders, as these form building blocks for the conditional autoencoder model.[1]

2.1. IPCA

In its general form, IPCA models an excess return $r_{i,t+1}$ as

$$\begin{aligned} r_{i,t+1} &= \alpha_{i,t} + \beta_{i,t} f_{t+1} + \epsilon_{i,t+1}, \\ \alpha_{i,t} = z'_{i,t}\Gamma_\alpha + \nu_{\alpha,i,t}, &\quad \beta_{i,t} = z'_{i,t}\Gamma_\beta + \nu_{\beta,i,t}, \end{aligned} \tag{1}$$

where $\beta_{i,t}$ are loadings, time-varying, and partially dependent on $z_{i,t}$, a $L \times 1$ vector of asset characteristics. f_{t+1} denotes a $K \times 1$ vector of factors. The system is estimated for a total of N assets and T periods by means of an alternating least squares procedure iterating between the first order conditions of Γ and f_{t+1}.

Similar to principal component analysis (Kozak et al., 2018; Pukthuanthong et al., 2019; Giglio and Xiu, 2021), IPCA does not require to pre-specify ad hoc factors as in the traditional approach of Fama and French (1993, 2015), in which observable factors are treated as true underlying factors, neglecting measurement and specification error. Additionally, IPCA incorporates time variation in factor loadings. Time-varying conditional factor loadings are estimated by assuming that loadings depend on observable characteristics. This not only contrasts with principal component analysis, but also with the standard approach of estimating conditional factor loadings via time-series regressions. Moreover, IPCA embeds dimension reduction which is helpful for several reasons (Kelly and Xiu, 2021). First, the traditional approach of specifying risk factors as long-short portfolios of characteristics has led to a "factor zoo" (Cochrane, 2011) as different studies have proposed their own distinct factors in isolation. IPCA allows the incorporation of

[1] IPCA is described in detail in Kelly et al. (2020b).

a vast set of potential characteristics and lets the data decide which characteristics are most informative. This also addresses the problem of aggregating information from a large set of informative, but noisy and imperfect characteristics.

2.2. Autoencoder

Autoencoders belong to the class of unsupervised machine learning algorithms. It is a special neural network and consists of two parts. In the first part, input variables are fed into hidden layer(s) containing a small number of neurons (encoding), which are then unpacked and mapped to the output layer (decoding) in the second part.

Generally, autoencoders with L hidden layers can be expressed as follows (Gu et al., 2021): Let K^L denote the number of neurons in each layer L. $r^{(l)} = \left(r_1^{(l)}, \ldots, r_{K^{(l)}}^{(l)}\right)$ denotes the vector of outputs for layer L, where $r_k^{(l)}$ is the output of neuron k in layer l. In each layer, the inputs from the previous layer are transformed using a non-linear activation function $g(\cdot)$ before being passed on to the next layer. Let $r = r^{(0)}$ denote the input to the first layer. Then, the recursive output formula for layer $l > 0$ is

$$r^{(l)} = g\left(b^{(l-1)} + W^{(l-1)} r^{(l-1)}\right), \tag{2}$$

where $W^{(l-1)}$ is a $K^{(l)} \times K^{(l-1)}$ matrix of weight parameters, and $b^{(l-1)}$ is a $K^{(l)} \times 1$ vector of bias parameters. I follow Gu et al. (2021) and use the rectified linear unit (ReLU), $g(y) = \max(y, 0)$, as the non-linear activation function. Each hidden layer contains $K^{(l)}(1 + K^{(l-1)})$ numbers of parameters. The output of the autoencoder is

$$G(r, b, W) = b^{(l)} + W^{(l)} r^{(l)}. \tag{3}$$

The output layer features the same dimension as the input layer and tries to approximate r itself.

2.3. Conditional Autoencoder

Gu et al. (2021) propose a new neural network architecture to extend the standard autoencoder to incorporate covariates such as in IPCA. Hence, the conditional autoencoder combines the non-linear dimension reduction of an autoencoder with the inclusion of conditioning variables for the identification of the factor structure.

Figure 1, taken from Gu et al. (2021), graphically describes the conditional autoen-

coder. Factor loadings are modeled as non-linear functions of asset characteristics (left-hand side), contrary to IPCA proposing a linear functional form. Let the $K \times 1$ vector $\beta_{i,t-1}$ denote the factor loadings. The conditional autoencoder assumes the following recursive formulation for the non-linear beta function:

$$z_{i,t-1}^{(0)} = z_{i,t-1}, \tag{4}$$

$$z_{i,t-1}^{(l)} = g(b^{(l-1)} + W^{(l-1)} z_{i,t-1}^{(l-1)}), l = 1, \ldots, L_\beta, \tag{5}$$

$$\beta_{i,t-1} = b^{(L_\beta)} + W^{(L_\beta)} z_{i,t-1}^{(L_\beta)}. \tag{6}$$

Equation (4) initializes the neural network with the observable asset characteristics, $z_{i,t-1}$. In Equation (5), characteristics are non-linearly transformed while they propagate through the hidden layer neurons. The terminal layer yields the K-dimensional vector of factor betas, as described in Equation (6). The factor specification on the right-hand side in Figure 1 proposes the use of a standard autoencoder. It is given as

$$r_t^{(0)} = r_t, \tag{7}$$

$$r_t^{(l)} = \tilde{g}\left(\tilde{b}^{(l-1)} + \tilde{W}^{(l-1)} r_{t-1}^{(l-1)}\right), l = 1, \ldots, L_f, \tag{8}$$

$$f_t = \tilde{b}^{(L_f)} + \tilde{W}^{(L_f)} r_t^{(L_f)}. \tag{9}$$

The network for the factor specification is initialized with a return vector r_t in Equation (7). In principle, r_t is taken to be the vector of individual asset returns. However, as Gu et al. (2021) point out, the cross-section of individual asset returns is highly unbalanced over time and usually orders of magnitude higher compared to the number of time series observations. I follow Gu et al. (2021) and initialize the factor specification with returns of characteristic-based managed portfolios. Büchner and Kelly (2022) adopt the same approach when modeling option returns via IPCA. Consequently, r_t in Equation (7) is substituted by

$$x_t = Z_{t-1}' r_t / N_{t-1}, \tag{10}$$

where Z_{t-1} denotes the matrix of asset characteristics and N_t the number of assets in the cross-section at time t. The ith entry in the $P \times 1$ vector x_t contains the return to a portfolio which is dynamically managed based on the ith asset characteristic $z_{i,t-1}$. As Gu et al. (2021) point out, initializing with x_t yields three advantages. First, it represents a preliminary dimension reduction, significantly reducing the amount of model parameters. Precisely, it can be viewed as adding a new initial layer to the factor neural network

which reduces the N returns into P neurons, x_t, before proceeding with the actual factor network. Second, it overcomes panel incompleteness, as portfolios based on characteristics can be formed each point in time. Third and finally, it is motivated by the literature on characteristic-managed portfolios (Kelly et al., 2019; Giglio and Xiu, 2021).

Equation (8) describe how returns propagate through the factor model. Finally, the output layer yields the $K \times 1$ dimensional vector f_t (Equation (9)). Note that the number of hidden layers for the beta function, L_β, and the factor specification, L_f, do not necessarily need to be the same. The final operation in Figure 1 connects the factors with the non-linear beta representation.

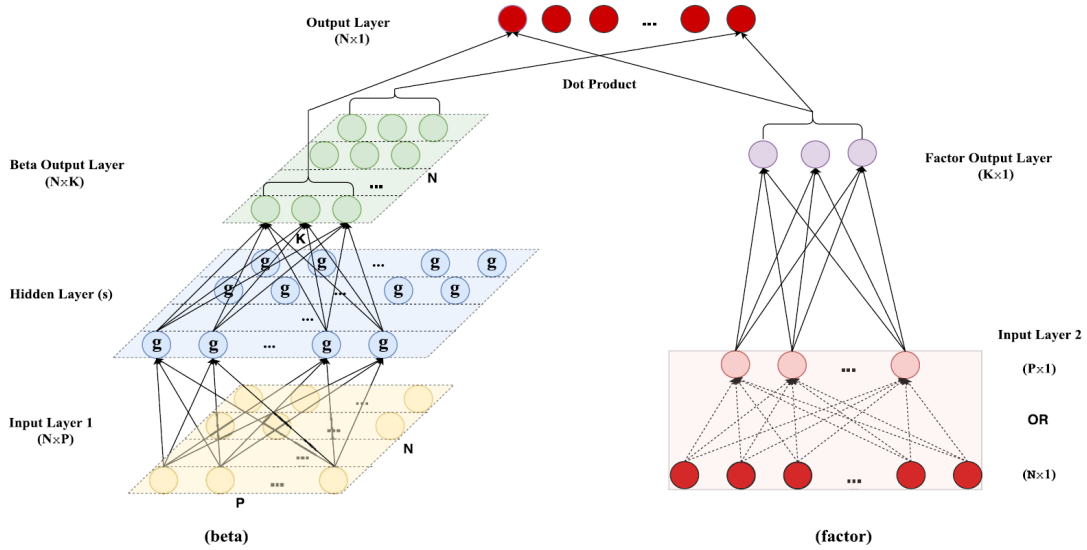


Fig. 1. **Conditional Autoencoder (Gu et al., 2021)**
The figure shows the network architecture of the conditional autoencoder model proposed by Gu et al. (2021). The left-hand side displays the network structure for the factor loadings. The right-hand side depicts the factor model. The figure is taken from Gu et al. (2021).

2.4. Model Set

I compare four different specifications of the conditional autoencoder (CA) with increasing complexity. As a benchmark model, I choose the instrumented principal component analysis (IPCA). The simplest conditional autoencoder, CA0, does not contain any hidden layer in the beta network. It is similar, but not equal to IPCA (Gu et al., 2021). CA1 adds a hidden layer, whereas CA2 adds a second hidden layer in the loadings network. Finally, the most complex model, CA3, exhibits three hidden layers on the beta network. I assume that the number of neurons in the first hidden layer is a hyperpa-

rameter and varies between 4 and 8.[2] A subsequent hidden layer halves the number of neurons compared to the previous layer.

Following Gu et al. (2021), I use a one-layer specification for the factor model, i.e., I assume that the factor network consists of one single layer. Economically speaking, the factors are themselves linear combinations, that is, portfolios of returns. The only source of variation is in the number of neurons, ranging from 1 to 5, representing the number of factors in the conditional autoencoder model.

2.5. Estimation Details

The high flexibility of autoencoders in selecting the most important and informative characteristics can be generally seen as an advantage. It comes, however, with the cost of potential overfitting. In this section, I outline how I try to mitigate overfitting. Thereby, I broadly follow Gu et al. (2021) in estimating the conditional autoencoder model.

Training, Validation, and Testing I split my data set into three non-overlapping periods maintaining the temporal order of the data. More specifically, I start by estimating model parameters on a training sample over five years (January 1996 – December 2000). This subsample is used to estimate the conditional autoencoder to a specific set of hyperparameters, the latter controlling model complexity. The second subsample, the validation sample, is used to tune the hyperparameters and spans the next two years (2001 – 2002). For a given set of hyperparameters, the estimated model is used to construct fitted values for data in the validation sample. Next, I calculate the mean squared error from the validation sample. I repeat this procedure and iteratively search for the set of hyperparameters that minimize the mean squared error in the validation sample. Lastly, I choose those hyperparameter combinations yielding the best fit in the validation sample to assess the predictive power in the one-year testing sample (2003). I replicate this procedure by extending the number of years in the training sample by one year in each iteration, for a total of 18 out-of-sample years (2003 – 2020).

Regularization Techniques I employ three machine learning regularization techniques. The first is called "early stopping". Early stopping makes use of the validation sample in the training. In each step of the optimization algorithm employed on the training data, the hitherto estimated model is evaluated on the validation data. The op-

[2]Section A provides an overview over all hyperparameters.

timization is ended in case the validation error increases. By ending optimization early, the estimated parameters are shrunken towards the initial guesses in the optimization algorithm, alleviating overfitting.

A a second technique, I add a penalty term to the objective function to discourage model complexity which could potentially lead to model overfitting. Precisely, let $\mathcal{L}(\theta;\cdot)$ denote the objective function for optimizing the conditional autoencoder, where θ denotes the weight parameters in the factor and beta networks. It is given as

$$\mathcal{L}(\theta;\cdot) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}||r_{i,t} - \beta'_{i,t-1}f_t||^2 + \phi(\theta;\cdot), \tag{11}$$

where $\phi(\theta;\cdot)$ is a penalty function. Following Gu et al. (2021), I employ LASSO penalization, i.e.,

$$\phi(\theta;\cdot) = \lambda\sum_{j}|\theta_j|. \tag{12}$$

Finally, I adopt an ensemble approach. That is, I use multiple random seeds, i.e., ten, in initializing the neural network estimation. Subsequently, I construct model predictions by averaging estimates from all trained networks, enhancing the stability of my results.

Optimization Algorithms I follow common practice in the machine learning literature in finance and use stochastic gradient descent (SGD) in training the neural network. Instead of using the full training sample to evaluate the gradient, SGD selects a random subset of the data accelerating the optimization routine at the expense of accuracy. The learning rate is a critical hyperparameter in SGD as it controls the step size in the gradient descent. Too small a value can lead to a long training process that could get stuck, while too large a value can lead to a suboptimal set of weights being learned too quickly or an unstable training process. I use the adaptive momentum estimation (ADAM) algorithm of Kingma and Ba (2014) which uses adaptive learning rates for individual parameters.

Besides SGD, I also adapt batch normalization. It is a technique to mitigate internal covariate shifts in which inputs of hidden layers follow different distributions to their analogs in the validation sample. Batch normalization re-centers and re-scales each of the layers' inputs in each training step.

2.6. Performance Evaluation

To identify the success of a given model, I follow Gu et al. (2021) in using standard R^2 statistics. First, I consider the total R^2 measure which assesses the explanatory power of contemporaneous factor realizations and factor loadings, defined as

$$R^2_{total} = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - z'_{i,t}\left(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{f}_{t+1}\right)\right)^2}{\sum_{i,t} r^2_{i,t+1}}. \tag{13}$$

Moreover, I employ the predictive R^2 which indicates how well the model describes panel variation in risk compensation. It is defined as

$$R^2_{pred} = 1 - \frac{\sum_{i,t}\left(r_{i,t+1} - z'_{i,t}\left(\hat{\Gamma}_\alpha + \hat{\Gamma}_\beta \hat{\lambda}\right)\right)^2}{\sum_{i,t} r^2_{i,t+1}}, \tag{14}$$

where $\hat{\lambda}$ is the unconditional time-series mean of the risk factors up to $t-1$.

3. Data and Variable Definitions

My primary data source is OptionMetrics IvyDB, which provides historical prices for all S&P 500 index options. Due to the starting date of this database, my sample covers the period from January 1996 through December 2020.

3.1. Option Filters

I rely on standard filters established in the literature to assure consistency of my analyses (Büchner and Kelly, 2022; Bollerslev et al., 2015).[3] First, I disregard observations for which the bid price is negative and the bid exceeds the ask. Second, I remove options for which no-arbitrage conditions are violated. Third, I retain only options with non-missing implied volatility as reported by OptionMetrics. Fourth, I keep options with positive open interest and time-to-maturities of less than one year. Fifth, I ensure that the absolute forward delta lies in the interval $[0.01, 0.5]$. Sixth, I retain only options with a standard third-Friday expiration. That is, I exclude weekly, quarterly, and LEAPS

[3]To make the investment process as realistic as possible, I apply the filters only at the start of the trade and assume that I have to use prevailing market quotes when the position is unwinded or regard the position as worthless.

options. Finally, I ensure convexity of option prices following Bollerslev et al. (2015). Specifically, I retain only those observations for a given maturity τ, for which the difference between the prices of two neighboring call (put) options with strike price $K_1 < K_2$ is ≤ 0 (≥ 0).

3.2. Option Returns

The main variable of interest is the excess return of buying an option that I delta-hedge on a daily rebalancing schedule.[4] I consider delta-hedged option gains following Bakshi et al. (2003) as the value of a self-financing portfolio consisting of a long option, hedged by a position in the underlying such that the portfolio is locally immune to changes in the stock price. Let $\Pi = \{t = t_0 < \cdots < t_N = t + \tau\}$ denote the partition of the interval from t to $t + \tau$. Assume that the long option position is hedged discretely N times at each of the dates $t_n, n = 0, \ldots, N - 1$. The discrete delta-hedged option gain over the period $[t, t + \tau]$ is given by:

$$\begin{aligned}\Pi(t, t+\tau) = V_{t+\tau} - V_t - \sum_{n=0}^{N-1} \Delta_{V,t_n} \times [S(t_{n+1}) - S(t_n)] \\ - \sum_{n=0}^{N-1} \frac{a_n r_n}{365} \left[V(t_n) - \Delta_{V,t_n} S(t_n)\right], \end{aligned} \tag{15}$$

where V_t denotes the price of the option at time t, r_n is the risk-free rate at t_n, a_n is the number of calendar days between rehedging dates t_n and t_{n+1}, which I set to $a_n = 1$, and Δ_{V,t_n} is the observed delta of the option. I consider investment horizons of one calendar month, or until maturity if the option expiration falls within the investment month.

Finally, I follow Cao and Han (2013) and Cao et al. (2021) and define option returns as:

$$r_{t,t+\tau} = \frac{\Pi(t, t+\tau)}{|\Delta_t S_t - V_t|}. \tag{16}$$

[4]Tian and Wu (2021) estimate that one-time delta-hedging at initiation removes an average of about 70% of the directional risks embedded in the option, whereas daily delta-hedging manages to eliminate upwards of 90% of these risks.

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: All Options (N=115,793)							
Delta-Hedged Return	-0.36	4.75	-2.77	-1.74	-0.82	0.04	1.39
Days to Maturity	124.54	95.85	28.0	56.0	91.0	182.0	301.0
Absolute Moneyness	1.65	1.4	0.23	0.57	1.22	2.39	3.8
Implied volatility	23.39	11.14	11.36	15.49	21.42	28.61	36.99
Absolute Delta	0.17	0.15	0.02	0.05	0.13	0.28	0.4
Panel B: All Options 1996-2002 (N=10,620)							
Delta-Hedged Return	-0.45	2.89	-2.92	-1.83	-0.79	0.41	2.19
Days to Maturity	143.02	102.73	28.0	56.0	119.0	217.0	301.0
Absolute Moneyness	1.07	0.88	0.16	0.39	0.86	1.5	2.32
Implied volatility	24.14	7.46	16.29	18.8	22.64	27.92	34.4
Absolute Delta	0.21	0.15	0.03	0.08	0.19	0.34	0.44
Panel C: All Options 2003-2020 (N=105,173)							
Delta-Hedged Return	-0.35	4.9	-2.75	-1.73	-0.82	0.01	1.31
Days to Maturity	122.68	94.93	28.0	56.0	91.0	182.0	280.0
Absolute Moneyness	1.71	1.43	0.24	0.6	1.27	2.5	3.91
Implied volatility	23.31	11.44	11.12	14.99	21.24	28.68	37.33
Absolute Delta	0.17	0.14	0.02	0.04	0.12	0.27	0.4

Table 1: Delta-Hedged Option Return Summary Statistics

The table reports descriptive statistics for delta-hedged option returns. Panel A reports statistics of returns and option characteristics over the period from 1996 to 2020. Delta-hedged option returns are measured over a period of one calendar month, or until option maturity. Delta-hedging is performed daily. Days-to-maturity is defined as the number of calendar days until option expiration. Moneyness is defined as $m = \log(K/S)/(\sigma^{ATM}\sqrt(\tau))$, where K is the strike price and τ the time to maturity of an option. S denotes the index level of the S&P 500, whereas σ^{ATM} is the at-the-money implied volatility. Option implied volatility is provided by OptionMetrics. Absolute delta follows the model of Black and Scholes (1973). Panel B reports summary statistics for the period 1996 to 2002 used exclusively in the training step of the models. Panel C reports summary statistics for the entire out-of-sample period, comprising the years 2003 through 2020.

3.3. Summary Statistics

My sample comprises 115,793 option-month observations for the period January 1996 – December 2020. Panel A of Table 1 shows that the average monthly delta-hedged option return is slightly negative at -0.36%, whereas the median monthly delta-hedged option return is -0.82%. The average absolute normalized moneyness is 1.65 and the average implied volatility is 23.39%. The average (median) days to maturity are 124 (91), while every fourth option exhibits a time to maturity of fewer than 56 days. Panel B in Table 1 depicts summary statistics for the years which are exclusively used in the training set (1996 – 2002), while Panel C shows summary statistics only for the testing

subsample from 2003 through 2020. Average monthly delta-hedged option returns tend to be slightly less negative for the more recent period. Additionally, the absolute normalized moneyness is higher for the more recent period. However, days to maturity is less for the more recent sample. Implied volatility is comparable across both periods.

Throughout this paper, I divide the time-to-maturity and moneyness domain of options into different regions. Specifically, I divide the time-to-maturity dimension into short-, medium- and long-term options ($\tau \leq 30$ vs. $\tau > 30$ and $\tau \leq 90$ vs. $\tau > 90$ days to maturity). Moreover, I consider at-the-money options (ATM, $|m| \leq 0.5$) and out-of-the-money options (OTM, $|m| > 0.5$), where moneyness m is defined as $m = \log(K/S)/(\sigma^{ATM}\sqrt(\tau))$, where K is the strike price and τ the time to maturity of an option. S denotes the index level of the S&P 500 and σ^{ATM} is the at-the-money implied volatility.

Table 2 shows the number of options falling into the time-to-maturity and moneyness categories, respectively. On average, I observe each month 405 options. Out of these, nearly half are long-term options, 30% are medium-term options and the rest expire within the next month. Short-term options are highly skewed towards OTM options, as two out of three belong to this category. Whereas the skew is decreasing in time to maturity, OTM options remain always over-represented on average.

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Short Term Otm	44.51	40.59	11.0	16.0	33.0	50.0	113.5
Short Term Atm	24.13	23.08	9.0	12.75	18.0	27.0	41.0
Medium Term Otm	76.91	87.37	10.0	16.5	44.0	102.5	193.8
Medium Term Atm	47.78	47.44	11.0	19.0	35.0	55.0	95.0
Long Term Otm	114.99	135.43	28.0	34.5	50.0	130.5	327.8
Long Term Atm	97.17	101.81	35.0	43.0	68.0	102.0	186.2

Table 2: Number of Options per Bucket

The table reports summary statistics on the number of options within certain regions of the time-to-maturity and moneyness domain. Moneyness is defined as $m = \log(K/S)/(\sigma^{ATM}\sqrt(\tau))$, where K is the strike price and τ the time to maturity of an option. S denotes the index level of the S&P 500, whereas σ^{ATM} is the at-the-money implied volatility. The time-to-maturity and moneyness domain is divided into short-, medium- and long-term options (short: $\tau \leq 30$ days to maturity, medium: $\tau > 30$ and $\tau \leq 90$ days to maturity, long: $\tau > 90$ days to maturity), at-the-money ($|m| \leq 0.5$) and out-of-the-money ($|m| > 0.5$) options. The sample period is from January 1996 to December 2020.

3.4. Option Characteristics

I follow Büchner and Kelly (2022) in selecting option-level characteristics. Precisely, I use the time-to-maturity and embedded leverage according to Frazzini and Pedersen (2021) and the Black and Scholes (1973) implied volatility, delta, theta, gamma, vega, and volga.[5] In all specifications, the eight characteristics are interacted with a put-call indicator variable and a constant. Furthermore, all characteristics are re-scaled to the range $[-1, 1]$. Scaling characteristics in this way ensures focusing on the cross-sectional distribution of characteristics and that portfolio weights sum up to zero as it is typical in the asset pricing literature (Kelly et al., 2019).

4. Empirics

4.1. Statistical Performance Evaluation

Table 3 shows the out-of-sample R^2_{total} for the testing sample from January 2003 through December 2020 using the four conditional autoencoders and IPCA. Overall, the best model utilizing the R^2_{total} is the IPCA model with five hidden factors. The R^2_{total} is 91.1%. There are several noteworthy observations in Table 3. First, the explanatory power is generally increasing in the number of factors regardless of the model. This relation is monotonic except for CA2 and CA3. Second, keeping the number of factors fixed in the cross-section of models, IPCA yields the highest R^2_{total}. CA1 comes second and CA0 third. Conditional autoencoders with higher complexity, i.e., with two or more hidden layers, perform worse. Third, the statistical performance is very close across IPCA, CA0, and CA1 for models with two or fewer factors. Whereas the improvement in the R^2_{total} is rather marginal with an increasing number of factors for CA0 and CA1, IPCA continues to substantially increase the R^2_{total} for models with three or four factors (77% for two factors vs. 84.1% for three factors vs. 89.3% for four factors).

Besides using single asset returns, the explanatory power of factor models is commonly assessed via portfolios. Natural candidates are the characteristic-managed portfolios. Given the predictions at the level of single option contracts, I can directly form predictions on managed portfolios by using the individual options' characteristics as these form the basis for the portfolio weights (Gu et al., 2020). Therefore, model refitting is not necessary. As Gu et al. (2021) point out, managed portfolios are extensive aggregations of

[5]Section B provides an overview of the characteristics and their computation.

Model	Test asset	K				
		1	2	3	4	5
CA0	r_t	68.7	75.0	74.2	76.9	77.2
	x_t	87.8	92.2	91.1	93.9	92.1
CA1	r_t	70.2	76.0	76.8	78.4	78.6
	x_t	91.9	90.7	91.1	91.9	91.7
CA2	r_t	66.7	68.4	73.1	71.5	71.7
	x_t	90.5	85.0	88.2	86.4	86.2
CA3	r_t	62.7	70.3	71.0	73.3	70.0
	x_t	89.0	87.5	87.1	89.2	87.4
IPCA	r_t	71.1	77.0	84.1	89.3	91.1
	x_t	92.9	95.7	98.0	99.1	99.5

Table 3: Out-of-sample R^2_{total}

The table reports the out-of-sample R^2_{total} (in %) for individual option contracts (r_t) and characteristics managed portfolios (x_t) using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA0 through CA3. In all models, the number of factors, K, varies from 1 to 5.

Model	Test asset	K				
		1	2	3	4	5
CA0	r_t	0.1	0.3	0.1	0.6	0.5
	x_t	0.9	1.2	0.4	2.0	2.6
CA1	r_t	0.6	0.4	0.3	0.4	0.4
	x_t	2.1	1.8	1.5	1.8	1.8
CA2	r_t	0.7	0.3	0.4	0.4	0.4
	x_t	2.2	1.7	1.8	1.7	1.5
CA3	r_t	0.8	0.4	0.4	0.4	0.4
	x_t	2.3	1.7	2.0	1.8	1.8
IPCA	r_t	0.3	0.1	0.1	-0.1	-0.3
	x_t	1.3	1.0	1.2	0.6	-0.2

Table 4: Out-of-sample R^2_{pred}

The table reports the out-of-sample R^2_{pred} (in %) for individual option contracts (r_t) and characteristics managed portfolios (x_t) using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA0 through CA3. In all models, the number of factors, K, varies from 1 to 5. The sample period is from January 20003 to December 2020.

single assets and consequently, much of the idiosyncratic risk of single assets is averaged out. Hence, higher readings of R^2_{total} are expected. Table 3 confirms this hypothesis. Again, IPCA achieves the highest R^2_{total} irrespective of the chosen number of factors. In all IPCA models, the R^2_{total} is above 90%. It is even higher than 98% for the models with three or more factors. CA0 and CA1 rank again second. Whereas CA0 achieves higher R^2_{total}s for models with higher factors, the R^2_{total} is larger than 90% for all CA1 models which is not the case for CA0. However, even the simplest IPCA model produces a higher R^2_{total} than most conditional autoencoder models irrespective of their number of factors. I employ Diebold and Mariano (1995) tests to determine whether outperformance in terms of higher R^2_{total} is statistically significant at conventional levels. Section C reports results and underpins the dominance of IPCA. Among the conditional autoencoder models, the outperformance of CA1 over the remaining models is generally statistically significant.

Next, Table 4 shows the out-of-sample R^2_{pred} for the IPCA and conditional autoencoder models. Whereas IPCA is the leading framework in terms of R^2_{total}, conditional autoencoder models, and in particular the more complex ones, are dominating in terms of R^2_{pred}. Any deep conditional autoencoder model, i.e., autoencoders with at least one hidden layer, achieves a higher R^2_{pred} irrespective of the selected number of factors and if single options contracts or managed portfolios are chosen as assets. Focusing on individual option contracts as test assets, Table C.2 in the appendix reveals that the underperformance of IPCA is statistically significant. Moreover, CA1 and CA2 statistically outperform CA0 for rather low-dimensional factor models., i.e., where the number of factors is less than three. An outperformance of CA1 over CA2 or vice versa is not apparent. This is especially the case for managed portfolios as test assets.

Why is IPCA the leading framework in terms of R^2_{total}, but lagging employing R^2_{pred}? As Büchner and Kelly (2022) state, IPCA directly targets the R^2_{total} in its objective function. The same applies to the conditional autoencoder models, but their fitting procedure differs. The estimation of IPCA does not require a validation sample. It is fit in-sample and then directly applied to the out-of-sample period. In contrast, the estimation of the conditional autoencoder models makes use of a validation sample. It is used to identify the set of model hyperparameters with the best tradeoff between bias and variance, i.e., the performance in-sample and in a simulated out-of-sample test. Hence, the focus on out-of-sample performance is higher in the estimation of conditional autoencoder models than in IPCA.

Finally, I find that IPCA yields lower readings for R^2_{total} and R^2_{pred} compared to the findings in Büchner and Kelly (2022). There might be three reasons for that. First, the

sample period in Büchner and Kelly (2022) ends in 2017. When restricting my sample period to the same as in Büchner and Kelly (2022), I observe considerably higher readings for the R^2_{total} and R^2_{pred} as shown in Table D.1 in the appendix. Second, my rolling out-of-sample testing framework slightly differs from Büchner and Kelly (2022). Precisely, the out-of-sample period starts in January 2001 in Büchner and Kelly (2022), whereas I start in January 2003. Moreover, Büchner and Kelly (2022) re-estimate the models every month when rolling forward in time, whereas I re-estimate the models every year. Third, I adopt the return convention (Cao and Han, 2013; Cao et al., 2021) in which delta-hedged gains are scaled by the absolute value of capital needed to set up the delta-hedging position. On the contrary, Büchner and Kelly (2022) scale by the level of the S&P 500 index.

Performance in the Cross-Section of Contracts So far, the statistical performance has been assessed by pooling all individual option contracts. Next, I study the performance of the models in different regions of the time-to-maturity and moneyness dimensions and for calls and puts, respectively. Table 5 breaks down the time-to-maturity into three bins: short-term, medium-term, and long-term. The moneyness region is binned into at-the-money (ATM) and out-of-the-money options (OTM). [6] Results for each bin are obtained by pooling all option-month observations falling into the respective bin. As Table 5 displays, IPCA continues to be the dominating model framework irrespective of the time-to-maturity category or if options are grouped by calls and puts in terms of R^2_{total}. The same applies to OTM options, but not ATM options. I observe large relative gains in using shallow conditional autoencoders for ATM options. Precisely, IPCA with one factor yields a negative R^2_{total} for ATM options, whereas CA0 and CA1 yield R^2_{total}s above 30%. Only for high-dimensional factor models, IPCA regains its superior performance. IPCA with five factors yield a R^2_{total} of 70.7%, whereas it is below 55% for conditional autoencoders. It is furthermore noteworthy that the best fit is achieved for longer-term, i.e., medium-term and long-term options, OTM options, and put options. This finding applies to all models under consideration. The relative underpeformance is especially pronounced for ATM and call options. For example, conditional autoencoder models barely reach a R^2_{total} above 40% for call options, but well above 70% for put options. Keeping the number of hidden layers and the number of factors fixed, differences in R^2_{total} across time-to-maturity bins generally do not exceed 10 percentage points for conditional autoencoders. A potential explanation might lie in the over-representation of

[6] See Section 3.3 for a definition of the time-to-maturity and moneyness dimensions.

Model	K				
	1	2	3	4	5
Panel A: Short-term					
CA0	64.1	68.2	68.2	68.7	71.0
CA1	62.2	68.8	69.3	70.1	69.0
CA2	59.3	61.3	64.2	61.1	60.5
CA3	58.7	64.2	61.2	68.2	61.5
IPCA	66.8	70.0	76.3	83.7	84.9
Panel B: Medium-term					
CA0	71.1	77.1	77.1	78.8	79.5
CA1	75.3	82.6	82.5	82.8	83.8
CA2	70.3	74.9	79.4	78.3	76.1
CA3	65.4	77.9	76.9	78.6	74.7
IPCA	72.1	77.5	84.6	91.3	92.9
Panel C: Long-term					
CA0	67.8	75.1	73.4	77.6	76.9
CA1	67.8	71.8	73.5	76.8	76.4
CA2	65.4	64.2	69.8	68.0	70.9
CA3	61.4	64.6	68.2	69.6	68.0
IPCA	71.5	78.7	86.1	89.1	91.2
Panel D: ATM					
CA0	35.3	54.2	45.6	43.8	49.6
CA1	34.7	46.2	49.0	52.0	55.0
CA2	29.2	44.9	48.4	48.3	49.8
CA3	24.0	42.4	42.1	46.7	50.5
IPCA	-1.0	-13.7	27.2	54.6	70.7
Panel E: OTM					
CA0	70.0	75.8	75.4	78.2	78.3
CA1	71.7	77.2	77.9	79.5	79.5
CA2	68.2	69.3	74.1	72.4	72.5
CA3	64.3	71.4	72.1	74.3	70.7
IPCA	74.0	80.6	86.4	90.7	91.9
Panel F: Calls					
CA0	19.0	37.4	38.7	38.4	39.9
CA1	20.1	38.1	37.8	38.1	39.2
CA2	17.0	33.4	38.6	31.4	36.5
CA3	15.1	34.7	33.8	37.5	40.1
IPCA	1.8	32.2	57.7	67.4	75.2
Panel G: Puts					
CA0	78.6	82.5	81.3	84.6	84.6
CA1	80.2	83.6	84.6	86.5	86.4
CA2	76.6	75.4	80.0	79.5	78.7
CA3	72.3	77.4	78.4	80.4	75.9
IPCA	85.0	85.9	89.4	93.7	94.2

Table 5: Out-of-sample R^2_{total} for Options Buckets

The table reports the out-of-sample R^2_{total} (in %) for individual option contracts within certain regions of the time-to-maturity and moneyness domain (see Section 3.3) using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA0 through CA3. In all models, the number of factors, K, varies from 1 to 5. The sample period is from January 2003 to December 2020.

Model	K				
	1	2	3	4	5
Panel A: Short-term					
CA0	0.9	1.1	1.1	1.6	1.1
CA1	1.1	1.1	0.9	0.9	1.1
CA2	1.3	1.0	1.0	1.1	1.0
CA3	1.2	1.1	1.3	1.0	1.1
IPCA	1.1	0.5	0.6	0.2	-0.5
Panel B: Medium-term					
CA0	-0.1	0.0	-0.4	0.2	0.0
CA1	0.3	0.1	0.1	0.0	0.0
CA2	0.5	0.1	0.1	0.1	0.0
CA3	0.6	0.2	0.2	0.2	0.1
IPCA	-0.0	-0.1	-0.1	-0.4	-0.8
Panel C: Long-term					
CA0	0.1	0.3	0.2	0.6	0.8
CA1	0.8	0.4	0.3	0.6	0.6
CA2	0.8	0.4	0.5	0.6	0.5
CA3	0.8	0.3	0.4	0.4	0.5
IPCA	0.4	0.1	0.1	-0.0	0.2
Panel D: ATM					
CA0	-0.6	-0.8	-2.9	-2.4	-1.7
CA1	-0.7	-1.7	-2.3	-1.5	-0.4
CA2	-0.7	-2.0	-2.5	-1.9	-1.8
CA3	-0.8	-2.7	-2.2	-2.3	-2.2
IPCA	-2.1	-1.9	-3.0	-4.5	-6.4
Panel E: OTM					
CA0	0.2	0.4	0.2	0.7	0.6
CA1	0.7	0.5	0.4	0.5	0.5
CA2	0.8	0.4	0.5	0.5	0.5
CA3	0.9	0.5	0.6	0.5	0.5
IPCA	0.4	0.2	0.3	0.0	-0.1
Panel F: Calls					
CA0	-0.2	0.3	-0.8	1.0	-0.3
CA1	0.4	0.1	0.1	0.4	1.0
CA2	0.4	-0.1	0.0	0.2	0.5
CA3	0.4	-0.1	0.1	-0.4	0.0
IPCA	-0.3	-0.6	0.0	-0.6	-0.6
Panel G: Puts					
CA0	0.2	0.3	0.2	0.5	0.7
CA1	0.7	0.4	0.3	0.4	0.3
CA2	0.8	0.4	0.5	0.5	0.4
CA3	0.9	0.5	0.5	0.5	0.5
IPCA	0.4	0.2	0.1	-0.0	-0.3

Table 6: Out-of-sample R^2_{pred} for Options Buckets

The table reports the out-of-sample R^2_{pred} (in %) for individual option contracts within certain regions of the time-to-maturity and moneyness domain (see Section 3.3) using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA0 through CA3. In all models, the number of factors, K, varies from 1 to 5. The sample period is from January 2003 to December 2020.

these bins as shown in Section 3.3. Table 6 confirms my findings in case of R^2_{pred}.

4.2. Economic Performance Evaluation

It is common practice in the asset pricing literature to assess the model performance in economic terms utilizing long-short portfolios (Gu et al., 2020, 2021; Bali, Beckmeyer, Moerke, and Weigert, 2021). Precisely, I study the return and risk-adjusted performances of portfolios based on the models' predictions. For each model and each month in my sample, I sort options into quintiles based on the model's forecasted out-of-sample delta-hedged excess return. Subsequently, I form a long-short portfolio with zero-net-investment by buying the options in the highest quintile and selling options in the lowest quintile. All invested options are delta-hedged. I consider equally-weighted and value-weighted portfolio returns. Value-weighting uses the option's market capitalization (open interest times price) following Büchner and Kelly (2022).

Model	K				
	1	2	3	4	5
Panel A: Equal-Weight					
CA0	0.08	0.22	0.25	0.32	0.30
CA1	0.17	0.17	0.25	0.28	0.29
CA2	0.19	0.11	0.32	0.26	0.28
CA3	0.18	0.20	0.21	0.17	0.19
IPCA	0.15	0.14	0.25	0.26	0.40
Panel B: Value-Weight					
CA0	0.02	0.13	0.30	0.28	0.26
CA1	0.10	0.16	0.22	0.34	0.35
CA2	0.12	0.05	0.27	0.40	0.49
CA3	0.16	0.23	0.24	0.17	0.20
IPCA	0.15	0.20	0.17	0.38	0.56

Table 7: Out-of-Sample Annualized Sharpe Ratios of Long-Short Portfolios

The table reports the out-of-sample annualized Sharpe ratios of long-short portfolios using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA1 through CA3. In all models, the number of factors, K, varies from 1 to 5.

Table 7 reports the annualized Sharpe ratios for all models and across all possible factor specifications. Generally, and consistent with the results on R^2_{pred}, conditional autoencoder models tend to yield higher Sharpe ratio portfolios than IPCA. Although the economic performances differ across the number of factors, CA1 and CA2 emerge as

the overall dominant models. This holds for either equally-weighted (Panel A) or value-weighted portfolios (Panel B). However, Sharpe ratios are usually rather low and below 0.4. A notable exception to the above findings is the IPCA model with five factors which achieves the highest Sharpe ratio (0.4 and 0.56 for equally-weighted and value-weighted portfolios, respectively).

As Section 4.1 has shown, the chosen sample period can have a substantial impact on overall model performance. Table D.2 in the Appendix reports annualized Sharpe ratios for all models and factor specifications when the sample ends in 2017, coinciding with the sample end in Büchner and Kelly (2022). As the table shows, Sharpe ratios are in all cases higher over the shorter sample period. Precisely, trading on CA1's predictions using value-weights yields the highest Sharpe ratio (0.76). Moreover, the outperformance of conditional autoencoder models over IPCA is more pronounced over this time.

Details on the Performance of Long-Short Portfolios Büchner and Kelly (2022) argue in favor of three factors when choosing among IPCA models for S&P 500 index options. In the remainder, I analyze the performance of long-short portfolios following by predictions of models with three factors.

Table 8 reports the predicted and realized delta-hedged option returns for the quintile and long-short portfolios for CA0, CA1, CA2, and IPCA.[7] Panel A reports results for equally-weighted portfolios, whereas Panel B for value-weighted portfolios. By construction, the predicted returns are increasing in the quintile portfolios. With few occasions for value-weights, this also holds for realized delta-hedged option returns. However, over the full sample period, only the value-weighted long-short return for CA2 is statistically significant from zero.[8] On the one hand, Table 8 confirms the superiority of the CA2 three-factor model, as also the Sharpe ratio is statistically significantly different from zero. On the other hand, the table shows that also CA0 and CA1 yield average returns of the long-short portfolio which are statistically significantly different from zero.

To further investigate the performance drivers in the out-of-sample long-short portfolios, I use the recently proposed performance attribution via Shapley values (Moehle, Boyd, and Ang, 2021). Contrary to other performance attribution methods such as one-at-a-time, sequential, leavel-one-out or permuted sequential, the Shapley attribution is the only method obeying desirable attribution methods.[9]

[7] I omit CA3 out of brevity as its performance is inferior to the remaining conditional autoencoder models.

[8] Table D.3 in the Appendix repeats the above analysis over the shorter sample period ending in 2017.

[9] These properties are full attribution, baseline correctness, fairness, and monotonicity. Section E

Model		Low	2	3	4	High	H-L	
Panel A: Equal-Weight								
CA0	Predicted	-1.02	-0.70	-0.49	-0.31	-0.04	0.98	
	Actual	-0.57	-0.47	-0.45	-0.38	-0.24	0.32	
	Std	5.58	3.88	2.52	2.00	1.79	4.49	
	SR	-0.35	-0.42	-0.62	-0.66	-0.47	0.25	
CA1	Predicted	-0.95	-0.61	-0.41	-0.27	-0.13	0.82	
	Actual	-0.61	-0.50	-0.44	-0.31	-0.26	0.35	
	Std	6.00	3.54	2.45	2.03	1.88	4.89	
	SR	-0.35	-0.49	-0.62	-0.52	-0.49	0.25	
CA2	Predicted	-0.85	-0.58	-0.42	-0.32	-0.23	0.63	
	Actual	-0.59	-0.60	-0.46	-0.33	-0.13	0.46	*
	Std	6.25	3.39	2.33	1.96	2.14	5.00	
	SR	-0.33	-0.62	-0.69	-0.58	-0.22	0.32	
IPCA	Predicted	-1.19	-0.64	-0.40	-0.22	0.03	1.21	
	Actual	-0.55	-0.48	-0.48	-0.33	-0.28	0.27	
	Std	5.45	3.06	2.34	2.38	2.57	3.87	
	SR	-0.35	-0.54	-0.71	-0.48	-0.37	0.25	
Panel B: Value-Weight								
CA0	Predicted	-0.97	-0.69	-0.49	-0.31	-0.03	0.94	
	Actual	-0.51	-0.40	-0.34	-0.28	-0.17	0.34	
	Std	4.98	3.59	2.74	2.18	1.58	3.88	
	SR	-0.35	-0.38	-0.43	-0.44	-0.38	0.30	
CA1	Predicted	-0.88	-0.60	-0.41	-0.26	-0.13	0.75	
	Actual	-0.45	-0.46	-0.33	-0.24	-0.19	0.27	
	Std	5.42	3.18	2.70	1.99	1.73	4.24	
	SR	-0.29	-0.50	-0.42	-0.41	-0.37	0.22	
CA2	Predicted	-0.81	-0.57	-0.42	-0.32	-0.23	0.58	
	Actual	-0.50	-0.52	-0.40	-0.30	-0.14	0.36	
	Std	5.80	3.10	2.26	1.66	1.68	4.69	
	SR	-0.30	-0.58	-0.62	-0.63	-0.28	0.27	
IPCA	Predicted	-1.12	-0.63	-0.39	-0.21	0.01	1.13	
	Actual	-0.38	-0.22	-0.36	-0.25	-0.21	0.17	
	Std	4.83	3.69	2.16	2.16	2.14	3.32	
	SR	-0.27	-0.21	-0.58	-0.41	-0.34	0.17	

Table 8: Performance of Out-Of-Sample Long-Short Portfolios

The table shows the returns to option portfolios sorted by the predictions of the instrumented principal component model (IPCA) and the conditional autoencoder models CA1 and CA2. Portfolios are either equally-weighted (Panel A) or value-weighted (Panel B). Predicted denotes the model's predicted returns. Actual denotes the realized returns. SD is the standard deviation of realized returns. SR denotes the realized annual Sharpe ratio. The last column indicates whether the realized returns (Sharpe ratios) of high-minus-low (H-L) portfolio is significantly different from zero. In all models, the number of factors, K, is fixed to three.

Model	Subset	Low	High	H-L
CA0	Short Term ATM	-0.02	-0.05	-0.03
	Short Term OTM put	-0.17	-0.10	0.07
	Short Term OTM call	-0.18	-0.11	0.07
	Medium Term ATM	-0.01	0.03	0.03
	Medium Term OTM call	-0.09	-0.08	0.01
	Medium Term OTM put	-0.21	-0.07	0.13
	Long Term ATM	0.05	0.22	0.18
	Long Term OTM call	-0.06	-0.07	-0.01
	Long Term OTM put	0.12	-0.00	-0.12
	Total	-0.57	-0.24	0.32
CA1	Short Term ATM	-0.03	-0.05	-0.02
	Short Term OTM put	-0.20	-0.08	0.12
	Short Term OTM call	-0.17	-0.10	0.07
	Medium Term ATM	0.01	0.02	0.01
	Medium Term OTM call	-0.13	-0.07	0.07
	Medium Term OTM put	-0.13	-0.05	0.07
	Long Term ATM	0.03	0.17	0.14
	Long Term OTM call	-0.08	-0.05	0.03
	Long Term OTM put	0.08	-0.06	-0.14
	Total	-0.61	-0.26	0.35
CA2	Short Term ATM	-0.04	-0.07	-0.02
	Short Term OTM put	-0.20	-0.07	0.13
	Short Term OTM call	-0.17	-0.12	0.05
	Medium Term ATM	0.00	0.05	0.05
	Medium Term OTM call	-0.11	-0.10	0.01
	Medium Term OTM put	-0.10	-0.07	0.04
	Long Term ATM	0.03	0.21	0.18
	Long Term OTM call	-0.07	-0.04	0.04
	Long Term OTM put	0.07	0.06	-0.01
	Total	-0.59	-0.13	0.46
IPCA	Short Term ATM	-0.03	-0.06	-0.02
	Short Term OTM put	-0.16	-0.10	0.06
	Short Term OTM call	-0.18	-0.10	0.09
	Medium Term ATM	0.02	0.03	0.01
	Medium Term OTM call	-0.11	-0.07	0.05
	Medium Term OTM put	-0.06	-0.08	-0.02
	Long Term ATM	0.02	0.13	0.11
	Long Term OTM call	-0.09	-0.03	0.06
	Long Term OTM put	0.05	-0.01	-0.06
	Total	-0.55	-0.28	0.27

Table 9: Performance Attribution of Out-Of-Sample Long-Short Portfolios

The table shows the performance attribution of out-of-sample long-short portfolios (equally-weighted) using the instrumented principal component analysis (IPCA) and the conditional autoencoder models CA0, CA1, and CA2. In all models, the number of factors is fixed to three. The out-of-sample performance for the short (Low), long (High), and long-short (H-L) are reported. Performance attribution is done with respect to the option subsets defined in Section 3 via Shapley values (Moehle et al., 2021).

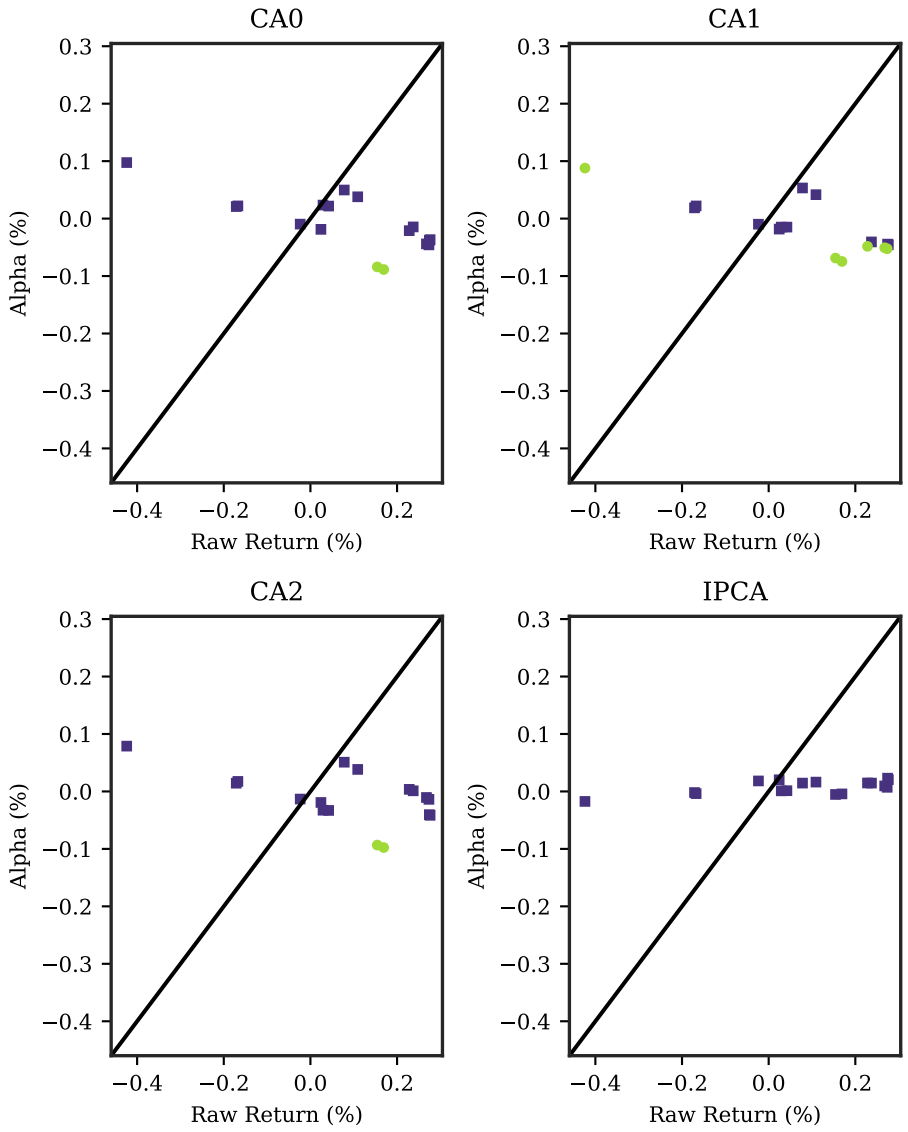


Fig. 2. **Out-of-sample Pricing Errors across Models**
The figure shows the out-of-sample pricing errors (alphas) for 17 characteristic-managed portfolios x_t, relative to the conditional autoencoders CA1 and CA2 and the instrumented principal component analysis (IPCA). The factors in all models are fixed to three.

Table 9 reports the performance attribution with respect to the option subsets defined in Section 3 for equally-weighted portfolios for CA0 to CA2 and IPCA.[10] Short-term OTM options and long-term ATM options generate the highest performance attribution, irrespectively of the model considered. This finding reflects the results in Table 5 but in economic terms. Medium-term OTM options contribute also greatly to the performance of long-short portfolios, however, mainly for shallow conditional autoencoder models. For the same class of models, long-term OTM put options are detrimental to performance, in contrast to CA2 and IPCA.

4.3. Risk Premia vs. Mispricing

IPCA allows the researcher to answer whether characteristics are predicting excess option returns through a mispricing channel or a risk compensation channel by setting $\Gamma_\alpha = 0$ in Equation (1). In contrast, the conditional autoencoder model is specified without an intercept in the non-linear factor model. Hence, it imposes that characteristics can only predict returns through systematic exposure. Economically speaking, the conditional autoencoder model enforces no-arbitrage (Gu et al., 2021).

To test whether the assumption of no-arbitrage is supported by the data, I adopt the pricing error tests suggested by Gu et al. (2021) and Büchner and Kelly (2022). Precisely, I calculate the unconditional out-of-sample pricing errors as the average difference between observed and model-predicted option returns. As in Büchner and Kelly (2022), I use the characteristic-managed portfolios instead of returns of single option contracts. As Gu et al. (2021) point out, these pricing errors can be understood as the average gain of a hedge portfolio having zero systematic exposure. If no-arbitrage holds, these hedge portfolios should not earn any excess return.

Figure 2 plots the average out-of-sample pricing errors for each characteristic-managed portfolio in the case of IPCA, CA0, CA1, and CA2, where the number of factors is fixed to three. The figure also shows if alphas are statistically significant (t-value above 2 in absolute terms).

Table 10 reports average out-of-sample pricing errors. The figure also highlights alphas of managed portfolios whose associated t-statistics are above two in absolute terms. The overall size of the alphas is comparable across the conditional autoencoder models CA0 to CA3. As Table 10 report, they amount to a monthly absolute average of 0.04%. Moreover,

describes the Shapley attribution in detail.

[10]Table E.1 in Section E reports results for value-weighted portfolios.

the alphas of the managed portfolios using embedded leverage (either for calls or puts) as the characteristic are statistically significant for CA0 to CA3. I observe the highest number of statistically significant alphas for CA1. Besides embedded leverage, CA1 yields significant alphas for three more characteristic-based portfolios (time-to-maturity for calls, and vega for calls and puts). In contrast, IPCA does not yield any significant alpha at the 5% percent level or below.

	CA0		CA1		CA2		IPCA	
	alpha	t-stat	alpha	t-stat	alpha	t-stat	alpha	t-stat
delta	-0.02	-1.00	-0.02	-1.03	-0.02	-1.12	0.02	1.45
delta:put	-0.01	-0.50	-0.01	-0.58	-0.01	-0.82	0.02	1.27
embedlev	-0.09	-2.18**	-0.07	-2.19**	-0.10	-2.49**	-0.00	-0.29
embedlev:put	-0.08	-2.14**	-0.07	-2.06**	-0.09	-2.43**	-0.01	-0.37
gamma	-0.04	-1.22	-0.04	-1.66*	-0.04	-1.28	0.02	1.94*
gamma:put	-0.04	-1.15	-0.05	-1.66*	-0.04	-1.30	0.02	1.84*
iv	0.04	1.21	0.04	1.65*	0.04	1.37	0.02	1.21
iv:put	0.05	1.45	0.05	1.79*	0.05	1.55	0.01	1.06
theta	0.02	0.88	0.02	0.96	0.02	0.76	-0.00	-0.21
theta:put	0.02	0.79	0.02	0.77	0.01	0.60	-0.00	-0.11
ttm	-0.02	-0.80	-0.05	-2.03**	0.00	0.21	0.01	1.19
ttm:put	-0.01	-0.48	-0.04	-1.87*	0.00	0.06	0.01	1.13
vega	-0.04	-1.87*	-0.05	-2.55**	-0.01	-0.49	0.01	0.89
vega:put	-0.05	-1.83*	-0.05	-2.64***	-0.01	-0.63	0.01	0.67
volga	0.02	0.42	-0.01	-0.39	-0.03	-0.76	0.00	0.07
volga:put	0.02	0.46	-0.01	-0.38	-0.03	-0.74	0.00	0.06
Avg. Abs. Alpha	0.04		0.04		0.04		0.01	

Table 10: Out-of-Sample Managed Portfolio Alphas

The table shows the individual characteristic-managed portfolios alphas in the instrumented principal component analysis model (IPCA) and the conditional autoencoder models CA1 and CA2. The number of factors is fixed to three. The portfolio alphas are obtained as time-series averages of the model residuals.

4.4. Characteristics Importance

Next, I analyze the importance of single characteristics. Ideally, I measure the importance of each characteristic by excluding it and re-estimating the models. As this procedure is computationally very intense, I adopt instead the approach of Lundberg and Lee (2017) who suggest using SHAPs (SHapley Additive exPlanations). SHAP is a game-theoretic approach to understanding the output of any machine learning model and can be used to gauge the importance of characteristics without re-fitting.

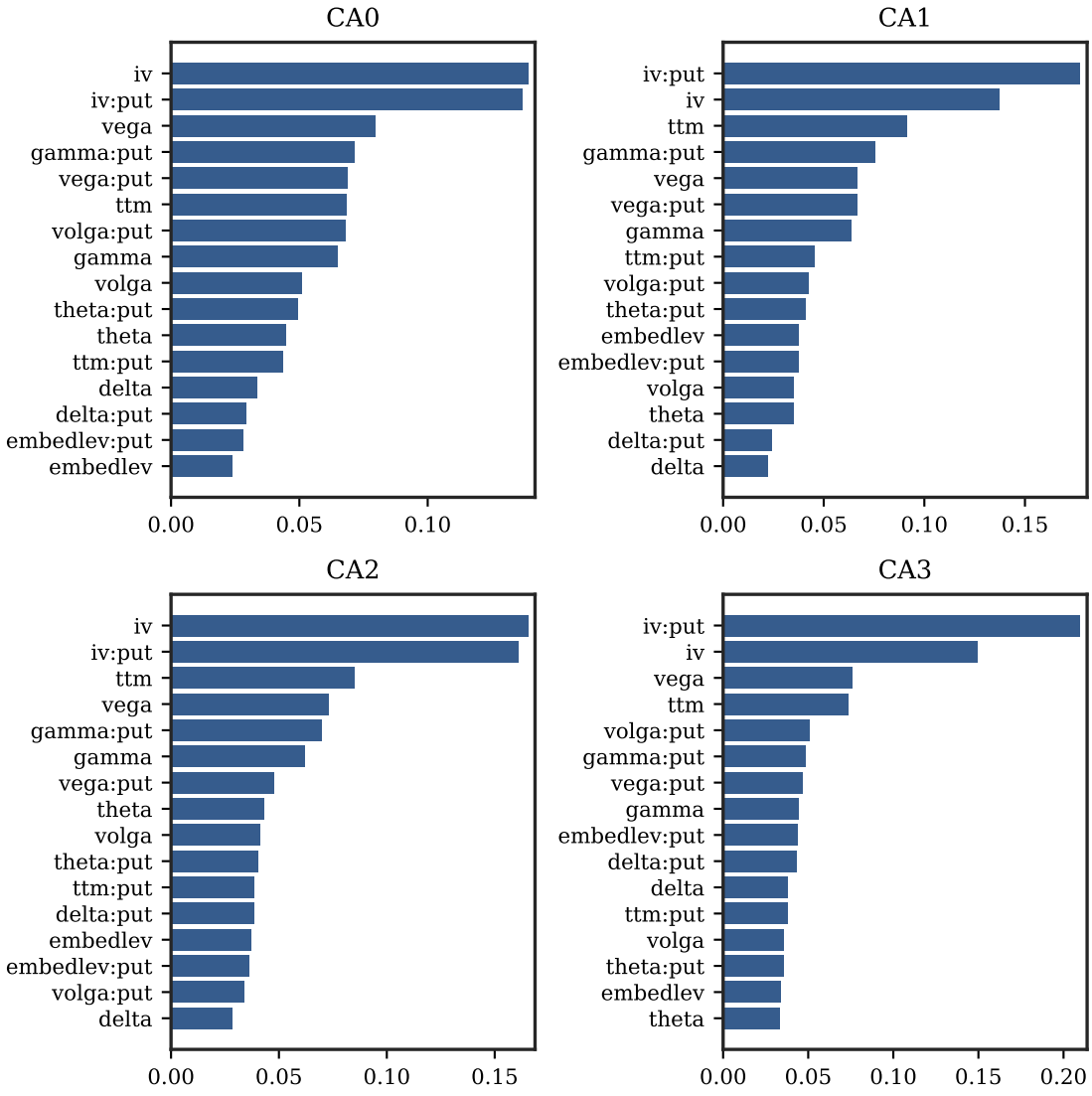


Fig. 3. **Characteristics Importance**
The figure shows the importance of single characteristics as estimated by SHAP values (Lundberg and Lee, 2017) for the conditional autoencoder models CA0 to CA3. The values are scaled such that they sum to one. The bars represent the mean feature group importance for the entire testing sample. The factors in all models are fixed to three.

Figure 3 shows the importance of the characteristics in the conditional autoencoder models CA0 to CA3, where the number of factors is fixed to three. Several noteworthy observations emerge from the figure. First, the most important characteristic is implied volatility. Its total contribution, summed over calls and puts, amounts to at least 30%. Second, my results are closely related, but not equal to the findings of Büchner and Kelly (2022) who report that implied volatility, vega, and gamma are the most important characteristics in IPCA models for S&P 500 index options. Precisely, whereas gamma and vega are also among the five most important characteristics in the conditional autoencoder models, time-to-maturity emerges as another highly important variable in my results. It is on par with gamma in CA0, but gains in importance for models with hidden layers and ranks among the top four therein. Contrarily, Büchner and Kelly (2022) reports it on the sixth rank, well behind implied volatility, vega, and gamma. This suggests that time-to-maturity rises in importance due to non-linearities and/or interaction effects. Additionally, theta seems to play an important role in findings of Büchner and Kelly (2022). In CA0 to CA3, theta plays a minuscule role and ranks in the bottom half of all characteristics. Its effect might be approximated by an interaction of other characteristics, including time-to-maturity, in the conditional autoencoder models.

5. Conclusion

In this study, I adopt the recently proposed conditional autoencoder factor model of Gu et al. (2021) and apply it to modeling S&P 500 index option returns. In contrast to instrumented principal component analysis, the former allows for a non-linear mapping between characteristics and factor exposures but inherits the latent structure and time-variation in factor loadings from the latter. Whereas conditional autoencoder models are inferior to instrumented principal component analysis in the explanatory power of contemporaneous factor realizations, they excel in purely predictive settings in statistical and economic terms. Trading on their option return predictions yields higher monthly average returns and higher risk-adjusted performance measures.

Appendix

Appendix A Hyperparameters

Table A.1 display the hyperparameters of the conditional autoencoder and their respective search spaces, if applicable.

Max epochs	150
Random search trials	128
Learning rate	$\in [0.0001, 0.001, 0.01, 0.1]$
Weight decay	$\mathcal{U}(0, 0.1)$
λ	$\mathcal{U}(1e^{-8}, 1e^{-3})$
Number of neurons in first hidden layer	$\in [4, 8]$

Table A.1: Hyperparameters for the Conditional Autoencoder

The table shows the hyperparameters of the conditional autoencoder model and their search spaces, if applicable. $\mathcal{U}$ refers to randomly drwaing from a uniform distribution.

Appendix B Option Characteristics

This section describes a broad set of option-level characteristics, motivated by Büchner and Kelly (2022). All of the below characteristics are interacted with dummy variable indicating one for put and zero for call options and a constant.

Time-to-maturity (*ttm*) The number of calendar years to maturity.

Implied volatility (*iv*) The Black and Scholes (1973) implied volatility of the option contract.

Delta (*delta*) The Black and Scholes (1973) delta of the option contract, i.e., the sensitivity of the option with respect to point-changes in the underlying.

Gamma (*gamma*) The Black and Scholes (1973) gamma of the option contract, i.e., the sensitivity of Δ with respect to changes in the underlying.

Theta (*theta*) The Black and Scholes (1973) theta, i.e., the time-decay of the option value.

Vega (*vega*) The Black and Scholes (1973) vega, i.e., the sensitivity of the option with respect to changes in the implied volatility.

Volga (*volga*) The Black and Scholes (1973) volga, i.e, the sensitivity of vega with respect to changes in the implied volatility.

Embedded leverage (*embedlev*) Following Frazzini and Pedersen (2021), the embedded leverage of the option contract, defined as

$$\Omega = \frac{S}{O} \times |\Delta|,$$

where S denotes the index level of the S&P 500, O the options price and Δ the delta of the option.

Appendix C DM Forecast Comparisons

C.1 DM Forecast Comparison for R^2_{total} for Individual Option Contracts

		CA0				CA1					CA2					IPCA				
		2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	**10.5**	**11.9**	**8.2**	**9.9**	**2.3**	**5.9**	**6.1**	**6.9**	**7.5**	**-2.8**	-0.2	**4.3**	**2.1**	**2.9**	**2.2**	**6.0**	**9.8**	**10.0**	**10.9**
	2		**-3.0**	**3.6**	**5.8**	**-5.6**	0.9	1.7	**3.2**	**3.6**	**-7.1**	**-3.9**	-1.7	**-2.4**	**-2.9**	**-4.2**	1.7	**7.6**	**8.9**	**10.0**
CA0	3			**3.9**	**5.7**	**-5.3**	1.6	**2.3**	**3.6**	**4.1**	**-7.2**	**-3.8**	-1.1	**-2.0**	**-2.5**	**-3.2**	**2.3**	**7.7**	**8.7**	**9.8**
	4				1.2	**-5.9**	-0.8	-0.1	1.8	**2.0**	**-6.5**	**-4.3**	**-3.0**	**-3.3**	**-3.8**	**-7.1**	0.1	**7.8**	**9.9**	**11.2**
	5					**-7.0**	-1.1	-0.4	1.5	1.7	**-7.3**	**-4.8**	**-3.6**	**-3.7**	**-4.4**	**-7.5**	-0.2	**6.8**	**9.1**	**10.4**
	1						**6.1**	**6.3**	**6.8**	**7.6**	**-4.1**	-1.6	**4.2**	1.2	1.7	0.9	**5.0**	**8.7**	**9.6**	**10.5**
	2							2.0	**3.8**	**4.0**	**-5.9**	**-5.6**	**-4.2**	**-5.3**	**-4.5**	**-4.3**	0.9	**6.1**	**8.9**	**10.0**
CA1	3								**4.7**	**5.1**	**-5.9**	**-5.3**	**-4.5**	**-4.9**	**-4.8**	**-5.4**	0.2	**6.1**	**9.5**	**10.8**
	4									0.4	**-6.2**	**-5.4**	**-4.9**	**-5.2**	**-5.2**	**-7.4**	-1.4	**5.4**	**9.9**	**11.4**
	5										**-6.6**	**-5.8**	**-5.8**	**-5.6**	**-5.9**	**-7.7**	-1.5	**5.5**	**9.4**	**10.8**
	1											1.6	**5.5**	**3.6**	**4.4**	**2.8**	**5.6**	**8.2**	**8.7**	**9.4**
	2												**5.4**	**4.5**	**4.2**	1.4	**4.3**	**6.9**	**7.8**	**8.5**
CA2	3													**-2.8**	**-3.1**	-1.5	**2.7**	**7.0**	**8.5**	**9.4**
	4														0.3	-0.2	**3.3**	**6.6**	**8.0**	**8.8**
	5															-0.4	**3.5**	**7.3**	**8.4**	**9.2**
	1																**6.8**	**11.4**	**11.9**	**13.1**
	2																	**8.2**	**9.4**	**10.8**
IPCA	3																		**6.3**	**8.8**
	4																			**6.0**

Table C.1: DM Forecast Comparison for R^2_{total} for Individual Option Contracts

The table shows the Diebold and Mariano (1995) test statistics for the instrumented principal component analysis (ICPA) and conditional autoencoder models CA1 and CA2. A positive number indicates that the model in the column outperforms the row model. The numbers are printed in bold if the outperformance is statistically significant at the 5% level. In all models, the number of factors, K, varies from 1 to 5.

C.2 DM Forecast Comparison for R^2_{pred} for Individual Option Contracts

		CA0				CA1					CA2					IPCA				
		2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	**4.0**	-0.7	**3.8**	**4.0**	**10.5**	**4.7**	**2.3**	**3.0**	**3.3**	**9.4**	**3.2**	**4.3**	**4.2**	**3.9**	**3.0**	-0.6	-0.1	**-2.3**	**-3.0**
	2		**-3.7**	**2.9**	**2.1**	**6.8**	1.4	-0.3	1.2	1.6	**5.7**	0.7	1.8	**2.1**	1.5	-0.1	**-2.9**	-1.8	**-4.0**	**-4.8**
CA0	3			**5.9**	**4.5**	**6.5**	**3.7**	**3.3**	**4.2**	**4.6**	**5.9**	**2.8**	**4.0**	**4.5**	**4.4**	**2.7**	0.1	0.5	**-2.0**	**-3.3**
	4				-0.5	0.6	**-2.1**	**-3.1**	-1.9	-1.4	1.2	**-2.2**	**-2.0**	-1.7	**-2.2**	**-2.2**	**-3.6**	**-4.2**	**-5.2**	**-6.7**
	5					1.2	-1.4	**-2.5**	-1.2	-0.9	1.6	-1.3	-1.1	-0.8	-1.3	**-2.6**	**-4.2**	**-3.8**	**-6.5**	**-6.4**
	1						**-6.0**	**-5.2**	**-2.7**	**-2.1**	1.8	**-5.2**	**-4.7**	**-3.8**	**-4.3**	**-4.6**	**-5.9**	**-4.5**	**-6.0**	**-6.3**
	2							-1.6	0.4	0.8	**5.8**	-0.6	0.6	1.3	0.1	-0.9	**-3.4**	**-2.5**	**-4.5**	**-5.0**
CA1	3								**3.2**	**3.3**	**4.9**	0.9	**2.2**	**3.6**	**2.3**	0.1	**-2.7**	**-2.4**	**-5.4**	**-6.1**
	4									1.0	**3.0**	-0.5	-0.1	0.5	-0.4	-1.1	**-3.4**	**-4.6**	**-8.0**	**-8.4**
	5										**2.6**	-0.9	-0.6	-0.2	-0.9	-1.4	**-3.7**	**-4.7**	**-8.6**	**-8.5**
	1											**-6.8**	**-5.3**	**-4.3**	**-4.7**	**-4.0**	**-5.4**	**-4.5**	**-6.0**	**-6.2**
	2												1.3	1.6	0.5	-0.5	**-2.6**	**-2.1**	**-4.0**	**-4.6**
CA2	3													1.0	-0.5	-1.1	**-3.5**	**-2.8**	**-4.9**	**-5.5**
	4														-1.6	-1.4	**-3.6**	**-3.6**	**-5.8**	**-6.3**
	5															-0.9	**-3.4**	**-2.9**	**-5.4**	**-6.0**
	1																**-4.0**	-1.5	**-3.7**	**-4.5**
	2																	0.4	**-2.1**	**-3.2**
IPCA	3																		**-4.6**	**-5.8**
	4																			**-2.6**

Table C.2: DM Forecast Comparison for R^2_{pred} for Individual Option Contracts

The table shows the Diebold and Mariano (1995) test statistics for the instrumented principal component analysis (ICPA) and conditional autoencoder models CA1 and CA2. A positive number indicates that the model in the column outperforms the row model. The numbers are printed in bold if the outperformance is statistically significant at the 5% level. In all models, the number of factors, K, varies from 1 to 5.

C.3 DM Forecast Comparison for R^2_{total} for Managed Portfolios

		CA0				CA1					CA2					IPCA				
		2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	**3.5**	**2.8**	**3.6**	**2.7**	**2.5**	1.2	1.3	2.0	1.7	1.3	-0.5	0.1	-0.3	-0.3	**2.6**	**3.8**	**4.2**	**4.0**	**4.1**
	2		-1.4	**2.3**	-0.1	-0.2	-0.6	-0.5	-0.2	-0.3	-0.8	-1.1	-0.9	-1.1	-1.1	0.7	**3.2**	**4.4**	**4.1**	**4.3**
CA0	3			**2.5**	1.1	0.7	-0.2	-0.0	0.6	0.4	-0.4	-1.0	-0.8	-1.0	-1.0	1.6	**3.7**	**3.9**	**3.6**	**3.7**
	4				**-3.2**	-1.2	-1.1	-1.0	-1.0	-0.9	-1.3	-1.3	-1.2	-1.3	-1.3	-0.8	1.6	**3.8**	**3.9**	**4.0**
	5					-0.1	-0.5	-0.4	-0.1	-0.2	-0.7	-1.1	-0.9	-1.0	-1.0	0.8	**3.0**	**4.6**	**4.4**	**4.5**
	1						-0.8	-0.5	-0.0	-0.2	-1.4	-1.2	-1.1	-1.2	-1.2	1.0	**3.2**	**3.3**	**3.0**	**3.2**
	2							1.0	1.0	1.3	-0.2	-1.3	-1.3	-1.4	-1.4	1.0	**2.2**	**2.4**	**2.4**	**2.4**
CA1	3								0.8	1.2	-0.5	-1.4	-1.3	-1.4	-1.5	0.9	**2.2**	**2.4**	**2.3**	**2.4**
	4									-0.3	-0.9	-1.3	-1.2	-1.3	-1.3	0.8	**2.6**	**2.9**	**2.7**	**2.8**
	5										-0.8	-1.4	-1.4	-1.4	-1.5	0.7	**2.2**	**2.5**	**2.4**	**2.5**
	1											-1.1	-0.8	-1.1	-1.1	1.3	**2.5**	**2.7**	**2.6**	**2.7**
	2												1.4	1.1	1.1	1.3	1.7	1.8	1.8	1.9
CA2	3													-1.4	-1.5	1.2	1.8	**2.0**	**2.0**	**2.0**
	4														-0.4	1.3	1.8	1.9	1.9	2.0
	5															1.3	1.8	1.9	1.9	**2.0**
	1																**3.9**	**4.1**	**3.6**	**3.7**
	2																	**2.6**	**2.4**	**2.6**
IPCA	3																		1.8	**2.2**
	4																			1.9

Table C.3: DM Forecast Comparison for R^2_{total} for Managed Portfolios

The table shows the Diebold and Mariano (1995) test statistics for the instrumented principal component analysis (ICPA) and conditional autoencoder models CA1 and CA2. A positive number indicates that the model in the column outperforms the row model. The numbers are printed in bold if the outperformance is statistically significant at the 5% level. In all models, the number of factors, K, varies from 1 to 5.

C.4 DM Forecast Comparison for R^2_{pred} for Managed Portfolios

		CA0				CA1					CA2					IPCA				
		2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
	1	1.4	-1.1	1.4	**3.1**	**3.8**	**4.0**	**2.7**	**2.8**	**3.0**	**4.2**	**3.4**	**4.7**	**3.2**	**3.3**	0.0	0.0	0.0	-0.0	-0.1
	2		**-2.2**	1.2	**2.3**	**3.1**	**2.5**	1.4	**2.4**	**2.1**	**3.0**	1.9	**2.9**	**2.5**	1.9	0.0	-0.0	-0.0	-0.0	-0.1
CA0	3			**2.2**	**2.6**	**3.2**	**2.8**	**2.3**	**2.6**	**2.5**	**3.2**	**2.5**	**3.0**	**2.6**	**2.6**	0.1	0.0	0.0	0.0	-0.0
	4				0.7	0.2	-0.2	-0.7	-0.3	-0.3	0.4	-0.3	-0.2	-0.4	-0.6	-0.0	-0.1	-0.0	-0.1	-0.1
	5					-1.0	-1.6	**-2.3**	-1.8	-1.8	-0.8	-1.9	-1.6	-1.8	**-2.1**	-0.1	-0.1	-0.1	-0.1	-0.2
	1						-1.4	**-2.5**	-1.2	-0.9	0.8	**-2.1**	-1.0	**-2.5**	**-2.0**	-0.0	-0.1	-0.1	-0.1	-0.1
	2							**-2.3**	-0.3	-0.3	**2.2**	-1.4	-0.1	-1.4	-1.6	-0.0	-0.1	-0.0	-0.1	-0.1
CA1	3								1.9	1.6	**2.8**	1.4	**2.9**	1.4	0.4	-0.0	-0.0	-0.0	-0.1	-0.1
	4									-0.0	1.5	-0.2	0.2	-0.7	-1.2	-0.0	-0.0	-0.0	-0.1	-0.1
	5										1.4	-0.2	0.3	-0.5	-1.4	-0.0	-0.0	-0.0	-0.1	-0.1
	1											**-3.0**	-1.6	**-2.5**	**-2.4**	-0.1	-0.1	-0.1	-0.1	-0.2
	2												0.6	-0.4	-0.9	-0.0	-0.0	-0.0	-0.1	-0.1
CA2	3													-1.0	**-2.2**	-0.0	-0.1	-0.0	-0.1	-0.1
	4														-0.9	-0.0	-0.0	-0.0	-0.1	-0.1
	5															-0.0	-0.0	-0.0	-0.1	-0.1
	1																-0.7	-0.3	-1.8	**-2.6**
	2																	0.5	-1.5	-2.0
IPCA	3																		-1.4	**-2.1**
	4																			-1.0

Table C.4: DM Forecast Comparison for R^2_{pred} for Managed Portfolios

The table shows the Diebold and Mariano (1995) test statistics for the instrumented principal component analysis (ICPA) and conditional autoencoder models CA1 and CA2. A positive number indicates that the model in the column outperforms the row model. The numbers are printed in bold if the outperformance is statistically significant at the 5% level. In all models, the number of factors, K, varies from 1 to 5.

Appendix D Performance Assessment for 2003 – 2017

D.1 Out-of-sample R^2s

Model	Test asset	K				
		1	2	3	4	5
Panel A: R^2_{total}						
CA0	r_t	62.7	67.2	66.9	69.8	68.6
	x_t	83.7	89.9	89.7	92.1	89.4
CA1	r_t	70.4	72.7	73.0	73.0	74.7
	x_t	92.4	93.1	93.8	93.0	93.9
CA2	r_t	67.2	72.6	73.3	73.3	71.7
	x_t	91.7	93.7	93.6	93.6	93.2
CA3	r_t	66.8	71.3	69.0	72.5	70.4
	x_t	91.3	93.1	91.9	93.7	93.9
IPCA	r_t	70.0	79.1	84.2	85.9	87.6
	x_t	93.4	96.7	98.2	99.0	99.3
Panel B: R^2_{pred}						
CA0	r_t	2.3	2.9	1.6	2.9	4.2
	x_t	3.0	3.6	2.0	4.5	6.5
CA1	r_t	3.6	3.1	3.1	3.5	3.3
	x_t	4.9	4.6	4.2	4.8	4.6
CA2	r_t	3.7	3.0	3.5	3.1	3.1
	x_t	5.0	4.4	4.9	4.4	4.2
CA3	r_t	3.6	3.2	3.7	3.3	3.4
	x_t	5.2	4.4	5.0	4.6	4.7
IPCA	r_t	3.0	2.7	1.9	1.9	1.2
	x_t	4.0	3.8	3.8	3.1	1.9

Table D.1: Out-of-sample R^2s

The table reports the out-of-sample total R^2s (in %) for individual option contracts (r_t) and characteristics managed portfolios (x_t) using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA1 through CA3. In all models, the number of factors,K, varies from 1 to 5. Panel A shows the R^2_{total}, whereas Panel B depicts R^2_{pred}. The sample period is from January 1996 to December 2017 as in Büchner and Kelly (2022).

D.2 Out-of-Sample Annualized Sharpe Ratios of Long-Short Portfolios

Model	K				
	1	2	3	4	5
Panel A: Equal-Weight					
CA0	0.19	0.52	0.40	0.52	0.63
CA1	0.33	0.36	0.42	0.43	0.56
CA2	0.38	0.29	0.71	0.42	0.43
CA3	0.36	0.45	0.51	0.42	0.42
IPCA	0.31	0.29	0.24	0.31	0.31
Panel B: Value-Weight					
CA0	0.11	0.39	0.60	0.62	0.63
CA1	0.26	0.39	0.32	0.45	0.76
CA2	0.29	0.22	0.64	0.72	0.69
CA3	0.31	0.61	0.61	0.50	0.44
IPCA	0.35	0.42	0.12	0.28	0.40

Table D.2: Out-of-Sample Annualized Sharpe Ratios of Long-Short Portfolios

The table reports the out-of-sample annualized sharpe ratios of long-short portfolios using the instrumented principal component analysis (IPCA) model and conditional autoencoders CA1 through CA3. In all models, the number of factors, K, varies from 1 to 5. The sample period is from January 1996 to December 2017 as in Büchner and Kelly (2022).

D.3 Performance of Out-Of-Sample Long-Short Portfolios

Model		Low	2	3	4	High	H-L	
Panel A: Equal-Weight								
CA0	Predicted	-1.06	-0.71	-0.49	-0.30	-0.02	1.04	
	Actual	-0.73	-0.61	-0.58	-0.50	-0.37	0.36	
	Std	4.14	3.00	2.04	1.74	1.66	3.13	
	SR	-0.61	-0.71	-0.99	-1.00	-0.78	0.40	
CA1	Predicted	-0.98	-0.62	-0.41	-0.26	-0.11	0.87	
	Actual	-0.78	-0.65	-0.59	-0.41	-0.37	0.41	*
	Std	4.56	2.43	1.93	1.80	1.85	3.45	
	SR	-0.60	-0.93	-1.05	-0.79	-0.69	0.42	
CA2	Predicted	-0.89	-0.59	-0.42	-0.32	-0.22	0.67	
	Actual	-0.82	-0.73	-0.60	-0.44	-0.22	0.60	***
	Std	4.50	2.37	1.80	1.70	2.16	2.90	
	SR	-0.63	-1.06	-1.15	-0.90	-0.36	0.71	*
IPCA	Predicted	-1.22	-0.65	-0.39	-0.21	0.04	1.27	
	Actual	-0.68	-0.60	-0.61	-0.48	-0.43	0.24	
	Std	4.47	2.69	1.90	1.74	2.02	3.59	
	SR	-0.53	-0.77	-1.12	-0.96	-0.75	0.24	
Panel B: Value-Weight								
CA0	Predicted	-1.00	-0.70	-0.49	-0.30	-0.01	0.99	
	Actual	-0.68	-0.52	-0.48	-0.40	-0.27	0.40	*
	Std	3.30	2.49	1.82	1.68	1.46	2.32	
	SR	-0.71	-0.73	-0.92	-0.82	-0.65	0.60	*
CA1	Predicted	-0.91	-0.61	-0.41	-0.26	-0.11	0.80	
	Actual	-0.58	-0.60	-0.48	-0.35	-0.28	0.30	
	Std	4.34	1.90	1.78	1.54	1.64	3.33	
	SR	-0.47	-1.09	-0.94	-0.78	-0.59	0.32	
CA2	Predicted	-0.83	-0.58	-0.42	-0.32	-0.22	0.61	
	Actual	-0.71	-0.64	-0.52	-0.38	-0.21	0.50	**
	Std	3.96	2.03	1.63	1.53	1.66	2.70	
	SR	-0.62	-1.10	-1.11	-0.87	-0.44	0.64	*
IPCA	Predicted	-1.15	-0.63	-0.38	-0.20	0.03	1.17	
	Actual	-0.45	-0.32	-0.44	-0.38	-0.35	0.10	
	Std	3.93	2.90	1.66	1.51	1.64	2.99	
	SR	-0.40	-0.39	-0.91	-0.87	-0.73	0.12	

Table D.3: Performance of Out-Of-Sample Long-Short Portfolios

The table shows the returns to option portfolios sorted by the predictions of the instrumented principal component model (IPCA) and the conditional autoencoder models CA1 and CA2. Portfolios are either equally-weighted (Panel A) or value-weighted (Panel B). Predicted denotes the models predicted returns. Actual denotes the realized returns. SD is the standard deviation of realized returns. SR denotes the realized annual Sharpe ratio. The last column indicates whether the realized returns (Sharpe ratios) of high-minus-low (H-L) portfolio is significantly different from zero. The number of factors, K, is fixed to three. The sample period is from January 1996 to December 2017 as in Büchner and Kelly (2022).

Appendix E Performance Attribution via Shapley Value

E.1 Methodology

In the following, I describe the proposed portfolio performance attribution of Moehle et al. (2021) which uses Shapley values.

Let a general investment process consist of n features. These features can be active or inactive. Examples for features are the choice of a benchmark index, the inclusion of sectors or the inclusion of a specific investment signal. $x_i \in \{0, 1\}$ denote the i-th feature. It is either active, i.e., $x_i = 1$, or inactive, i.e., $x_i = 0$. Moehle et al. (2021) calls a collection of features a configuration which is given as a Boolean vector $x = (x_1, \ldots, x_n)$. There are in total 2^n possible configurations. The configuration $x = (1, \ldots, 1)$ means that all features are enabled. It is called full configuration and assumed to be the configuration which is actually used in the investment process. Contrarily, the configuration $x = (0, \ldots, 0)$ is called the baseline or benchmark configuration.

In order to evaluate the performance of (hypothetical) configurations, a backtesting engine is used which generates a real-valued performance metric, e.g., the return, risk, risk-adjusted return, turnover, etc. The backtesting is represented as a function

$$f : \{0, 1\}^n \to \mathbb{R}, \tag{E.1}$$

with

$$y = f(x) = f(x_1, \ldots, x_n). \tag{E.2}$$

Let denote by y^{real} the performance of the full configuration.

Following Moehle et al. (2021), the lift or marginal contribution of a feature is the change in performance given the inclusion of this feature. Let x be a configuration with $x_i = 0$. Then $\tilde{x} = x + e_i$ is the configuration by enabling feature i. Consequently, the lift or marginal contribution is defined as

$$f(x + e_i) - f(x). \tag{E.3}$$

Finally, an attribution method determines real values a_i for the ith feature and a baseline amount b. The attributes a_i determine how much of the overall performance is caused by feature i. The attribution is represented as vector $a = (a_1, \ldots, a_n)$.

Moehle et al. (2021) define several desirable properties of an attribution method:

- Full attribution: The observed performance is fully attributed to the features plus the baseline, i.e, $y^{\text{real}} = f(1, \ldots, 1) = a_1 + \ldots a_n + b$.
- Correct baseline value: The baseline value corresponds to the performance of the baseline configuration, ie., $b = f(0, \ldots, 0)$.
- Fairness: An attribution method is called fair, if for any permutation of the features, the attribution is permuted in the same way.
- Monotonicity: An attribution method obeys monotonicity if for any change in f such that the marginal contribution of a single feature does not decrease, the attribution of this feature neither decreases.

Moehle et al. (2021) propose the Shapley method for portfolio attribution. It is the average of permuted sequential attributions over all $n!$ permutations. Let a_π and b denote the attribution for permuted sequential attribution with permutation π. Then, the Shapley attribution is

$$a = \frac{1}{n!} \sum_{\pi} a_\pi, \tag{E.4}$$

where the sum is over all $n!$ permutations. As the authors argue, the Shapley method is the only method satisfying all four properties described above, but requires 2^n backtests. The attribution for feature i, i.e., a_i, can be exactly computed as

$$a_i = \left(\sum_{x \in \mathcal{X}_i} \frac{(1^T x)!(n - 1^T x - 1)!}{n!} \left(f(x + e_i) - f(x) \right) \right) - b, \tag{E.5}$$

where $\mathcal{X}_i$ is the sef of configurations with feature i disabled, i.e., $\mathcal{X}_i = \{x | x_i = 0\}$. As evident in Equation (E.5), the ith Shapley attribution is the weighted average of the lift of adding feature i.

E.2 Performance Attribution of Out-Of-Sample Long-Short Value-Weighted Portfolios

Model	Subset	Low	High	H-L
CA0	Short Term ATM	0.02	-0.05	-0.07
	Short Term OTM put	-0.18	-0.08	0.09
	Short Term OTM call	-0.14	-0.10	0.04
	Medium Term ATM	-0.03	0.01	0.04
	Medium Term OTM call	-0.08	-0.07	0.01
	Medium Term OTM put	-0.25	-0.05	0.20
	Long Term ATM	0.09	0.21	0.12
	Long Term OTM call	-0.05	-0.04	0.01
	Long Term OTM put	0.12	-0.01	-0.13
	Total	-0.51	-0.17	0.34
CA1	Short Term ATM	0.00	-0.04	-0.04
	Short Term OTM put	-0.23	-0.07	0.15
	Short Term OTM call	-0.15	-0.09	0.06
	Medium Term ATM	0.06	0.02	-0.04
	Medium Term OTM call	-0.11	-0.07	0.04
	Medium Term OTM put	-0.10	-0.05	0.05
	Long Term ATM	0.08	0.17	0.09
	Long Term OTM call	-0.06	-0.02	0.04
	Long Term OTM put	0.06	-0.02	-0.08
	Total	-0.45	-0.19	0.27
CA2	Short Term ATM	-0.03	-0.06	-0.03
	Short Term OTM put	-0.19	-0.06	0.13
	Short Term OTM call	-0.15	-0.11	0.05
	Medium Term ATM	0.02	0.05	0.03
	Medium Term OTM call	-0.09	-0.10	-0.01
	Medium Term OTM put	-0.12	-0.07	0.06
	Long Term ATM	0.07	0.20	0.14
	Long Term OTM call	-0.06	-0.03	0.03
	Long Term OTM put	0.05	0.03	-0.02
	Total	-0.50	-0.14	0.36
IPCA	Short Term ATM	0.02	-0.05	-0.08
	Short Term OTM put	-0.15	-0.09	0.06
	Short Term OTM call	-0.15	-0.09	0.06
	Medium Term ATM	0.03	0.02	-0.01
	Medium Term OTM call	-0.09	-0.08	0.01
	Medium Term OTM put	-0.10	-0.08	0.03
	Long Term ATM	0.05	0.14	0.09
	Long Term OTM call	-0.09	-0.03	0.06
	Long Term OTM put	0.09	0.04	-0.05
	Total	-0.38	-0.21	0.17

Table E.1: Performance Attribution of Out-Of-Sample Long-Short Portfolios

The table shows the performance attribution of out-of-sample long-short portfolios (value-weighted) using the instrumented principal component analysis (IPCA) and the conditional autoencoder models CA0, CA1 and CA2. In all models, the number of factors is fixed to three. The out-of-sample performance for the short (Low), long (High), and long-short (H-L) are reported. Performance attribution is done with respect to the option subsets defined in Section 3 via Shapley values (Moehle et al., 2021).

Bibliography

Andreou, P. C., Bali, T. G., Kagkadis, A., Lambertides, N., 2021. Firm growth potential and option returns. Georgetown McDonough School of Business Research Paper No. 3874674.

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. Review of Financial Studies 16, 527–566.

Bali, T. G., Beckmeyer, H., Moerke, M., Weigert, F., 2021. Option return predictability with machine learning and big data.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Bianchi, D., Babiak, M., 2021. A factor model for cryptocurrency returns.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. Journal of Financial Economics 118, 113–134.

Büchner, M., Kelly, B., 2022. A factor model for option returns. Journal of Financial Economics 143, 1140–1161.

Campbell, J. Y., Cochrane, J. H., 1999. By force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Cao, J., Han, B., 2013. Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics 108, 231–249.

Cao, J., Han, B., Tong, Q., Zhan, X., 2021. Option return predictability. Review of Financial Studies, forthcoming.

Carr, P., Wu, L., 2004. Time-changed lévy processes and option pricing. Journal of Financial economics 71, 113–141.

Christoffersen, P., Fournier, M., Jacobs, K., 2018. The factor structure in equity options. Review of Financial Studies 31, 595–637.

Cochrane, J. H., 2011. Presidential address: Discount rates. Journal of finance 66, 1047–1108.

Constantinides, G. M., Jackwerth, J. C., Perrakis, S., 2009. Mispricing of s&p 500 index options. Review of Financial Studies 22, 1247–1277.

Coval, J. D., Shumway, T., 2001. Expected option returns. Journal of Finance 56, 983–1009.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343–1376.

Eisdorfer, A., Goyal, A., Zhdanov, A., 2020. Cheap options are expensive.

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

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fournier, M., Jacobs, K., Orłowski, P., 2021. Modeling conditional factor risk premia implied by index option returns.

Frazzini, A., Pedersen, L. H., 2021. Embedded Leverage. Review of Asset Pricing Studies 12, 1–52.

Giglio, S., Xiu, D., 2021. Asset pricing with omitted factors. Journal of Political Economy 129, 1947–1990.

Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. Journal of Financial Economics 94, 310–326.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450.

Gu, S., Kelly, B. T., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Heston, S. L., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.

Horenstein, A. R., Vasquez, A., Xiao, X., 2020. Common factors in equity option returns.

Israelov, R., Kelly, B. T., 2017. Forecasting the distribution of option returns.

Jones, C. S., 2006. A nonlinear factor analysis of s&p 500 index option returns. Journal of Finance 61, 2325–2363.

Karakaya, M. M., 2014. Characteristics and expected returns in individual equity options.

Kelly, B. T., Palhares, D., Pruitt, S., 2020a. Modeling corporate bond returns.

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kelly, B. T., Pruitt, S., Su, Y., 2020b. Instrumented principal component analysis.

Kelly, B. T., Xiu, D., 2021. Factor models, machine learning, and asset pricing. Working Paper.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Kozak, S., Nagel, S., Santosh, S., 2018. Interpreting factor models. Journal of Finance 73, 1183–1223.

Lundberg, S. M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777.

Moehle, N., Boyd, S., Ang, A., 2021. Portfolio Performance Attribution via Shapley Value.

Ofek, E., Richardson, M., Whitelaw, R. F., 2004. Limited arbitrage and short sales restrictions: Evidence from the options markets. Journal of Financial Economics 74, 305–342.

Pohl, W., Schmedders, K., Wilms, O., 2018. Higher order effects in asset pricing models with long-run risks. Journal of Finance 73, 1061–1111.

Pukthuanthong, K., Roll, R., Subrahmanyam, A., 2019. A protocol for factor identification. Review of Financial Studies 32, 1573–1607.

Tian, M., Wu, L., 2021. Limits of arbitrage and option risk premium.

Curriculum Vitae

Mathis Rudolf Werner Mörke, born on June 1st, 1988, in Gütersloh (Germany)

Education

2023: PhD in Finance, University of St.Gallen

2013: Master of Science UZH ETH in Quantitative Finance, ETH Zurich and University of Zurich

2011: Bachelor of Science in Mathematics with Minor in Business (Wirtschaftsmathematik), University of Bielefeld

Professional Experience

2015-2017: Senior Consultant, Fintegral Consulting AG, Zurich

2014-2015: Junior Quant Engineer, swissQuant Group AG, Zurich

2011-2014: Quantitative Investment Analyst, Vescore Solutions AG, Zurich