Option Return Predictability with Machine Learning and Big Data*

Turan G. Bali†, Heiner Beckmeyer‡, Mathis Moerke§, 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. The nonlinear machine learning models generate statistically and 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.

JEL classification: G10, G12, G13, G14

Keywords: Machine learning, big data, option return predictability

*We are grateful to the editor, Stefano Giglio, and two anonymous referees for their constructive and insightful comments. We also benefited from discussions with Manuel Ammann, Nicole Branger, Jie Cao, Jerome Detemple, Ilias Filippou, Amit Goyal, Alexander Kempf, Sebastiano Manzan, Andreas Neuhierl, Seth Pruitt, Alberto Rossi, Paul Söderlind, Sebastian Stöckl, Yinan Su, Allan Timmermann, Martin Wallmeier, Guofu Zhou, and seminar participants at the Virtual Derivatives PhD Workshop organized by the University of Illinois Urbana-Champaign and Michigan State University, the 14th Annual Hedge Fund Conference at Imperial College, the annual meeting of the German Academic Association of Business Research 2022, the SFI Research Days 2022, an internal seminar at Goldman Sachs, an internal seminar at Hull Tactical Asset Allocation, the BVI-CFR 2021 seminar, the Federal Reserve Board, Georgetown University, the University of Fribourg, the University of Liechtenstein, the University of Minnesota, and the University of Muenster.

†Robert S. Parker Chair Professor of Finance, McDonough School of Business, Georgetown University, Washington, D.C. 20057. Email: turan.bali@georgetown.edu.

‡School of Business and Economics, Finance Center Münster, University of Münster, Universitätsstr. 14-16, 48143 Münster, Germany. E-mail: heiner.beckmeyer@wiwi.uni-muenster.de.

§Swiss Institute of Banking and Finance, University of St.Gallen, Unterer Graben 21, 9000 St.Gallen, Switzerland. Email: mathis.moerke@unisg.ch.

¶Institute of Financial Analysis, University of Neuchâtel, Rue A.-L. Breguet 2, 2000 Neuchâtel, Switzerland. Email: florian.weigert@unine.ch. Florian Weigert is also affiliated with the Centre of Financial Research (CFR) Cologne and thankful for their continuous support.

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 characteristic-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 nonlinearities 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 and Kapadia (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 273 variables composed of 80 option-based characteristics (e.g., option illiquidity, time-to-maturity, and 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 (2022)). 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 combine 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 pre-

[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 how well a model predicts cross-sectional option return spreads.

diction 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*) producing out-of-sample R^2s of 2.26% and 1.96%.[3] Moreover, the equal-weighted ensemble of all nonlinear models (denoted N-En) outperforms the ensemble of all linear models (denoted L-En) by more than 1.7% in out-of-sample R^2 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 all other models and most other models with statistical significance at the 5% level (the only exceptions to this finding are *GBR*, *Dart*, and *feed-forward neural networks* which all produce forecasts highly correlated with the nonlinear ensemble model (correlations amount to 0.95, 0.93, and 0.77, respectively). The outperformance of nonlinear models compared to linear models is stable over time with a higher predictability for future option returns in 69.8% of the months in our sample (86.0% when considering the cross-sectional out-of-sample R^2). Notably, we also find better predictions for the nonlinear models during the December 2019 – December 2020 period in which the COVID19 pandemic shook financial markets worldwide.[4] The higher predictability of nonlinear models not only holds for the complete set of options investigated in our sample, but also for different option buckets, such as options sorted by maturity (i.e., short-term and long-term options) and moneyness (i.e., out-of-the money, at-the-money, and in-the-money options).

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 the long-short portfolios based on L-En's and N-En's forecasts of expected returns generate economically significant return spreads of 1.30% and 2.04% per month, respectively, both statistically significant at the 1% level.[5] The long-short return spread

[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. This discrepancy may be driven by the different (and shorter) sample period that we consider, which is restricted by the availability of information on single equity options.

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

[5]The respective monthly Sharpe ratios amount to 1.03 and 1.28.

of the nonlinear ensemble outperforms the return spread of the linear ensemble by statistically significant 0.74% per month, stressing the importance of nonlinearities. This result holds also for the subset of call and put options separately, does not depend on earnings announcements, and persists over time. Moreover, the profitability of the long-short return spread of the nonlinear ensemble exceeds existing and newly proposed measures of expected return benchmarks, and is robust to risk adjustments of established asset pricing models, accounting for time-varying leverage, as well as changes in the length of the training window, return frequency, and different samples of big and liquid stocks on which options can be traded. The results also remain significant across different states of the economy.

Digging deeper into the compositions of the different spread portfolios, we find that the short leg contains more puts and short-term options than the long leg. Interestingly, the short leg of the spread portfolio also displays strong differences to the other portfolios in terms of complexity of the characteristics that determine the allocation of options into portfolios. In this sense, options that are selected in the short leg are determined by the least number of characteristics, but have the highest number of nonlinearities and interaction effects among these characteristics.

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 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 25% – 100% of the quoted bid and ask spread, which we denote as the effective spread (Eisdorfer et al., 2022). In addition, we also incorporate the costs of trading of the delta-hedging procedure by accounting for a similar percentage of the underlying's quoted spread.[6] Our results show that the returns of the nonlinear machine learning trading strategy remain sizeable (0.67% per month) even if investors have to pay the

[6]To simulate a realistic investment process, we add an estimate of the transaction costs at the time of the trade initiation to the return prediction and then sort options into decile portfolios.

full effective spread for transactions and delta-hedging on all options.[7] Margins are an important consideration when investing in the options market. On top of the transaction costs arising from bid-ask spreads, we also include different margin requirements for setting up hedged long and short option positions. Realized returns and Sharpe ratios of the predictions made by the nonlinear ensemble decrease, but only turn insignificant if the investor had to pay 100% of the quoted spread for each option and delta-hedge. Importantly, the predictions by the nonlinear ensemble significantly outperform those by its linear counterpart in all cases.

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. To do so, we classify our 273 option and stock predictor variables into 12 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*. If we assess the functional form of the impact of the three most important single characteristics on the model-predicted delta-hedged return, our results reveal that higher implied volatility negatively affects returns, whereas higher bid-ask spreads and industry momentum predict returns positively.[8]

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

[7] It is important to note that in this case (i.e., 100% effective spread for transaction costs and delta-hedging), we do not observe an outperformance of the trading strategy based on linear machine learning models any more. This again illustrates the importance of incorporating nonlinearities and interactions when forming option portfolios.

[8] Investigating this functional form reveals that each of the ten most important characteristics impacts returns in a nonlinear fashion.

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.[9] 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. However, the benefit of adding stock-based characteristics to option-based characteristics is substantial and helps to obtain more accurate forecasts for future option returns.

Finally, we determine possible sources of option return predictability. 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. To test this conjecture, we create different indices of information frictions based on stock- and option-level information. In line with our prediction, we find that predictability of option returns increases with higher informational frictions. Our results reveal that the out-of-sample R^2 for the nonlinear ensemble model equals 5.32% (0%) for options whose underlyings fall within the highest (lowest) quintile of stock-level information frictions. Options exhibiting the highest (lowest) level of information frictions show an R^2 of 4.00% (1.36%). We also estimate the level of mispricing per option contract, again using a composite mispricing score. Consistent with the intuition that machine learning models manage to identify misvaluation in the options market, overall predictability is increasing in the mispricing score.

The remainder of the paper is organized as follows. Section 2 reviews the literature and outlines our contribution. Section 3 discusses possible benefits of including nonlinearities and interaction effects when predicting option returns. Section 4 describes alternative machine learning methods implemented in this study and explains how we evaluate the

[9] 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.

models' predictive power. In Section 5, we introduce the option return data and describe the option and stock characteristics used for prediction. Section 6 presents the main empirical results. Section 7 investigates the sources of option return predictability. We conclude in Section 8.

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. By 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. Roll, Schwartz, and Subrahmanyam (2010) examine trading volume in option markets relative to the volume in underlying stocks and relate it to contemporaneous returns. Cao and Han (2013) show that delta-hedged option returns decrease monotonically with an increase in the idiosyncratic volatility of the underlying stock. Bali and Murray (2013) find a strong negative relation between risk-neutral skewness and delta- and vega-neutral equity option returns, consistent with a positive skewness preference. An, Ang, Bali, and Cakici (2014) show that the cross-section of stock returns predicts future changes in option implied volatilities. Byun and Kim (2016) find that call options written on the most lottery-like stocks underperform otherwise similar call options on the least lottery-like stocks. Christoffersen, Fournier, and Jacobs (2018a) show that equity options display a strong factor structure, which is highly correlated to volatility, skew and term structure of S&P500 index options. Christoffersen, Goyenko, Jacobs, and Karoui (2018b) 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) examine the impact of short-sale constraints on the pricing of options. Zhan, Han, Cao, and Tong (2022) 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.

Heston, Jones, Khorram, Li, and Mo (2022) investigate the phenomenon of option momentum and reversal. In contrast to their study, which focuses on extracting trading signals from the insight that options that appreciated over a certain past horizon tend to continue to do so in the future, we are agnostic about which characteristics explain future option returns and instead propose a way to extract information simultaneously from 273 option and stock characteristics. Finally, Goyenko and Zhang (2021) apply machine learning techniques to analyze which characteristics drive option and stock returns.

We also extend the literature on the usage of machine learning techniques in empirical asset pricing. So far, the majority of papers applies machine learning models to predict the cross-section of individual stock returns.[10] Rapach, Strauss, and Zhou (2013) use Lasso in predicting market returns across countries, while 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) apply instrumented principal component analysis (IPCA), detailed in Kelly, Pruitt, and Su (2020b), to model the cross-section of returns which allows for latent factors and time-varying loadings.

Gu et al. (2020) perform a comparative analysis of machine learning methods to measure equity risk premia based on a large set of stock characteristics. The authors use a broad set of stock characteristics following Green et al. (2017), whereas Murray, Xiao, and Xia (2021) focus solely on historical price data. Neuhierl, Tang, Varneskov, and Zhou (2021) examine the predictive power of option characteristics for the cross-section of stock returns. Kozak, Nagel, and Santosh (2020) impose an economically motivated prior on stochastic discount factor coefficients that shrinks contributions of low-variance principal components for the cross-section of stock returns and Chen, Pelger, and Zhu (2020) add to

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

these insights, using deep neural networks to estimate an asset pricing model for individual stock returns. Martin and Nagel (2022) show that asset returns may appear predictable in-sample when analyzing the economy ex-post and stress the importance of out-of-sample tests. Feng, Giglio, and Xiu (2020) propose a new model selection method which accounts for model selection mistakes that produce a bias due to omitted variables, and Lettau and Pelger (2020) construct a new estimator that generalizes principle component analysis by including a penalty on the pricing error in expected returns. A nonparametric method to dissect characteristics based on the adaptive group Lasso is proposed by Freyberger, Neuhierl, and Weber (2020). Giglio, Liao, and Xiu (2021) perform "thousands of alpha tests" to develop a new framework to rigorously perform multiple hypothesis testing in linear asset pricing models. 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 analyse return prediction factors in the Chinese stock market. Recent research also expands the application of machine learning models for the prediction of other asset classes: Kelly, Palhares, and Pruitt (2020a) propose a conditional factor model for corporate bonds returns based on the IPCA approach. Hull, Li, and Qiao (2021) build a predictive model of breakeven implied volatilities of S&P 500 index options. Bali, Goyal, Huang, Jiang, and Wen (2021) find that machine learning models substantially improve the out-of-sample performance of stock and bond characteristics when predicting the cross-section of corporate bond returns. Bianchi, Büchner, and Tamoni (2021) apply similar techniques to Treasury securities, whereas Filippou, Rapach, Taylor, and Zhou (2020) employ them in the context of exchange rates. DeMiguel, Gil-Bazo, Nogales, and Santos (2021) show that machine learning helps to select future outperforming mutual funds and Wu, Chen, Yang, and Tindall (2021) establish similar conclusions for hedge funds. Finally, Li and Rossi (2020) apply machine learning to select mutual funds on the basis of their exposure to a large set of stock characteristics.

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. We also 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. Finally, we introduce an expected return benchmark for delta-hedged options based on the instrumented principal component analysis and show that an ensemble of nonlinear machine learning models produces abnormal economical and statistical profits to the benchmark.

3. Importance of Nonlinearities and Interactions for Predicting Option Returns

The existent academic literature on financial derivatives does not provide clear theoretical guidance on what delta-hedged expected returns should look like. In its simplest form – the Black and Scholes (1973) world – delta-hedged excess returns should be zero. In this section, we shortly discuss what to expect when predicting delta-hedged option returns with characteristics and why we believe that nonlinearities and interactions between predictors should matter in this task.

First, both (delta-hedged) call and put options are financial assets whose payoff structure depends on whether the respective underlying is above or below a pre-defined strike price. As a consequence, the returns of these securities are characterized by strong nonlinearities. Table 1 documents that delta-hedged option returns display high absolute skewness and kurtosis as well as significantly fail the Jarque-Bera test to conform to a normally-distributed random variable.

Second, it has been known since Heston (1993) that option prices have non-constant implied volatilities, so that log returns are not normally distributed under the risk-neutral measure. Thus, option prices may be significantly affected by time-varying skewness and kurtosis. The practical consequence is that using the Black-Scholes formula that assumes zero skewness and excess kurtosis distorts option prices. To address departures from normality in the empirical return distribution, earlier studies rely on nonlinear

Gram-Charlier expansions that allow for additional flexibility over a normal density by introducing adjustments for observed skewness and kurtosis of the assumed distribution.[11] Thus, standard linear models might be unable to capture significant departures from normality in option return distribution. Nonlinear models, on the other hand, provide a more accurate characterization of the skewed, fat-tailed distribution of option returns.

Third, we expect that interactions between variables — in addition to nonlinearities — should matter for predicting future delta-hedged option returns. As an example, Black-Scholes option prices (with constant volatility estimates) will be distorted when implied volatility jointly varies with different predictor variables. To understand how interactions can affect the prediction of option returns, we examine the monthly aggregate time-series of options' implied volatilities and the underlyings' idiosyncratic volatility in our sample. Figure IA1.1 plots the rolling correlation (over a time horizon of 60 months) between both variables.

Over our sample period, the average correlation between these two variables is close to zero (-0.03). However, we observe that the correlation fluctuates over time between a minimum of -0.70 and a maximum value of 0.70. As a consequence, it is likely that not only each set of characteristics predicts future delta-hedged option returns, but it is the interaction between them, that particularly matters. Previous research on the topic supports this idea: Ramachandran and Tayal (2021) document a monotonic relation between short-sale constraints and delta-hedged put returns for overpriced stocks. Hence, the documented predictability is an interaction between three characteristics: the level of mispricing of the underlying stock, the level of short-sale constraints of the underlying stock, and if the option is a put or a call option.

Fourth, results from asset pricing theory show that "realistic" preferences of the representative investor lead to a nonlinear pricing kernel for the cross-section of asset returns. More specifically, Pratt and Zeckhauser (1987), Kimball (1993), and Dittmar (2002) show

[11] A rich literature exists regarding option pricing models with stochastic volatility, stochastic interest rates, and the inclusion of jumps (e.g., Bakshi, Cao, and Chen (1997) and Heston and Nandi (2000)). An alternative way of pricing options is the Gram-Charlier approach, which encompasses the distribution's first four moments (see Jondeau and Rockinger (2001) and Schlögl (2013)). Gram-Charlier distributions capture skewness and kurtosis, while retaining a lot of the tractability of the normal distribution.

that a nonlinear pricing kernel is the outcome if an investor's four derivatives of the utility function have altering signs about final wealth, i.e., investors show non-satiation, they are risk-averse, they like skewness, and dislike kurtosis. Using a different theoretical setup, Bekaert, Engstrom, and Xu (2021) document that time-varying risk aversion and aversion to economic uncertainty of investors translate into expected asset returns which can only be poorly captured by assuming linearity of the pricing kernel. Supporting these theoretical considerations, Büchner and Kelly (2022) find that unconditional factor exposures appear to be nonlinear in the moneyness domain of S&P 500 index option returns. Hence, we expect that also forecasts for individual option returns should take into account nonlinearities and interactions between predictor variables to accurately translate investor preferences.

Due to all these reasons, machine learning methods, which allow for nonlinearities and interactions of predictor variables, seem particularly promising to forecast delta-hedged option returns and to derive trading strategies based on the predictability. In our empirical analysis we also propose an adequate benchmark for expected option returns. Precisely, we compare our results to (*i*) a newly constructed benchmark based on instrumented principal component analysis, (*ii*) forecasts based on univariate predictions, and (*iii*) different asset pricing models consisting of stock and option-based risk factors.

4. Methodology and Performance Evaluation

In its most general form, we express future option returns as the sum of expected returns and the 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). Appendix IA3 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 combine 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 combine 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). Appendix IA4 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,

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

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 $(r_{i,s,t+1} - \bar{r}_{i,s,t+1})$.

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:

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

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:

$$c_{t+1}^{(j)} = \frac{1}{n_{\mathcal{T}_3}} \sum_{(i,t)\in\mathcal{T}_3} \left[r_{i,s,t+1}^2 - (\hat{e}_{i,s,t+1}^{(j)})^2 \right]. \tag{7}$$

Here, $n_{\mathcal{T}_3}$ is the number of observations in the testing sample and $\hat{e}_{i,s,t+1}^{(j)}$ is the forecast error on option i on underlying s 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.[12]

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

[12]The main assumption underlying the Diebold and Mariano (1995) test statistic is that the loss differential between two competing forecasts is covariance stationary. However, if one tests equal predictive accuracy of two models across all time-periods and cross-sections in panel data, she may need to make a further assumption about the degree of cross-sectional correlations. Following Gu et al. (2020), we conduct the DM test on the time-series of cross-sectional averages of loss differentials. Thus, in our case, the presence (or degree) of cross-sectional dependence does not matter for the purpose of the test and we only need to account for serial correlation. Following Diebold (2015), we therefore perform the DM test using Newey and West (1987) to account for serial correlation and heteroscedasticity in the time-series of cross-sectional averages. Thus, in our setting, the covariance stationarity of each time series in the panel is sufficient for the DM test to be valid, which is also confirmed by Qu, Timmermann, and Zhu (2022).

error of the mean difference between squared forecast errors $d^{(1,2)}$:

$$d_{t+1}^{(1,2)} = \frac{1}{n_{\mathcal{T}_3}} \sum_{(i,t)\in\mathcal{T}_3} \left[(\hat{e}_{i,s,t+1}^{(1)})^2 - (\hat{e}_{i,s,t+1}^{(2)})^2 \right]. \tag{9}$$

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.

5. 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. Whereas option returns are calculated using historical option prices, the interpolated volatility surface data are only used for constructing option-based characteristics. 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 Zhan et al., 2022), we purposely do not remove stocks with nominal prices below \$5 per share, as Eisdorfer et al. (2022) 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.[13]

Option returns are notoriously noisy, especially for underlyings with few outstanding option contracts and relatively low 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; Zhan et al., 2022).[14] 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).

5.1. 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 and Kapadia (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

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

[14] 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.

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 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.[15] 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 Zhan et al. (2022) as:

$$r_{t,t+\tau} = \frac{\Pi(t, t+\tau)}{|\Delta_t S_t - V_t|}. \tag{12}$$

5.2. Summary Statistics

After the data filters discussed earlier, our sample comprises 4,867,767 options on 6,942 unique underlyings, for a total of 11,983,005 option-month observations for the period January 1996 – December 2020. Our sample is made up of roughly 54% call and 46% put options. Panel A of Table 1 shows that the average monthly delta-hedged option return is

[15] In case we have missing option data at the time we close an option position, we take the intrinsic value of the option as a conservative estimate of the option price (Vasquez, 2017).

0.17%, whereas the median monthly delta-hedged option return is −0.37%. The average moneyness is 1.03 and the average implied volatility is 47.99%. The average (median) days to maturity are 173 (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 0.44% 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 IA7.1 in the Internet Appendix reports summary statistics of the 6,942 stocks in our sample. Our sample includes on average 1,706 optionable stocks per month, which comprise 76% 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 71 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 IA7.1.

5.3. 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: $m^{stand} > 1$ for puts, $m^{stand} < 1$ for calls), out-of-the-money (OTM: $m^{stand} < 1$ for puts, $m^{stand} > 1$ for calls) calls and puts, and at-the-money options (ATM: $-1 \leq m^{stand} \leq 1$), as well as time-to-maturity and moneyness combinations. The moneyness buckets are based on standardized moneyness, i.e., $m^{stand} = \log(K/S)/(\sigma^{atm}\sqrt{\tau})$, where σ^{atm} is the at-the-money implied volatility for time

	Mean	Sd	10-Pctl	Q1	Q2	Q3	90-Pctl	Skew	Kurt	JB
Panel A: All Options (N=12,136,401)										
Delta-Hedged Return	0.11	602.77	-4.45	-1.98	-0.37	1.26	4.46	1.31	10.81	0.0
Days to Maturity	172.88	179.68	21.0	50.0	113.0	204.0	444.0	1.84	3.07	
Moneyness	1.03	0.37	0.77	0.89	1.0	1.11	1.29	2.66	17.06	
Implied volatility	47.89	26.17	23.48	30.23	41.25	58.18	80.48	1.81	5.05	
Absolute Delta	0.46	0.25	0.13	0.25	0.45	0.66	0.82	0.17	-1.01	
Panel B: Call Options (N=6,559,893)										
Delta-Hedged Return	0.35	819.87	-4.91	-2.08	-0.33	1.52	5.22	1.26	11.59	0.0
Days to Maturity	176.75	182.37	21.0	50.0	113.0	204.0	449.0	1.81	2.91	
Moneyness	1.06	0.28	0.82	0.93	1.03	1.15	1.33	2.17	11.13	
Implied volatility	47.37	24.95	23.27	30.06	41.12	57.96	79.62	1.58	3.52	
Absolute Delta	0.51	0.24	0.18	0.32	0.51	0.71	0.84	-0.03	-1.0	
Panel C: Put Options (N=5,576,508)										
Delta-Hedged Return	-0.17	4.37	-3.99	-1.88	-0.41	1.0	3.62	1.46	8.87	0.0
Days to Maturity	168.33	176.35	21.0	50.0	112.0	204.0	417.0	1.86	3.18	
Moneyness	0.99	0.45	0.72	0.85	0.96	1.06	1.22	3.69	30.27	
Implied volatility	48.51	27.51	23.73	30.43	41.39	58.43	81.57	2.05	6.72	
Absolute Delta	0.4	0.25	0.1	0.2	0.36	0.58	0.78	0.47	-0.75	
Panel D: All Options 1996-2002 (N=2,027,277)										
Delta-Hedged Return	1.34	1474.68	-5.96	-2.41	-0.19	2.31	7.02	1.1	11.63	0.0
Days to Maturity	152.63	174.13	22.0	50.0	108.0	176.0	351.0	2.36	5.69	
Moneyness	1.1	0.54	0.78	0.9	1.02	1.17	1.42	3.99	28.44	
Implied volatility	65.39	29.68	32.22	42.95	60.67	82.31	103.98	0.99	1.49	
Absolute Delta	0.49	0.24	0.18	0.31	0.48	0.68	0.83	0.11	-0.94	
Panel E: All Options 2003-2020 (N=10,109,124)										
Delta-Hedged Return	-0.14	9.32	-4.19	-1.92	-0.4	1.1	3.97	1.3	9.54	0.0
Days to Maturity	176.94	180.5	21.0	50.0	113.0	206.0	449.0	1.74	2.62	
Moneyness	1.02	0.33	0.77	0.89	0.99	1.1	1.27	1.97	11.16	
Implied volatility	44.38	23.91	22.7	28.9	38.77	52.96	71.62	2.15	7.77	
Absolute Delta	0.45	0.25	0.13	0.24	0.44	0.65	0.82	0.19	-1.02	

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). Skew denotes skewness. Kurt denotes excess kurtosis. JB denotes the p-value in percent of the Jarque-Bera test if delta-hedged option returns conform to a normal distribution. 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 τ. 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 IA7.3 shows that on average, 15 long term options and 11 short term options belong to each underlying stock per month, with the majority 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 80 we compute, 43 characteristics operate on the level of the underlying (e.g., iv-rv-spread; Goyal and Saretto, 2009), 20 on the level of option buckets (e.g., the Amihud, 2002, illiquidity measure) and 17 on the level of individual option contracts (e.g., the option's time-to-maturity). Appendix IA5 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., 2022), and realized volatility (Cao, Vasquez, Xiao, and Zhan, 2019).[16]

We are left with a total of 273 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 (2022) and Green et al. (2017). Classification group *Contract* contains five option-based characteristics, the time-to-maturity, moneyness, implied volatil-

[16]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).

ity, 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. Appendix IA6 provides a detailed list of all 273 characteristics, their origin from the literature, the primary information source (option- vs. stock-based) and the feature group we have assigned them to.

6. Predicting Option Returns

6.1. 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 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 PLS to -0.18% for elastic net 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.[17] 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, at or above 2% 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 higher R^2_{OS}. The best call and put return predictions are both made by GBR, with $R^2_{OS} \geq 2.0\%$, the worst call predictions are generated by Lasso, and the worst

[17] We find that our results remain unchanged 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.

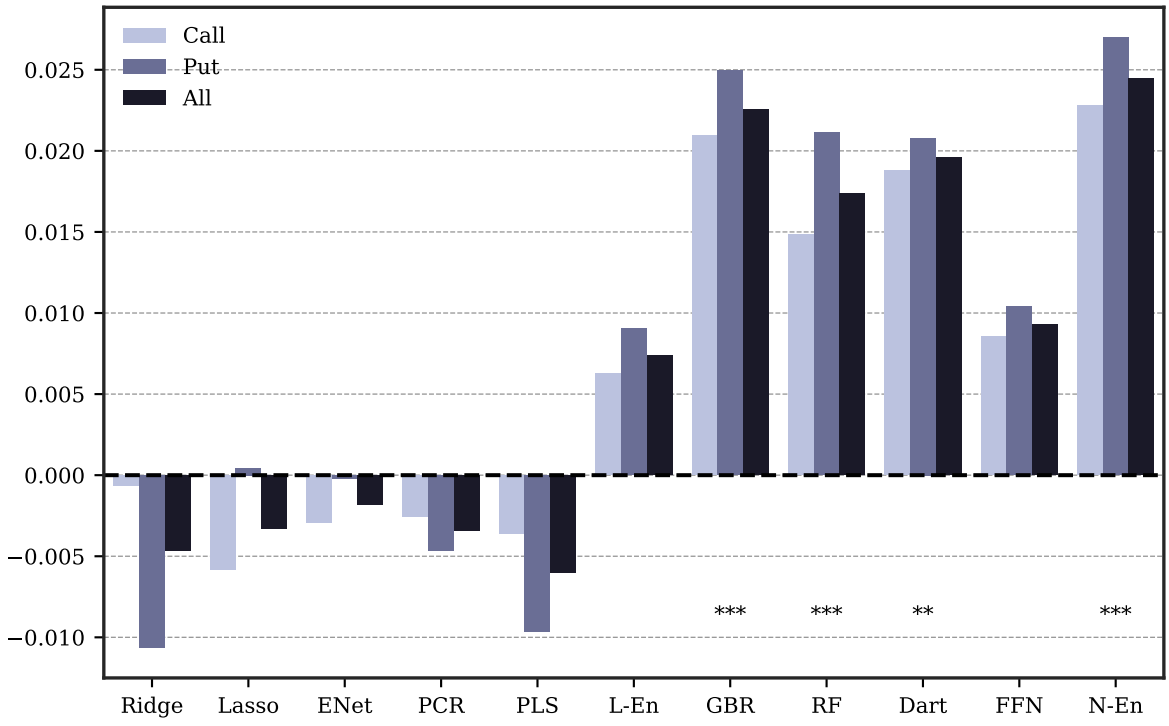

Fig. 1. R^2_{OS} Model Comparison

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.

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 the complex structure does not generalize well to the testing sample of option returns. In contrast, tree-based methods tend to outperform, such that histogram-based estimation including *nonlinear interactions* trumps model complexity in this market.[18]

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

[18]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.

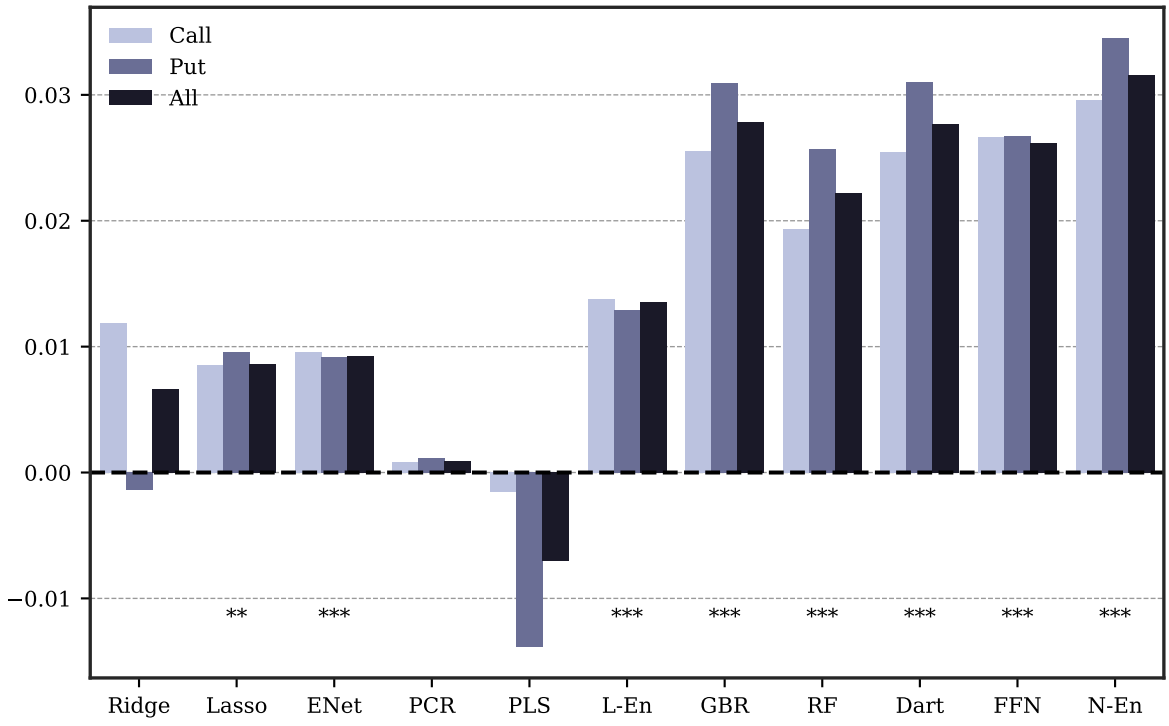

Fig. 2. $R^2_{OS;XS}$ Model Comparison

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.

Hinton, 2012; Lakshminarayanan, Pritzel, and Blundell, 2016).[19] 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 0.9% 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, no model outperforms the ensemble, highlighting the importance of averaging predictions of multiple methods. Just as all tree-based methods, the resulting predictions comfortably outperform 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

[19] Steel (2020) discusses its uses in economics.

predictability on the full sample, especially penalized regressions are able to generate realistic return spreads. For Lasso and ENet these are 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 with an $R^2_{OS;XS}$ of 3.2%.

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 insights 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 around the median, or more exposed to the downside. Notable exceptions from this are GBR and Dart, which manage to produce the best predictability most consistently, with an interquartile range of 1.5% to 4.0%. Leveraging the informational content uncovered by these tree-based models potentially 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 -7.6% to $+4.4\%$ and for N-En between -0.7% and $+6.9\%$. 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

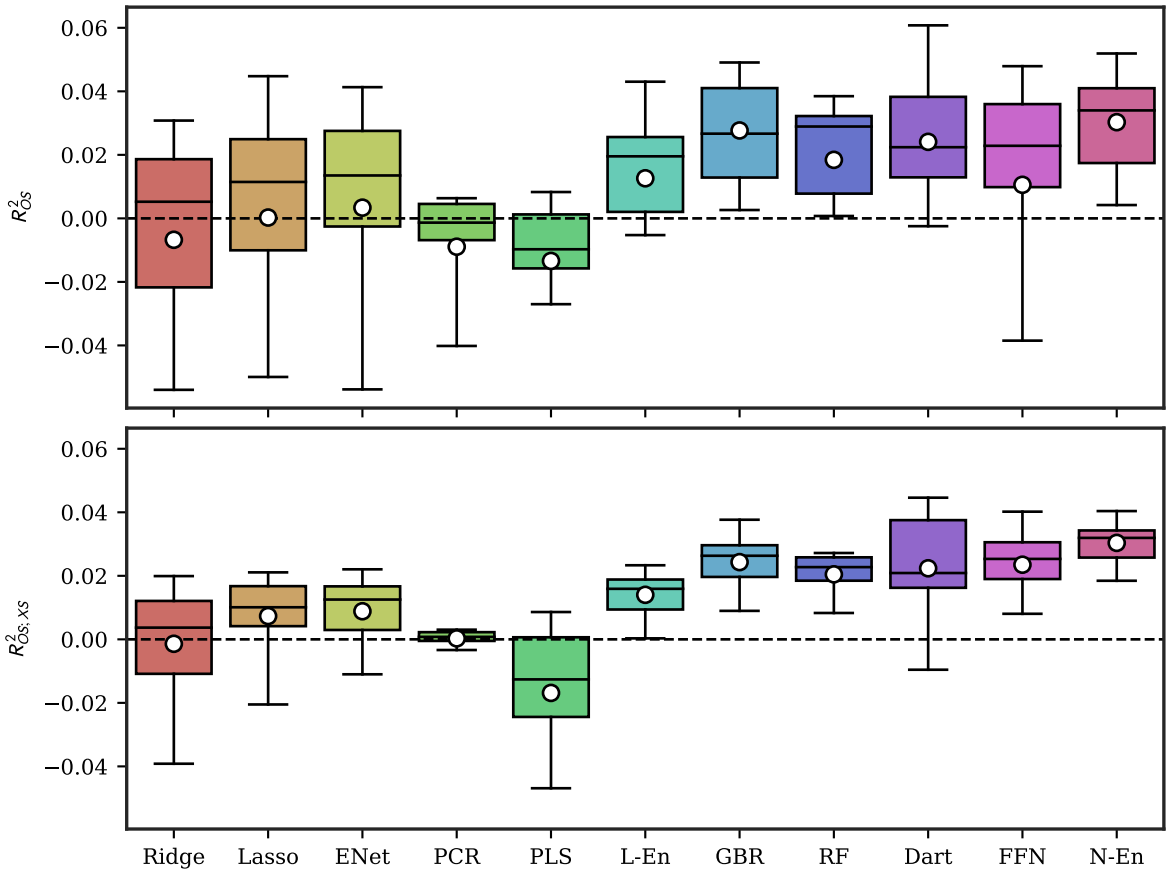

Fig. 3. Time-Series Dispersion of R^2_{OS} and $R^2_{OS;XS}$

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).

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	1.21	1.91	−0.02	−0.36	**3.29**	**2.87**	**2.21**	**2.01**	**3.58**	**3.53**
Lasso		1.67	−0.36	−0.67	**2.43**	**2.27**	1.68	1.59	**3.06**	**2.78**
ENet			−0.56	−0.89	**2.20**	**2.22**	1.56	1.52	**2.06**	**2.74**
PCR				−1.34	**2.26**	**6.68**	**7.27**	**3.11**	1.24	**7.25**
PLS					**2.79**	**7.15**	**9.48**	**3.47**	1.58	**8.14**
L-En						**2.03**	1.04	1.02	−0.23	**2.79**
GBR							**-2.51**	−0.53	−1.30	1.12
RF								0.57	−0.74	**3.36**
Dart									−0.83	0.86
FFN										1.65

Panel B: Forecast Correlation

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	**0.80**	**0.84**	−0.03	0.53	**0.91**	0.48	0.45	0.47	**0.77**	0.61
Lasso		**0.95**	−0.02	0.51	**0.91**	0.41	0.40	0.39	**0.78**	0.56
ENet			−0.02	0.53	**0.92**	0.43	0.42	0.41	**0.78**	0.57
PCR				0.25	0.13	0.18	0.27	0.15	0.03	0.17
PLS					**0.76**	0.51	0.54	0.52	0.58	0.61
L-En						0.54	0.54	0.53	**0.83**	0.68
GBR							**0.86**	**0.91**	0.59	**0.95**
RF								**0.79**	0.57	**0.89**
Dart									0.58	**0.93**
FFN										**0.77**

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 comfortably with a t-statistic of 2.03. Interestingly, the forecasts of GBR and Dart, which have produced the highest single-model predictability, are indistinguishable from one another. FFN manages to outperform Ridge, Lasso and ENet, but none of the nonlinear models.[20] 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

[20]The Diebold and Mariano (1995) test provides a statistical measure of the differences in the total forecast errors. Comparing cross-sectional forecast errors in Appendix IA8.1 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.

t-statistic of 2.79. 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 = 1.65).

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 with either of these methods ($\rho > 90\%$), as does FFN ($\rho > 75\%$), 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 $\rho > 75\%$ 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 $\rho^{\text{N-En,FFN}} = 77\%$ and $\rho^{\text{N-En,RF}} = 89\%$.

Impact of Nonlinearities. The results underscore 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 predicts returns much more consistently. Third, for N-En, the predictive power has stayed roughly constant over time, confirming not only that nonlinear models enlarge the information embedded in option characteristics, but also that the resulting predictability patterns are highly persistent over time.[21] At the

[21] We keep increasing the training period by one year each time we roll forward, such that no historical information is ever discarded.

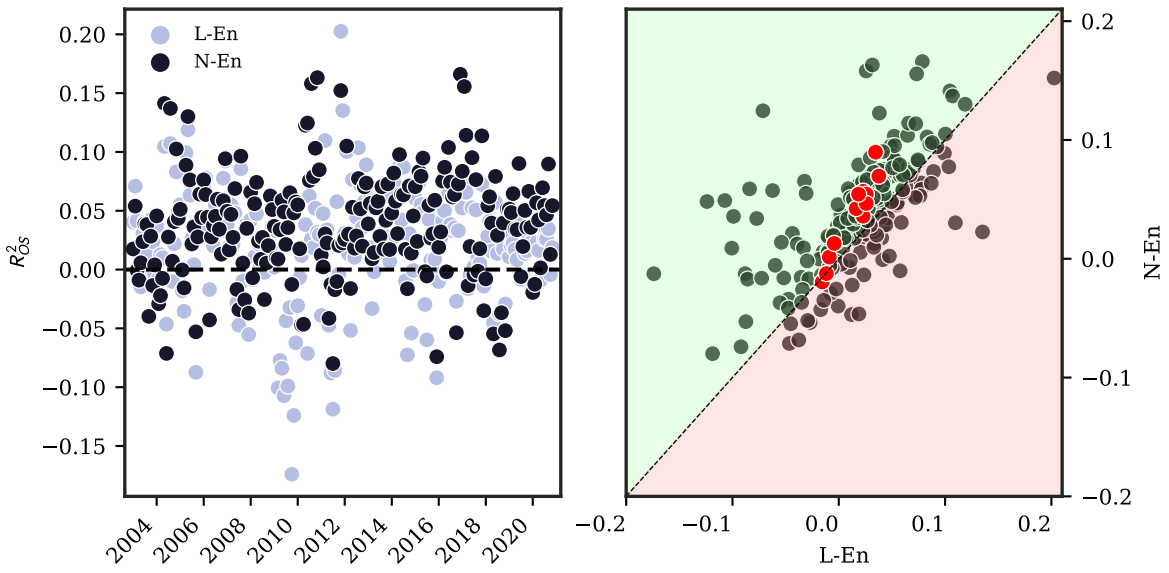

Fig. 4. Comparing Linear and Nonlinear Ensembles

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 red circles represent the Coronavirus selloff and subsequent recovery from December 2019 through December 2020.

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 69.8% 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 the end of our sample in December of 2020, which are the months surrounding the Coronavirus selloff in February and March 2020, and the subsequent recovery. Nonlinear ensembles were better able to deal with this huge exogenous shock, with a performance on-par or exceeding that of L-EN.[22] The R^2_{OS} for N-En dips slightly below zero only in January and February and reaches pre-crisis levels of 4.8% afterwards. While

[22]Given the short Corona-sample period of 13 months, we cannot assess if one model statistically outperforms the other. However, we generally find higher predictability for N-En in an absolute sense.

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 IA8.1 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 86.0% of the months with a positive $R^2_{OS;XS}$ in more than 90% of months.

6.2. 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. 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 2.04% with a Sharpe ratio of 1.28.[23] Figure IA10.1 shows that N-En generates sizable returns each year. Realized Sharpe ratios consistently exceed 0.5 per month and often

[23] Our findings are robust to equal-weighting as shown in Table IA10.1.

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.351	−1.087	1.398	−0.778	−1.730	−1.649	1.942	−0.849	***
2	−0.775	−0.528	1.535	−0.344	−0.781	−0.700	1.529	−0.458	
3	−0.542	−0.365	1.434	−0.255	−0.460	−0.415	1.368	−0.303	
4	−0.369	−0.259	1.466	−0.177	−0.280	−0.266	1.255	−0.212	
5	−0.224	−0.196	1.497	−0.131	−0.155	−0.166	1.295	−0.128	
6	−0.092	−0.122	1.509	−0.081	−0.050	−0.119	1.325	−0.090	
7	0.038	−0.061	1.494	−0.041	0.052	−0.075	1.425	−0.053	
8	0.174	−0.027	1.510	−0.018	0.166	−0.031	1.491	−0.021	
9	0.337	0.046	1.436	0.032	0.324	0.090	1.555	0.058	
Hi	0.637	0.216	1.485	0.146	0.791	0.391	1.835	0.213	
H-L	1.988	1.303	1.270	1.026	2.521	2.040	1.598	1.277	***
		(13.27)		(8.95)		(13.27)		(8.83)	
call	1.864	1.400	1.614	0.867	2.596	2.290	1.941	1.180	***
put	1.943	1.232	1.274	0.967	2.264	1.971	1.663	1.185	***

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.

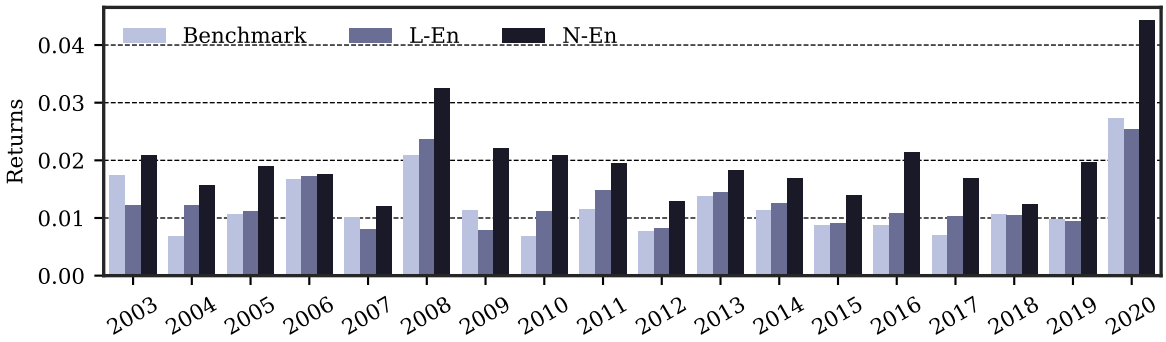

Fig. 5. Trading on Machine Learning Predictions vs. Return Benchmark

The figure shows the year-by-year performance of the trading strategy following predictions by the linear (L-En) and nonlinear ensemble N-En, as well as a benchmark relying on IPCA by Kelly et al. (2019). We show average realized monthly returns of decile-based long-short portfolios.

exceed 1.5. Figure IA10.2 replicates this analysis for put and call options separately. We find that N-En manages to produce significant return spreads for each year in the sample and for both put and call options.[24]

Option Return Benchmark We have highlighted in Section 3 that there is little

[24]Figure IA9.1 and Figure IA9.2 in the Internet Appendix investigate how consistent expected returns of N-En are for put and call options across different maturities.

theoretical guidance on what delta-hedged option returns should look like. As a simplistic benchmark, excess returns on delta-hedged options are zero in the world of Black and Scholes (1973). To provide a more realistic benchmark, we proceed in two ways. First, akin to the capital asset pricing model for risky securities, we propose a linear factor model for delta-hedged option returns with two factors. Following the decomposition in Appendix IA2, which expresses expected delta-hedged option returns as a function of expected returns of the underlying and expected raw option returns, we include the excess return on the value-weighted equity market (CRSP) index (r^M), as well as the excess return on an option market index, which we calculate as the dollar open interest-weighted delta-hedged return of all options available at time t (r^O). Since options are short-lived, we cannot reliably estimate beta-exposures to the two risk factors on a per-option basis. Thus, in the spirit of Büchner and Kelly (2022), we make use of instrumented principal component analysis (IPCA) proposed by Kelly et al. (2019), to instrument the time-variation in betas by observable option- and stock-level characteristics.[25] Since our objective is to provide a simple and interpretable benchmark for expected returns in the options market, we use the stock and option market factors detailed above, instead of estimating latent factors from cross-sectional option returns, i.e., $f_{t+1} = [r^M_{t+1}, r^O_{t+1}]$. Then, we express next-period's option returns as:

$$r_{i,s,t+1} = \alpha_{i,s,t} + \beta_{i,s,t} f_{t+1} + \varepsilon_{i,s,t+1} \tag{13}$$
$$\alpha_{i,s,t} = z'_{i,s,t}\Gamma_\alpha + \nu_{\alpha,i,s,t} \qquad \beta_{i,s,t} = z'_{i,s,t}\Gamma_\beta + \nu_{\beta,i,s,t}.$$

After we estimate $\Gamma_\alpha, \Gamma_\beta$ and f on an expanding sample at the start of each January, we obtain expected option returns in an out-of-sample fashion analogous to the procedure used for N-En and L-En, by fixing the factor means observed until time t, denoted by $\widehat{f_t}$:

$$\mathbb{E}_t[r_{i,s,t+1}] = \alpha_{i,s,t} + \beta_{i,s,t}\widehat{f_t} \tag{14}$$

[25]Since IPCA does not include a sparsity or regularization mechanism, we limit z to include the 50 option and stock characteristics that produce the largest decile return spreads between 2003 and 2020.

With expected delta-hedged option returns at hand, we again invest in the options in the highest expected return decile and short options in the lowest expected return decile. The comparison of year-by-year realized monthly returns for L-En, N-En, and the aforementioned benchmark shown in Figure 5 produces a number of interesting results. First, realized returns of both machine learning ensembles, as well as the benchmark are consistently positive in each year. Second, each method produces the largest return spreads in years of crises, namely 2008 and 2020. Third, the nonlinear ensemble outperforms both the competing linear ensemble, as well as the proposed factor-based benchmark by a large margin. N-En's average monthly return amounts to 2.04%, which compares well to L-En's 1.30% and the benchmark's 1.20%. In fact, N-En manages to outperform L-En and the benchmark at the 1% significance level, whereas the average realized returns of L-En and the benchmark are statistically indistinguishable from one another.

The second benchmark we propose mimics a stylized quantitative investment strategy, by performing univariate sorts on stock- and option-level characteristics. With this, we investigate which characteristics provide independent information about delta-hedged option returns (Green et al., 2017).[26] To do so, we sort options into decile portfolios by each of the 80 characteristics described in Appendix IA5 as well as the 94 stock-level characteristics we obtain from Green et al. (2017).[27] We then form long-short portfolios, as the difference between the highest and lowest decile, wherein option returns in each decile portfolio are weighted by the contract's dollar open interest at trade initiation. We order the sorting in a way that assures that the average return over the testing sample between 2003 and 2020 of the long-short portfolios is positive.

The upper panel of Figure 6 shows a histogram of the average monthly returns that can be achieved using this approach. Three characteristic-sorted long-short portfolios manage to produce monthly returns north of, but close to, 1%. These include the option's implied volatility, the maturity-specific at-the-money volatility, as well as *pzeros* – an illiquidity measure based on the proportion of zero return days of individual option buckets

[26] Green et al. (2017) investigate the question of which stock characteristics provide information about stock returns in a regression setup, as opposed to the univariate sorts we use.

[27] We exclude the industry dummies for this exercise.

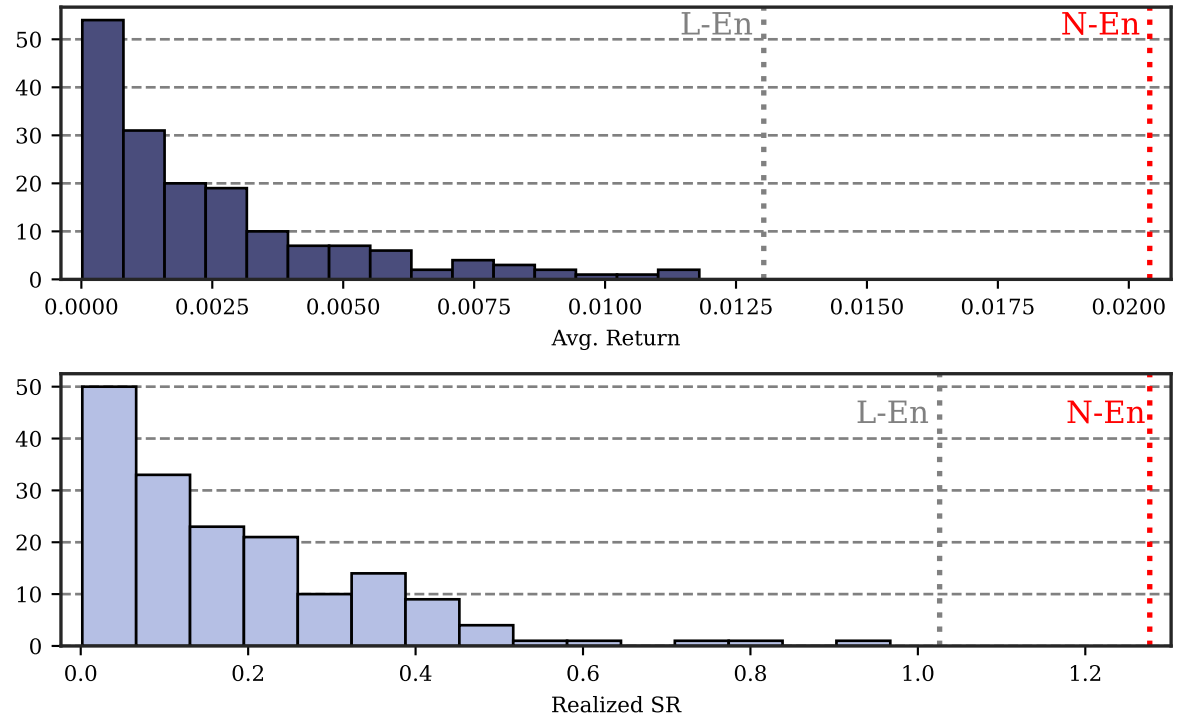

Fig. 6. Average Realized Returns and Sharpe Ratios of Characteristic-Sorted Long-Short Portfolios

The figure shows a histogram of the average realized returns (upper panel) and Sharpe ratios (lower panel) of characteristic-sorted long-short portfolios, using both option- and stock-level characteristics. Each month, we sort options into decile portfolios based on each of the 80 characteristics described in Appendix IA5 and the 94 stock-level characteristics obtained from Green et al. (2017) and calculate long-short portfolio returns, as the difference between the highest and lowest decile, wherein the of each decile portfolio returns are weighted by each contract's dollar open interest at trade initiation. We order the sorting in a way that the average long-short returns are on positive over the testing sample between 2003 and 2020.

(Lesmond, Ogden, and Trzcinka, 1999). The average return of the long-short portfolios amounts to 0.25% per month, with 26 portfolios averaging returns above 0.50%. In comparison, the red (gray) line denotes the average realized return of N-En (L-En) of 2.04% (1.30%), demonstrating a significant outperformance of both methods but especially of the nonlinear machine learning method when compared to the crude portfolio sorting approach. When comparing the realized returns of each univariate sort and the two ensembles, we find that L-En significantly (at the 5% level) dominates all but the sorts on the implied volatility, the at-the-money implied volatility, as well as *pzeros*, whereas N-En outperforms all of the univariate sorts. In the lower panel of Figure 6, we show the realized Sharpe ratios, confirming these findings. Both benchmark analyses highlight the efficacy of the proposed machine learning methods, and stress the importance of allowing

for nonlinearities and interaction effects when predicting option returns.

Portfolio Decomposition. Table 4 provides summary statistics for the decile portfolios based on the nonlinear ensemble N-En. The high and low portfolios tend to include options from $N^U = 502$ and $N^U = 637$ 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 = 0.44), while the high portfolio includes more long-term call options (% Calls = 0.78). 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 portfolio 9 and the high portfolio. For the latter, it stands at 52%. Figure IA10.4 to Figure IA10.6 in the Online Appendix provide a visual representation of the portfolio decomposition, split by the put and call contracts included.

Figure IA10.7 shows the relative share of call options in the machine learning portfolios for each year in our testing sample. 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 exploits expected return differentials of options across or within underlyings. Figure IA10.8 shows the share of options of a given underlying that end up in the same portfolio. We find that this measure is highest for the high and low portfolios being slightly below 50% suggesting that N-En exploits return differentials of options across underlyings. To further shed light on this, we consider alternations of the trading strategy. Enforcing that all options of an underlying are ranked in the same decile portfolio yields a monthly average excess return of 1.93% for N-En (Table IA10.2).

	Lo	2	3	4	5	6	7	8	9	Hi
N^U	637.04	801.99	829.53	813.09	783.93	760.82	742.42	708.65	642.98	502.24
m	1.04	1.01	1.00	1.00	1.01	1.01	1.01	1.02	1.03	1.05
ttm	125.12	154.03	173.00	182.08	187.37	189.01	187.04	183.64	182.52	199.59
$\hat{r} > 0$	0.31	0.35	0.38	0.39	0.41	0.42	0.43	0.44	0.47	0.52
$s(r) = s(\hat{r})$	0.69	0.65	0.62	0.60	0.57	0.53	0.49	0.46	0.48	0.52
% Calls	0.44	0.40	0.42	0.42	0.45	0.49	0.55	0.62	0.69	0.78
Spread	0.20	0.16	0.14	0.12	0.11	0.11	0.11	0.12	0.12	0.14
OI	0.06	0.08	0.10	0.11	0.12	0.12	0.12	0.11	0.10	0.08
delta	0.03	0.00	0.01	−0.00	0.00	0.03	0.08	0.15	0.21	0.29
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	0.21	0.20	0.20	0.21
theta	−0.32	−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 $x \in [vega, theta]$).

Specifically, for each underlying, we compute the average expected return of all options in a given month and use this average to sort underlyings into ten portfolios. Building decile portfolios on the underlying level and then averaging across underlyings yields a monthly excess return of 1.02% for N-En (Table IA10.3), suggesting that N-En primarily exploits return differences *across* underlyings as opposed to *within* the same underlying. However, as the unrestricted full sample return spread amounts to 2.04% per month, N-En tends to combine within and across underlying return predictability.[28]

Prediction Persistence and 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.[29] With this, we can understand

[28]Selecting only one option per underlying, i.e., the option with the highest absolute return prediction, is the most extreme case for exploiting return differentials within and across underlyings. In this case, N-En yields monthly average excess returns of 5.10% (Table IA10.4).

[29]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.

more about the persistence of the machine learning predictions. Figure IA10.9 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. The options of a given stock-bucket combination of the low and high portfolios exhibit the strongest persistence where the low portfolio is again more persistent than the high portfolio.

Importance of Complexity. Table 3 shows that the difference in the high-minus-low returns between L-En and N-En is highly significant at 0.74% per month. This spread is primarily driven by the low portfolios, with significantly lower realized returns for the options shorted by N-En. At the same time, the average prediction error for this portfolio is much smaller for N-En, whereas the errors are comparable for the high decile portfolio. These findings suggest that nonlinearities seem to play a particularly important role in uncovering options with negative returns over the next month.

We investigate the importance of nonlinearities and complexity in general for the predictions of N-En per decile portfolio in Table 5. Complexity comprises three facets: the number of influential characteristics, the importance of nonlinearities, and the degree of interaction effects between characteristics. For the low portfolio, fewer characteristics are important on average. However, the predictions of expected option returns rely much more on nonlinearities for the low portfolio. We assess this by fitting a linear model to the functional form of the impact of each characteristic on expected returns and calculating the standard deviation of the residuals from the linear fit.

Besides nonlinearities, Table 5 reveals that interaction effects play an important role in forming predictions for the low portfolio, represented by a higher degree of local dispersion. At the same time, the portfolio with the second-highest dependence on interaction effects is the high portfolio, highlighting the importance of including nonlinearities and interactions when predicting option returns of both legs of the spread portfolio.

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

	Nbr Imp. Features	Nonlinearities	Interactions (Std)	Interactions (IQR)
Lo	17.859	2.256	1.832	2.613
2	22.170	0.979	0.653	0.850
3	22.374	0.825	0.491	0.624
4	22.030	0.798	0.453	0.582
5	21.625	0.780	0.443	0.578
6	21.450	0.771	0.443	0.573
7	21.610	0.764	0.448	0.572
8	21.735	0.773	0.459	0.578
9	22.049	0.810	0.497	0.625
Hi	22.791	0.971	0.756	0.978

Table 5: The Importance of Complexity for the N-En Decile Portfolios

The table shows the importance of complexity for the predictions made by the nonlinear ensemble (N-En) method. The importance of complexity is measured for each decile portfolio. Complexity constitutes three aspects: the number of important features (Nbr Imp. Features), the degree of nonlinearities (Nonlinearities), and the degree of interaction effects between characteristics. The degree of nonlinearities is assessed by the standard deviation of residuals obtained from fitting a linear function to the functional form of each characteristic. The functional form is estimated via SHAP (SHapley Additive exPlanations; Lundberg and Lee, 2017) values. The degree of interactions is assessed by the local dispersion of SHAP values for each characteristic. Dispersion is either measured by the standard deviation (Interactions (Std)) or the interquartile range (Interactions (IQR)). Higher values denote a higher degree of nonlinearities and interaction effects, respectively.

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. We measure transaction costs by realized effective spreads, which we vary between 25% and 100% of the quoted spread of the option contract or the underlying. Additionally, we also consider the impact of having to post different margins for long and short option positions. Details on the calculation of margin requirements are provided in Appendix IA16.1. When incorporating transaction costs into the sorting step, which places options into the respective decile portfolios, we adjust predicted returns by expected transaction costs before sorting into decile portfolios. In case an option does not mature at the end of the investment period, we assume that the observed spread at trade initiation also applies to unwinding the position after a month. For the expected transaction costs of the hedging portfolio, we assume that we pay on average three times the bid-ask spread of the underlying, quoted at trade initiation and multiplied by the absolute value of the option's delta.[30]

Results are reported in Table 6. The upper panel shows results if we only consider

[30]Varying the assumptions underlying time-t expected transaction costs does not alter our results.

Eff. Spread	L-En			N-En			N vs. L
	H-L	t	SR	H-L	t	SR	
			Option Costs				
25%	0.591	(7.82)	0.514	1.271	(8.06)	0.801	***
50%	0.349	(4.35)	0.286	0.992	(5.47)	0.538	***
75%	0.183	(2.28)	0.148	0.806	(3.89)	0.408	***
100%	0.047	(0.56)	0.037	0.591	(2.79)	0.288	***
			Option And Delta-Hedging Costs				
25%	0.541	(7.73)	0.460	1.235	(8.27)	0.782	***
50%	0.303	(3.67)	0.233	0.916	(5.38)	0.503	***
75%	0.130	(1.48)	0.098	0.684	(3.93)	0.352	***
100%	0.009	(0.11)	0.007	0.500	(2.87)	0.245	***
			Option And Delta-Hedging Costs with Long/Short Margin Requirements				
25%	0.747	(5.81)	0.304	1.711	(6.22)	0.593	***
50%	0.312	(2.19)	0.121	1.218	(4.64)	0.400	***
75%	0.034	(0.22)	0.013	0.712	(2.75)	0.227	**
100%	−0.244	(−1.30)	−0.086	0.352	(1.29)	0.114	**

Table 6: Trading on Machine Learning Predictions – Transaction Costs

The table shows the returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods after accounting for transaction costs through effective spreads. Effective spreads are defined as a fraction of the quoted spread. We consider effective spreads for option prices (upper panel) and option prices and the underlyings' prices (middle panel). Effective spreads for the underlyings' prices account for delta-hedging costs. We additionally consider the impact of having to post different margins for long and short options positions (lower panel). Details on margin requirements are given in Appendix IA16.1. Predictions made by L-En and N-En are adjusted by expected transaction costs before sorting in the high-minus-low portfolios. Each contract is weighted by its dollar open interest at the time of investment. H-L denotes the average monthly realized returns of the high-minus-low portfolio. t denotes the corresponding t-statistic and SR the resulting monthly 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.

transaction costs of the option position. In the middle panel we consider both the transaction costs for trading the option and setting up and maintaining the hedge portfolio. Finally, in the lower panel, we additionally account for margin requirements of long and short hedged positions in the options market. Table 6 shows that predictions by N-En yield statistically significant positive monthly returns for all levels of transaction costs, except for when we assume that the investor has to pay 100% of the quoted spreads on all positions and incorporate long and short margin requirements. In contrast, high-minus-low return spreads following L-En's predictions turn insignificant for moderate to high levels of transaction costs. Trading on N-En's predictions is always superior to L-En's

predictions gauging by the average monthly returns, the Sharpe ratio, and the statistical assessment of this outperformance. Finally, Table 6 shows that delta-hedging costs are generally of second order importance compared to transaction costs of trading the option, whereas margin requirements in combination with high levels of transaction costs are an important impediment to consider.

Muravyev and Pearson (2020) argue that effective spreads are much lower than the ones we measure 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 exploitable.

6.3. 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 group sequentially. This approach, however, is infeasible, given the large number of characteristics (a total of 273) in our estimation and hence, the large computational burden required.[31] Instead, we SHAPs 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 7, 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 surface and knowing where that implied volatility surface lies relative to the market is essential when making option return forecasts. Measures of 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 in-

[31] We use this approach in Section 6.4 and re-estimate each model for three subsamples of the input characteristics.

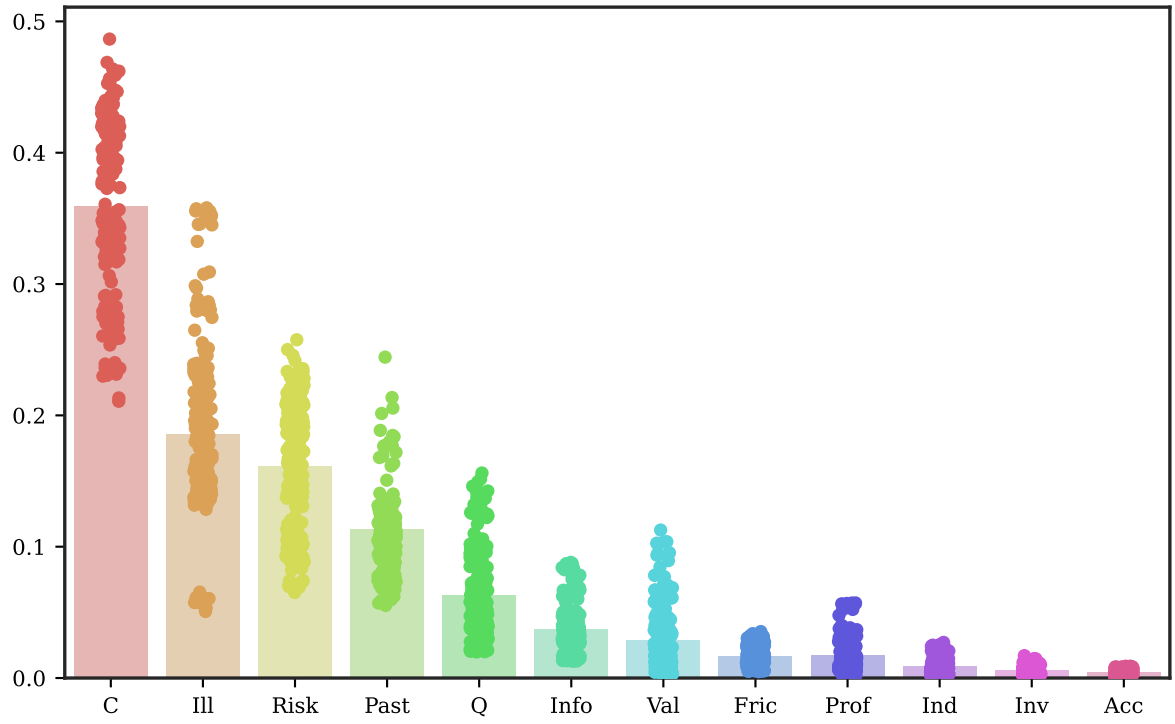

Fig. 7. Feature Group Importance for Nonlinear Forecast Ensemble

The figure shows the feature group importance for the twelve feature groups defined in Appendix IA6 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.

cludes measures of stock and option momentum.[32] 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 (e.g., accruals or investment).

Single Characteristics. Figure 8 shows the ten most important characteristics, along with the dispersion of their relative importance across the testing sample. The

[32]Figure IA11.1 and Figure IA11.2 provide evidence that the above findings are robust when we focus on single option buckets and on single years in our sample, respectively.

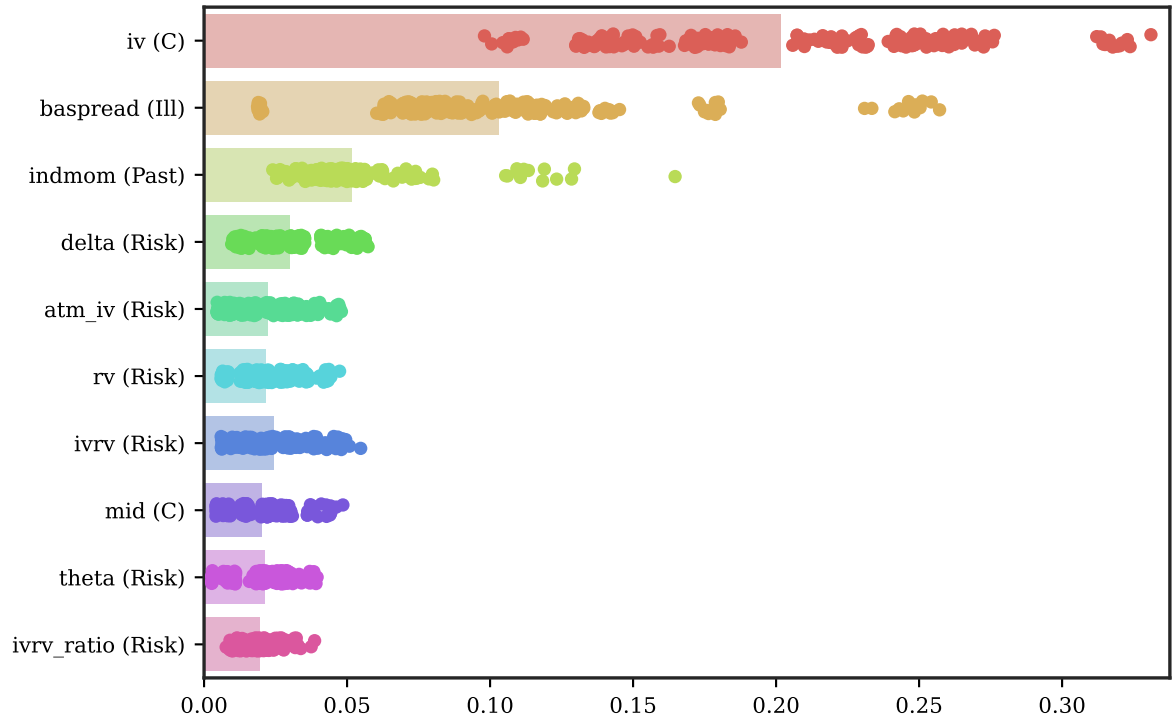

Fig. 8. Most Important Characteristics for N-En

The figure shows the ten most influential characteristics 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 273 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. The feature group associated with the feature as defined in Appendix IA6 is given in parentheses. The abbreviations used: C=Contract, Ill=Illiquidity, Past=Past Prices.

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 delta (delta), the maturity-specific at-the-money implied volatility (atm_iv), 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).[33]

Characteristic Sensitivity. Next, we seek to understand how the most important characteristics affect the direction of the predictions made by N-En. For this, Figure 9 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

[33]Figure IA11.3 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.

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 IA11.4 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 IA11.5 provides further insights on the functional from of the impact of single features on L-En's and N-En's predicted delta-hedged returns. The impact for N-En is highly nonlinear for each characteristic among the ten most important. On the contrary and as expected, we observe a linear relationship for L-En. Additionally, Figure IA11.5 reveals that there are instances where L-En's predictions appear as averages of N-En's predictions (e.g., for delta), restricted by the imposed linear functional form. On the other hand, we observe instances with an entire level shift in the functional form (e.g., for rv and mid).

Sensitivity to Volatility and Jump Risk. A large part of the literature manifests 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.[34] 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 IA11.1 reports these results. In general we find that jump risk proxies impact short-term and especially out-of-the-money 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

[34]For volatility risk, we use the volatility of the at-the-money implied volatility per underlying, as in Baltussen, van Bekkum, and van der Grient (2018) (ivvol), 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.

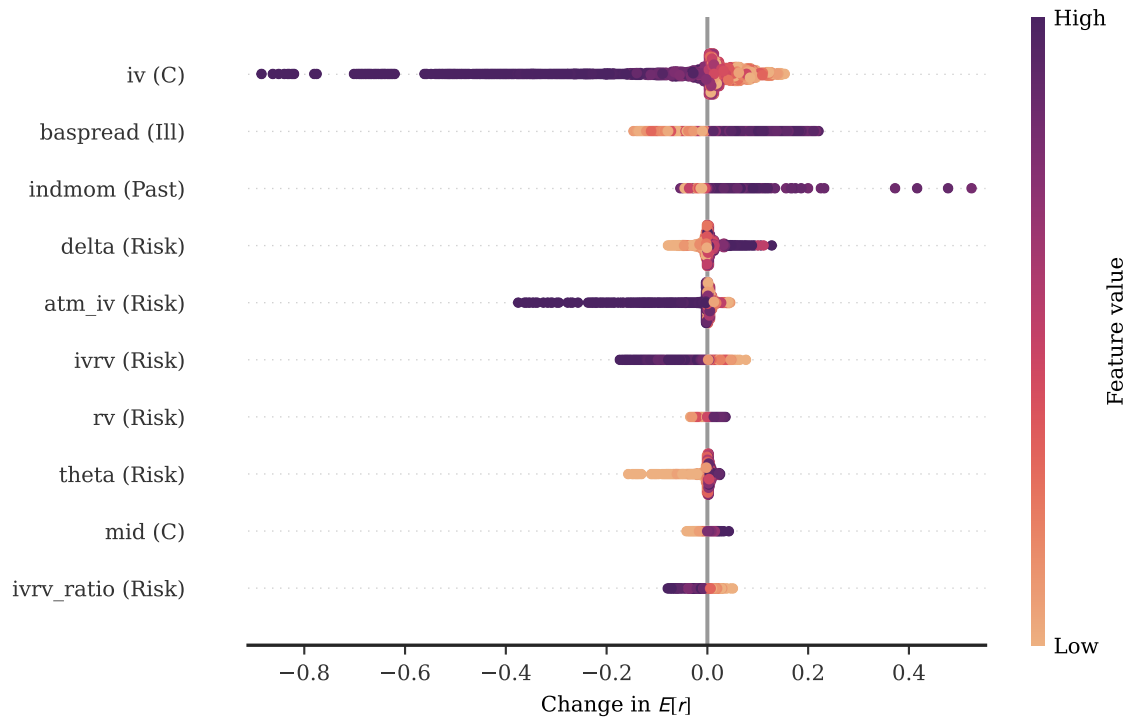

Fig. 9. Impact of the Most Important Characteristics on Predicted Delta-Hedged Returns for N-En

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). The feature group associated with the feature as defined in Appendix IA6 is given in parentheses. The abbreviations used: C=Contract, Ill=Illiquidity, Past=Past Prices.

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 10 visualizes the functional form of how sensitive N-En's predictions are to changes in the most important volatility (ivrv) and jump risk proxies (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 within the five highest ivrv percentiles. 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

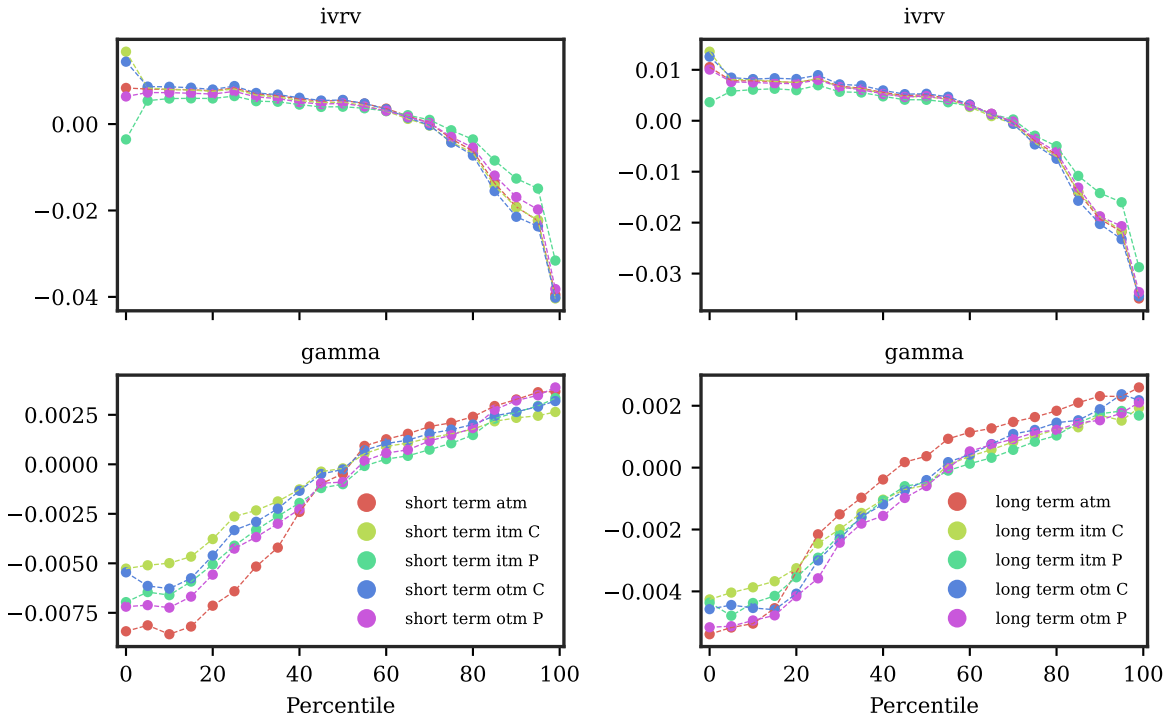

Fig. 10. Relative Impact of Volatility and Jump Risk Premia Per Bucket

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 Bali and Hovakimian (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 exhibit the largest and most diverse sensitivity to changes in gamma. The functional form for long-term options differs little for different moneyness regions. Figure IA11.6 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.

6.4. 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 individual

contract-level (B+I) and contrast the resulting predictability with that of the full information set.[35]

Figure 11 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 80 option-based characteristics is substantial, given that the R^2_{OS} drops from 2.45% to 1.99% (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 0.11%. 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 1.69% (B+I) to 1.99% (O). Consistent with the feature group importance shown in Figure 7, contract-based characteristics are highly important, but alone do not suffice to forecast future single-equity option returns.[36]

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

[35]Note that individual contract-level characteristics ("I") in this setting are not to be confused with group "C" in Figure 7. The groups in Figure 7 relate to a characteristic's information. For example, the vega of an option is placed in group "Risk". In contrast, in this section, we refer to individual contract-based characteristics ("I") as those that operate on the level of the contract, and do not relate to the underlying stock or a bucket of options. Hence, the vega of the option is placed in group "I" in this setting.

[36]Our findings are robust to using the cross-sectional out-of-sample $R^2_{OS;XS}$ (Figure IA12.1) and focusing on option buckets (Table IA12.1). 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.

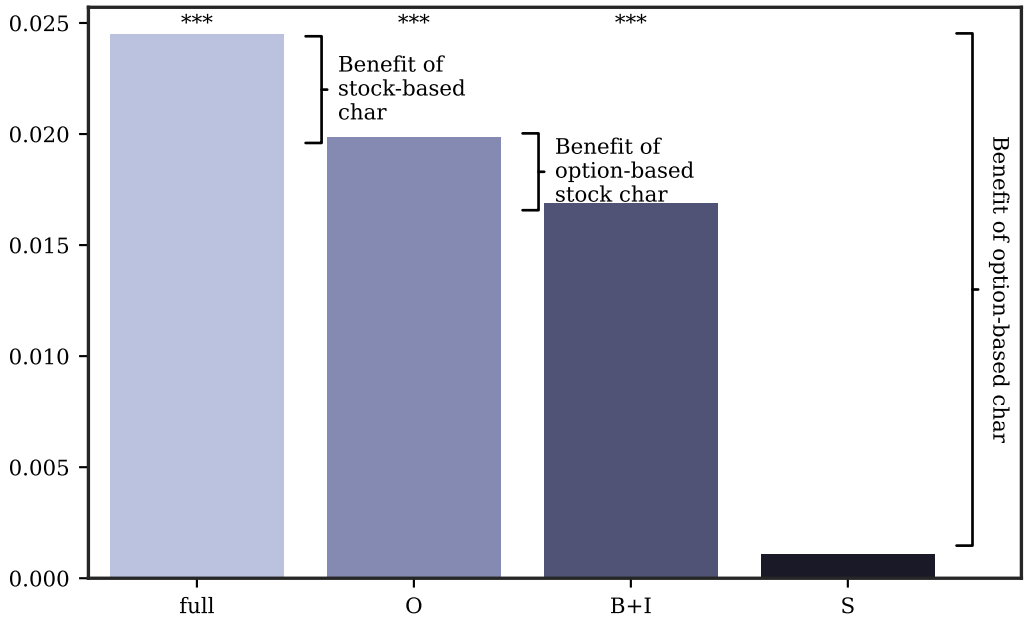

Fig. 11. Restricting the Information Set for N-En

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 individual contract-based information (B+I) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA6. ***, **, * 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.

allowing for nonlinearities helps when using only a restricted set of characteristics.

Panel A of Table 7 reports the 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+I) 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 forecasts of the linear and nonlinear ensemble, we find that N-En restricted to option-based information (O) manages to surpass all L-En models at the 10% significance level or better. 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 (*i*) more information is always better and (*ii*) that nonlinear interactions

Panel A: Diebold and Mariano (1995) Forecast Comparison

	L-En			N-En			
	O	S	full	B+I	O	S	full
L-En: B+I	**3.07**	**-2.87**	**4.39**	**2.26**	**2.80**	0.05	**3.48**
L-En: O		**-4.82**	**3.59**	1.84	**2.47**	−0.63	**3.23**
L-En: S			**7.04**	**3.42**	**4.08**	1.24	**4.78**
L-En: full				1.17	1.88	−1.74	**2.79**
N-En: B+I					**2.38**	**-3.97**	**3.26**
N-En: O						**-5.13**	**2.13**
N-En: S							**7.42**

Panel B: Forecast Correlation

	L-En			N-En			
	O	S	full	B+I	O	S	full
L-En: B+I	**0.90**	0.50	**0.80**	0.62	0.57	0.33	0.55
L-En: O		0.47	**0.86**	0.60	0.65	0.33	0.62
L-En: S			0.58	0.32	0.31	0.61	0.35
L-En: full				0.55	0.61	0.41	0.68
N-En: B+I					**0.87**	0.38	**0.79**
N-En: O						0.38	**0.85**
N-En: S							0.53

Table 7: 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 individual contract-based information (B+I) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA6. 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%).

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 findings from the Diebold and Mariano (1995) comparison carry over to evidence from forecast correlations in Panel B of Table 7. We find the largest similarity of the resulting predictions for linear or nonlinear ensembles estimated on option-based information (both B+I 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.[37]

[37]For the sake of completeness, Table IA12.1 replicates this analysis for different option buckets. For most buckets we find that the full model produces the most accurate predictions.

6.5. Robustness Checks

6.5.1. Trading on Machine Learning Portfolios

Earnings Announcements. Engelberg, McLean, and Pontiff (2018) document that stock market anomalies are 50% higher on corporate news days and six times higher on earnings announcement days, indicating that stock return predictability is more consistent with a mispricing explanation or driven by biased expectations which are partly corrected upon news arrival.

To gauge the effect of earnings announcements and news on return spreads in our setting, we use a subsample analysis similar to Zhan et al. (2022), but on the set of weekly short-term at-the-money options (Section 6.5.4). Precisely, in each month, one subsample contains short-term at-the-money options whose underlying stocks experience an earnings (news) announcement during that week, while the other subsample contains short-term at-the-money options on stocks without an earnings (news) announcement during that week. News days are identified using the Dow Jones version of Ravenpack News Analytics. News are only recorded if the relevance score is 100 and if they are highly positive (sentiment score above 0.75) or highly negative (sentiment score below 0.25).

Figure IA10.3 shows the realized return spreads of the high-minus-low portfolio for the different subsamples and produces a number of interesting results: first, the realized return spread is nearly three times larger for earnings announcement weeks for predictions made by both L-En and N-En. Second, the returns of the high-minus-low portfolio are still significantly different from zero for non-earnings announcement weeks. Third, N-En yields higher high-minus-low portfolio returns regardless whether earnings are announced or not. Fourth, we find a similar, but less pronounced effect for news announcements as the realized returns of high-minus-low portfolios are approximately 50% higher for news weeks. These findings suggest that our uncovered option return predictability is at least partially driven by option mispricing.

Subsamples. We investigate in how far the profitability of our machine learning portfolios changes with the state of the economy. The results in Table IA10.5 show that

realized returns are generally amplified during bad states of the economy and that N-En manages to outperform 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.

Risk Attribution. A possible explanation for our 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 Hi-Lo 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 including the CAPM, the Fama and French (2015) five factor model enhanced by the momentum factor of Carhart (1997) (Fama and French, 2018) and additionally by the liquidity factor of Pástor and Stambaugh (2003), a model using the Agarwal and Naik (2004) option-market factors, a model with the intermediary leverage bearing capacity of Grünthaler, Lorenz, and Meyerhof (2022), and finally the factor model of Bali, Chabi-Yo, and Murray (2022) proposed for stock returns based on option prices.

The results are provided in Table IA10.6. Risk-adjusted returns for all option types are virtually unchanged compared to average raw returns. The risk exposure picked up by the candidate models does not suffice to explain the return spreads generated by the nonlinear machine learning methods.

6.5.2. Variations to the Training Window

Fixed-Length Training Window. As in Gu et al. (2020), we train the models using an expanding training sample. We refit the models each year and correspondingly increase the size of the training window by one year after each iteration. As a robustness check, we consider a rolling training sample with a fixed length of 10 years. Figure IA13.1 depicts the statistical performance for N-En. The R^2_{OS} declines from 2.5% to 1.75% for the testing sample between 2008 and 2020, when compared to using an expanding

training sample.[38] The reduction is also prevalent for calls and puts, as well as for the cross-sectional predictability, $R^2_{OS;XS}$. Whereas realized returns of the high-minus-low portfolio are higher for the expanding training sample, the realized Sharpe ratio is in fact higher using a rolling training scheme, driven by a lower fluctuation of realized returns (see Table IA13.1). However, neither model statistically dominates the other, neither for the full sample, nor individually for put or call options.

Excluding vs. Including the Great Financial Crisis. Next, we re-estimate the models excluding information about option returns during the financial crisis in 2008 and 2009. Subsequently, we compare the resulting performance with our baseline model specification for the testing years of 2010 through 2020. We observe a slightly higher statistical performance (Figure IA13.2) and return spread of the Hi-Lo portfolio when including 2008 and 2009 in the training sample (Table IA13.2). However, the resulting return spreads of the high-minus-low portfolios are statistically indistinguishable.[39]

Summing up, these results suggest that the model's ability to estimate a meaningful connection between option characteristics and future option returns does not materially hinge on the way we set up the training and validation samples and instead is robust to alterations of it.

6.5.3. Restricting to the 500 Largest CRSP Stocks

Next, we focus only on the 500 largest stocks of the entire CRSP universe determined by market capitalization of individual stocks in each month. While the overall level of predictability R^2_{OS} is lower, the return predictions made by the nonlinear ensemble remain useful. N-En yields a positive R^2_{OS} and a significantly positive $R^2_{OS;XS}$ (Figure IA14.1). This compares well to L-En, which fails to yield significant predictability. The average realized excess returns of N-En's predictions are 1.73% per month with a monthly Sharpe ratio of 0.76, significantly outperforming the high-minus-low portfolio returns of the L-En (Table IA14.1). These conclusions are robust to considering call and put options

[38] The testing sample lasts from 2008 to 2020 as we require 10 years to initially train the competing models and 2 years for validation.

[39] In unreported results we find that the agreement between expected returns and decile portfolio assignments between the models fitted with and without information about the financial crisis is high.

separately.

6.5.4. Weekly Investment Period

In this section we consider a shorter investment period of one week instead of one month. We focus the analysis on short-term at-the-money options which are most actively traded and which have low transaction costs (Garleanu et al., 2009; Zhan et al., 2022).

Figure IA15.1 reports the resulting predictability of the linear and nonlinear ensemble when predicting weekly option returns. We find higher predictability compared to the sample of monthly option returns. While L-En consistently yields out-of-sample R^2_{OS} at or above 3%, predictability of N-En is almost twice that high. These findings are robust to the cross-sectional $R^2_{OS;XS}$ and hold regardless of whether we consider puts, calls, or both. In line with higher predictability, we also document an improvement in the economic performance (Table IA15.1). Both L-En and N-En generate *weekly* Sharpe ratios above 1.3, doubling the monthly Sharpe ratios obtained in the baseline results. Furthermore, both N-En's short and long portfolios yield significantly higher returns than predictions based on L-En, stressing the importance of modeling nonlinear and interaction effects for option returns also at higher frequencies.

6.5.5. Different Return Definitions

Margin-adjusted Returns. The main analysis throughout the paper uses delta-hedged gains scaled by the cash requirement of opening the delta-hedged option position. Garleanu and Pedersen (2011) and Hitzemann, Hofmann, Uhrig-Homburg, and Wagner (2021) show that margin requirements for option positions are large and suggest evaluating returns using margin requirements as the denominator in the return definition of Equation (12). We consequently adopt the CBOE minimum margin for costumer accounts. Details on the calculation of margin requirements are given in Appendix IA16.1.

Refitting the models on returns scaled by margin requirements of long options positions leaves our baseline results mostly unchanged. Statistical performance is shown in Figure IA16.1. Table IA16.1 reports the economic performance with high-minus-low

portfolios before transaction costs whereas Table IA16.2 includes transaction costs and accounts for margin requirements of short option positions.

Delevered Returns. Besides margin-adjusted returns, we assess how the trading strategy profits change if we account for time-variation in the leverage of the traded options following Frazzini and Pedersen (2022). Table IA16.3 shows average realized excess returns and Sharpe ratios of the decile portfolios sorted by N-En's predictions, after we account for the embedded leverage. We account for leverage in three ways: first, we scale the realized profits of each portfolio by the average leverage of the options contained in portfolio p. We find that realized returns when expressed per unit of leverage decrease to 0.55% per month. However, the Sharpe ratio is virtually unchanged at 1.25 (1.28 with leverage), suggesting that risk-adjusted profits are not driven by differences in the average leverage of the high or low portfolio. Separate results for calls and puts confirm this finding. Second, we account for time-variation of the portfolio leverage, by scaling the realized returns by the average leverage of portfolio p measured in the month of trade initiation (t). We find that month-specific differences in the portfolio leverage result in slightly lower Sharpe ratios for the high-minus-low portfolio (0.95). However, the Sharpe ratio and average realized returns are still highly significant. Lastly, we also account for the time-varying leverage of each individual option contract o. The resulting high-minus-low returns per unit of leverage amount to 1.21% per month with a Sharpe ratio of 0.65. The realized monthly Sharpe ratio decreases by about half if we account for time-variation in the contract-specific leverage, but is still large for unit leverage of both the short and long leg.

6.5.6. 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 only at-the-money and short-term contracts.

We address this potential criticism in three ways. First, Figure IA17.1 shows the

R^2_{OS} comparison between the linear and nonlinear ensemble for each option bucket considered. Table IA17.1 provides the trading strategy 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 machine learning models. Throughout all option buckets but short-term otm calls, we find significantly higher raw and risk-adjusted returns for N-En than for L-En.

Second, we investigate the predictive power of N-En for predicting the returns of option portfolios and compare it with predicting returns for the individual contracts within that portfolio (Figure IA17.2). The construction of portfolios follows the bucket definition in Section 5. It yields 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) contracts 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; Zhan et al., 2022; Goyenko and Zhang, 2021). We find that we can predict a higher share of portfolio returns within the nonlinear ensemble (which is fitted on individual contracts), especially for short-term options. Overall, however, predictability, measured both by R^2_{OS} and $R^2_{OS;XS}$, is comparable for individual contracts and option portfolios.

Third, we run a full-on comparison between N-En estimated using all options in our sample and a nested model specializing on short-term at-the-money options. Appendix IA18 shows comparable levels of overall predictability and little drawbacks of predicting returns of all options simultaneously.

7. Sources of Option Return Predictability

Hong and Stein (1999) propose a theoretical model in which gradual diffusion of information among investors explains the observed predictability of asset returns. In their model, at least some investors can process only a subset of publicly available information because either they have limited information-processing capabilities or searching over all

possible forecasting models using publicly available information itself is costly (Hirshleifer and Teoh, 2003), and there are limits to arbitrage (Shleifer and Vishny, 1997; Pontiff, 2006). Due to investors' limited attention and informational frictions, new informative signals are incorporated into asset prices partially because at least some investors do not adjust their demand by recovering informative signals from observed prices. As a result of this failure on the part of some investors, asset returns exhibit predictability. In this section, we investigate potential economic mechanisms underlying the sources of option return predictability. In particular, we test whether informational frictions and option mispricing provide an explanation to the observed return predictability in the options market.

As shown in Appendix IA2 of the Internet Appendix, the expected return to selling a delta-neutral call (put) is the weighted average of the expected return on underlying stock and the expected return on call (put) option. Hence, we argue that both stock and option characteristics can be viewed as potential determinants of the cross-sectional differences in delta-hedged option returns. Since the literature on option pricing does not provide clear theoretical guidance on how delta-hedged expected returns should look like, and combined with the considerations put forth in Section 3, we conjecture that the expected return on delta-hedged option positions can be a highly nonlinear function stock and option characteristics. Thus, it is possible that delta-hedged option return predictability can be driven by information frictions, limits-to-arbitrage, and mispricing in both the underlying equity and option markets.

7.1. Informational Frictions

We analyze if the return predictability is concentrated in options with higher levels of informational frictions. 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.

Rather than relying on a single proxy for information frictions, we follow Atilgan, Bali, Demirtas, and Gunaydin (2020) and construct an arbitrage cost index using a number of

indicators known to capture several dimensions of limits-to-arbitrage. First, we build an index on the stock-level, for which we include firm size, firm age, idiosyncratic volatility of the underlying, institutional ownership (Ofek et al., 2004; Nagel, 2005; Eisdorfer et al., 2022), and analyst coverage (Zhang, 2006). To construct the arbitrage cost index, we sort stocks in increasing order based on their idiosyncratic volatility. Similarly, we sort stocks in decreasing order based on their level of institutional ownership, analyst coverage, size and age, since lower values of these variables indicate higher arbitrage costs. Each stock is given the corresponding score of its decile rank for each variable. Finally, the arbitrage cost index on the stock-level is the sum of the five scores such that it ranges from 5 to 50. A higher value implies tighter limits-to-arbitrage.

Similarly, we build an alternative arbitrage cost index on the option-level. We use option illiquidity, option bucket trading volume, dollar open interest, and margin requirements. Moreover, Tian and Wu (2021) argue that delta-hedging an option position exposes the option investor to three primary risks: the delta-hedging costs, stochastic volatility risk, and random jump risk. Consequently, we add the bid-ask spread of the underlying stock, volatility of implied volatility, and excess kurtosis of the underlying stock as variables for the index. While constructing the index on the option-level, we adjust the sorting mechanism to ensure that higher values indicate stricter limits to arbitrage.

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 the arbitrage index on the stock-level, or on the option-level (Q1–Q5). Subsequently, we contrast the predictability of options written on stocks with different levels of the arbitrage indices. Figure 12 depicts the R^2_{OS} values for the sub-samples. Higher informational frictions directly translate to higher predictability of option contracts by N-En, confirming our hypothesis. Options on underlyings with the highest levels of information frictions show the highest predictability at $R^2_{OS} = 5.32\%$. Instead, for options written on stocks with the lowest level of information frictions, R^2_{OS} is indistinguishable from zero. The predictability uncovered is statistically significant only for the three highest information friction quintiles. Turning

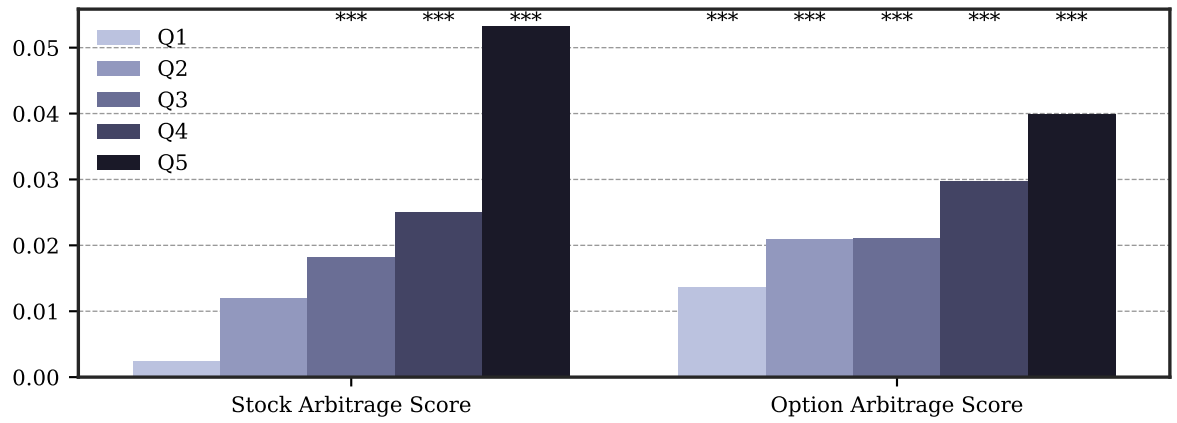

Fig. 12. Predictability Conditional on Information Frictions

The figure shows the predictability using R^2_{OS} of the nonlinear ensemble for options sorted into quintiles by an index of informational frictions on the underlying-level and an index of informational frictions on the option-level. Index constructions follows Atilgan et al. (2020). Firm size, firm age, idiosyncratic volatility, institutional ownership, and analyst coverage are used to construct the index on the stock-level. The option contract's bid-ask spread, margin requirement, dollar open interest, bucket volume, the volatility of the implied volatility, the historical excess-kurtosis of the underlying, and the underlying's bid-ask spread are taken for the index construction on the option-level. Since the level of institutional ownership, analyst coverage, firm age, firm size, dollar open interest, and bucket volume are inversely related to informational frictions, these characteristics are inversely sorted in the indices construction.

to the arbitrage index on the option-level, option return predictability is monotonically increasing in the contract-specific limits to arbitrage. In the highest quintile, we find $R^2_{OS} = 4.00\%$, which drops to 1.36% in Q1. Figure IA19.1 in the Online Appendix shows similar findings in the case of cross-sectional out-of-sample predictability.

Additionally, we perform bivariate portfolio sorts to study how realized option returns depend on informational frictions. We sort options into quintiles based on the friction index on the stock or option-level 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 8 presents the results for the 25 (5x5) portfolios sorted by expected returns conditional on information frictions on the stock-level. High-minus-low return spreads are significantly positive for each of the stock arbitrage quintiles. However, realized excess returns are an increasing function of informational frictions in the underlying. For the lowest friction levels, average returns amount to 0.75% per month, for the highest level of frictions to 3.34%, which leads to an economically and statistically significant diff-in-diff return spread of 2.59% per month.

	Low Pred.	2	3	4	High Pred.	H-L
			Stock Arbitrage Score			
Low	-0.650***	-0.273**	-0.168	-0.092	0.103	0.754***
2	-0.891***	-0.335**	-0.199	-0.082	0.221	1.112***
3	-1.125***	-0.402***	-0.190	-0.080	0.349	1.474***
4	-1.371***	-0.436***	-0.191	0.043	0.592**	1.963***
High	-2.529***	-0.857***	-0.313*	0.062	0.813***	3.342***
H-L	-1.878***	-0.584***	-0.145**	0.154**	0.710***	2.588***
			Option Arbitrage Score			
Low	-1.105***	-0.316**	-0.142	-0.023	0.275	1.380***
2	-1.214***	-0.266*	-0.076	0.038	0.399**	1.613***
3	-1.249***	-0.276**	-0.049	0.089	0.483**	1.732***
4	-1.388***	-0.293**	-0.037	0.122	0.615***	2.003***
High	-1.487***	-0.362***	-0.073	0.188	0.767***	2.255***
H-L	-0.382***	-0.046	0.069	0.211***	0.492***	0.874***

Table 8: Bivariate Portfolios of Information Frictions and Expected Returns

The table shows the returns to option portfolios first sorted by an index of informational frictions, either on the stock-level (upper panel) or on the option-level (lower panel) and then by the return predictions made by the nonlinear ensemble method. Index constructions follows Atilgan et al. (2020). Firm size, firm age, idiosyncratic volatility, institutional ownership, and analyst coverage are used to construct the index on the stock-level. The option contract's bid-ask spread, dollar open interest, bucket volume, the volatility of the implied volatility, the historical excess-kurtosis of the underlying, and the underlying's bid-ask spread are taken for the index construction on the option-level. Since the level of institutional ownership, analyst coverage, firm age, firm size, dollar open interest, and bucket volume are inversely related to informational frictions, these characteristics are inversely sorted in the indices construction. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level.

Panel B of Table 8 replicates this analyses for the option-level friction score. The high-minus-low return spreads are monotonically increasing in the level of the friction index, confirming our hypotheses that option return predictability is directly related to information frictions. Furthermore, this leads to a significant difference between the high-minus-low spreads of the highest and lowest limits-to-arbitrage quintile of 0.87% per month.

7.2. Option Mispricing

We expect to find higher levels of predictability for options that are priced incorrectly as the nonlinear ensemble manages to identify these opportunities and correctly proposes shorting over- and purchasing undervalued contracts. In the spirit of the previous section on limits-to-arbitrage, we again refrain from taking a stand on which variable constitutes option-level mispricing, but instead use a composite score. Note that we are investigating

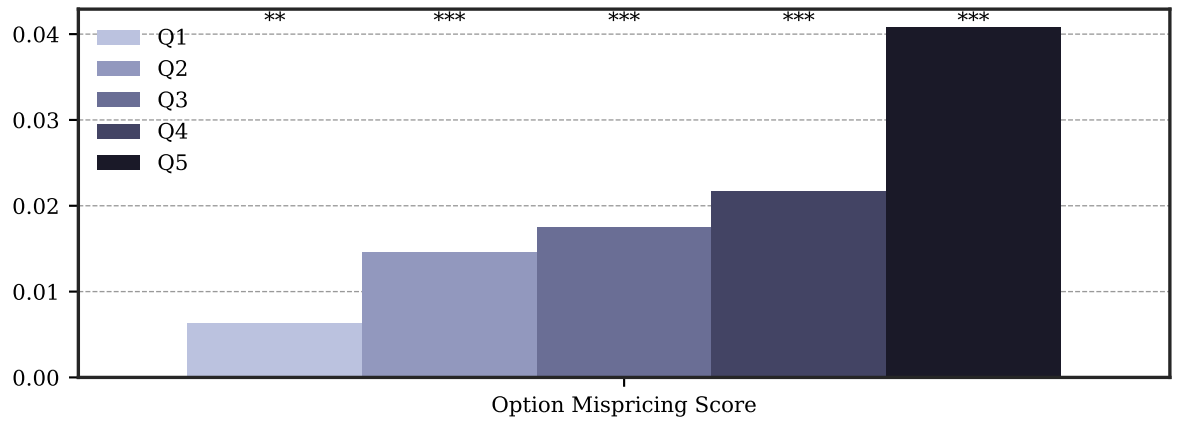

Fig. 13. Predictability and Profitability Conditional on Option Mispricing – R^2_{OS}

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 absolute option mispricing using a composite mispricing score. As inputs, we use $iv - rv$ (Goyal and Saretto, 2009; Carr and Wu, 2009), the mispricing measure by Eisdorfer et al. (2022), as well as the absolute return prediction of the nonlinear ensemble. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7).

absolute levels of mispricing.

The first input to the mispricing index is $iv - rv$, studied by Goyal and Saretto (2009) and Carr and Wu (2009). Next, we follow Eisdorfer et al. (2022) 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., $|\text{Mispricing}| = \left|\log(O/\widetilde{O})\right|$, where $\widetilde{O}$ denotes the theoretical and O the observed mid price. Averaging over all short-term at-the-money options, we obtain one level of mispricing per underlying stock at each point in time. Our last mispricing measure is the absolute value of the return forecast of the nonlinear ensemble made in month t.

We once again sort options into quintiles, this time by the composite mispricing score, and show the resulting predictability patterns for future returns over the next month in Figure 13.[40] For low levels of mispricing, R^2_{OS} is close to zero. In contrast, for the options with the highest mispricing score, R^2_{OS} is 4.09% per month.

[40]Results for the cross-sectional $R2_{OS;XS}$ are provided in Figure IA19.2.

	Low Pred.	2	3	4	High Pred.	H-L
			Option Mispricing Score			
Low	-0.424***	-0.095	0.003	0.069	0.335*	0.759***
2	-0.721***	-0.141	0.003	0.141	0.509**	1.230***
3	-0.935***	-0.241*	-0.028	0.171	0.580**	1.515***
4	-1.506***	-0.500***	-0.214	0.064	0.650***	2.156***
High	-2.348***	-0.952***	-0.553***	-0.238	0.392*	2.740***
H-L	-1.924***	-0.857***	-0.555***	-0.307***	0.056	1.981***

Table 9: 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 absolute option mispricing using a composite mispricing score. As inputs, we use $iv - rv$ (Goyal and Saretto, 2009; Carr and Wu, 2009), the mispricing measure by Eisdorfer et al. (2022), as well as the absolute return prediction of the nonlinear ensemble. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level.

In Table 9, we look at the resulting realized returns of portfolios first sorted by the option mispricing score and then by N-En's expected return forecast. While the high-minus-low return spread is significant for all levels of option mispricing, it monotonically increases with the degree of mispricing. In fact, the monthly return difference between the H-L spreads using options with the highest versus the lowest mispricing scores amounts to 1.98% per month.

Overall, these results suggest that option return predictability is largely driven by informational frictions, limits-to-arbitrage, and mispricing in the options market. However, given the connection between risk, arbitrage cost, and proxies for information frictions, we cannot rule out potential risk-based explanations. Moreover, as discussed in Section 6.3, we find that nonlinear option risk measures such as jump and volatility risk are important determinants of option returns. It is also well-known that illiquid stocks tend to have high market beta and high firm-specific volatility. Moreover, illiquid stocks and options are known to exhibit skewed fat-tailed return distributions with significant volatility and jump risk premia.[41] Thus, higher values of the information frictions index proposed in the paper indicate stricter limits-to-arbitrage and higher level of riskiness so that the significantly large abnormal returns on portfolios of equity options can partly be compensation for volatility, jump, and liquidity risks in the options market.

[41]The interested reader may wish to consult Amihud (2002), Xing et al. (2010), An et al. (2014), Cremers et al. (2015), Baltussen et al. (2018), Christoffersen et al. (2018b), Atilgan et al. (2020), and Zhan et al. (2022) for empirical evidence on the cross-sectional relations between liquidity, market risk, volatility, and higher-order moments of individual stocks and options.

8. 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 techniques to predict individual U.S. equity option returns using a set of 80 option-based and 193 stock-based characteristics in the period from 1996 to 2020. Empirically, we derive several results that enhance our knowledge on the cross-sectional pricing of equity options. First, we show that the complexity of the machine learning models matters 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 and underlyings with higher limits-to-arbitrage and higher degree of mispricing.

References

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). Working paper.

Atilgan, Y., Bali, T. G., Demirtas, K. O., Gunaydin, A. D., 2020. Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns. Journal of Financial Economics 135, 725–753.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

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., Chabi-Yo, F., Murray, S., 2022. A factor model for stock returns based on option prices. Working paper.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2021. Predicting corporate bond returns: Merton meets machine learning. Working paper.

Bali, T. G., Hovakimian, A., 2009. Volatility spreads and expected stock returns. Management Science 55, 1797–1812.

Bali, T. G., Murray, S., 2013. Does risk-neutral skewness predict the cross-section of equity option portfolio returns? Journal of Financial and Quantitative Analysis 48, 1145–1171.

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.

Bates, J. M., Granger, C. W., 1969. The combination of forecasts. Journal of the Operational Research Society 20, 451–468.

Bekaert, G., Engstrom, E. C., Xu, N. R., 2021. The time variation in risk appetite and uncertainty. Management Science 68, 3975–4004.

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.

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.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Büchner, M., Kelly, B., 2022. A factor model for option returns. Journal of Financial Economics 143, 1140–1161.

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., Vasquez, A., Xiao, X., Zhan, X., 2019. Volatility Uncertainty and the Cross-Section of Option Returns. Working paper.

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

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

Chen, L., Pelger, M., Zhu, J., 2020. Deep learning in asset pricing. Working paper.

Christoffersen, P., Fournier, M., Jacobs, K., 2018a. The factor structure in equity options. Review of Financial Studies 31, 595–637.

Christoffersen, P., Goyenko, R., Jacobs, K., Karoui, M., 2018b. 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.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J., Santos, A. A. P., 2021. Can machine learning help to select portfolios of mutual funds? Working paper.

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. Working paper.

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., 2015. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. Journal of Business & Economic Statistics 33, 1–1.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 20, 134–144.

Dittmar, R. F., 2002. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. Journal of Finance 57, 369–403.

Eisdorfer, A., Goyal, A., Zhdanov, A., 2022. Limited attention and option prices. Working paper.

Engelberg, J., McLean, R. D., Pontiff, J., 2018. Anomalies and news. Journal of Finance 73, 1971–2001.

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. Working paper.

Frazzini, A., Pedersen, L. H., 2022. Embedded leverage. Review of Asset Pricing Studies 12, 1–52.

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., 2011. Margin-based asset pricing and deviations from the law of one price. Review of Financial Studies 24, 1980–2022.

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 .

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. Working paper.

Grammig, J., Hanenberg, C., Schlag, C., Sönksen, J., 2020. Diverging roads: Theory-based vs. machine learning-implied stock risk premia. Working paper.

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., 2022. The leverage bearing capacity: A new tool for intermediary asset pricing. Working paper.

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. Working paper.

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.

Heston, S. L., Jones, C. S., Khorram, M., Li, S., Mo, H., 2022. Option momentum. Journal of Finance, forthcoming.

Heston, S. L., Nandi, S., 2000. A closed-form garch option valuation model. Review of Financial Studies 13, 585–625.

Heston, S. L., Sadka, R., 2008. Seasonality in the cross-section of stock returns. Journal of Financial Economics 87, 418–445.

Hirshleifer, D., Teoh, S. H., 2003. Limited attention, information disclosure, and financial reporting. Journal of Accounting and Economics 36, 337–386.

Hitzemann, S., Hofmann, M., Uhrig-Homburg, M., Wagner, C., 2021. Margin requirements and equity option returns. Working paper.

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.

Hull, B., Li, A., Qiao, X., 2021. Option pricing via breakeven volatility. Working paper.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2022. Is there a replication crisis in finance? Journal of Finance, forthcoming.

Jondeau, E., Rockinger, M., 2001. Gram–charlier densities. Journal of Economic Dynamics and Control 25, 1457–1483.

Kanne, S., Korn, O., Uhrig-Homburg, M., 2020. Stock illiquidity and option returns. Working paper.

Kelly, B. T., Palhares, D., Pruitt, S., 2020a. Modeling corporate bond returns. Journal of Finance, forthcoming.

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. Working paper.

Keloharju, M., Linnainmaa, J. T., Nyberg, P., 2016. Return seasonalities. Journal of Finance 71, 1557–1590.

Kimball, M. S., 1993. Standard risk aversion. Econometrica 61, 589–611.

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. Working paper.

Leippold, M., Wang, Q., Zhou, W., 2021. Machine-learning in the chinese factor zoo. Journal of Financial Economics 145, 64–82.

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. Working paper.

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., 2022. Market efficiency in the age of big data. Journal of Financial Economics 145, 154–177.

Moritz, B., Zimmermann, T., 2016. Tree-based conditional portfolio sorts: The relation between past and future stock returns. Working paper.

Muravyev, D., Pearson, N. D., 2020. Options trading costs are lower than you think. Review of Financial Studies 33, 4973–5014.

Murray, S., Xiao, H., Xia, Y., 2021. Charting by machines. Working paper.

Nagel, S., 2005. Short sales, institutional investors and the cross-section of stock returns. Journal of Financial Economics 78, 277–309.

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. Working paper.

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.

Pratt, J. W., Zeckhauser, R. J., 1987. Proper risk aversion. Econometrica 55, 143–154.

Qu, R., Timmermann, A., Zhu, Y., 2022. Comparing forecasting performance with panel data. Working paper.

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

Schlögl, E., 2013. Option pricing where the underlying assets follow a gram/charlier density of arbitrary order. Journal of Economic Dynamics and Control 37, 611–632.

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 primary risk taking in derivative securities. Working paper.

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. Working paper.

Vilkov, G., Xiao, Y., 2012. Option-implied information and predictability of extreme returns. Working paper.

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.

Zhan, X., Han, B., Cao, J., Tong, Q., 2022. Option return predictability. Review of Financial Studies 35.

Zhang, X. F., 2006. Information uncertainty and stock returns. Journal of Finance 61, 105–137.

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.

Internet Appendix

Option Return Predictability with Machine Learning and Big Data

by Turan G. Bali, Heiner Beckmeyer, Mathis Moerke, Florian Weigert

Table of Contents:

- Appendix IA16 alters the return definition taking **margin requirements** and **deleveraged returns** into account, respectively.
- Appendix IA17 provides additional information on statistical and economic performance for **options buckets**.
- Appendix IA18 provides additional information on predicting **ATM** options only.
- Appendix IA19 provides additional information for the **sources of option return predictability**.

Appendix IA1. Importance of Nonlinearities and Interactions for Predicting Option Returns

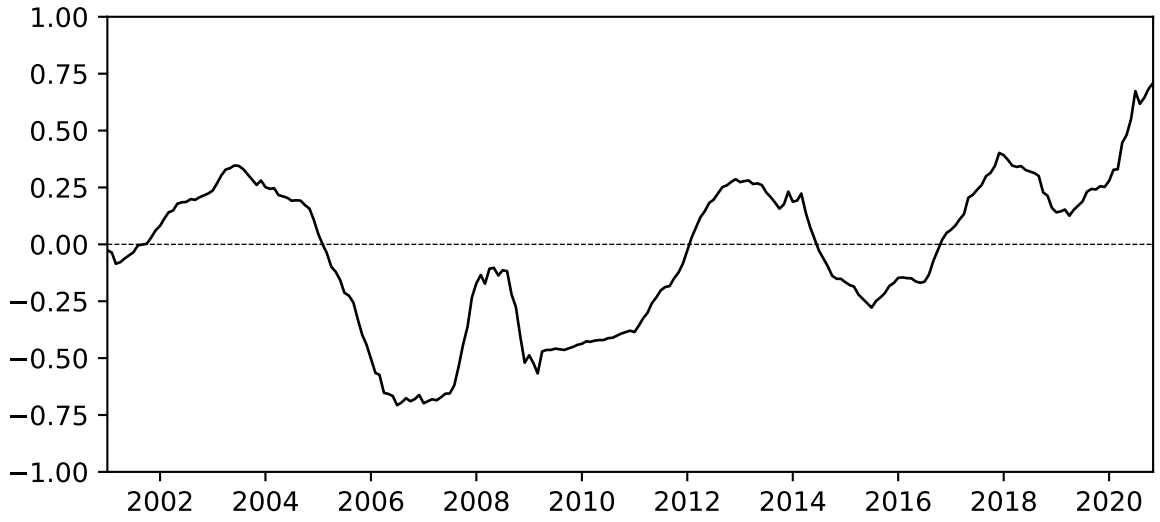

Fig. IA1.1. Rolling Correlation Between Implied Volatility and Underlyings' Idiosyncratic Volatility

The figure shows the rolling correlation between aggregate implied volatility and aggregate underlyings' idiosyncratic volatility. Aggregation uses value-weighting. The correlation is calculated over five years of monthly data on a rolling basis.

Appendix IA2. Expected Delta-Hedged Option Returns

The return to selling a delta-neutral call over $[t, t+1]$ is

$$HPR_{t+1} = \frac{\Delta_t S_{t+1} - C_{t+1}}{\Delta_t S_t - C_t} - 1, \tag{IA1}$$

with $C_t = \frac{1}{R_{f,t+1}} \mathbb{E}_t \left[\max(S_{t+1} - K, 0)\right]$, $C_{t+1} = \frac{1}{R_{f,t+2}} \mathbb{E}_{t+1} \left[\max(S_{t+2} - K, 0)\right]$, and K the option's strike price.

The initial investment cost is $\Delta_t S_t - C_t$ and the payoff at the end of the holding period is $\Delta_t S_{t+1} - C_{t+1}$, such that we can rewrite the holding period return at time $t+1$ as:

$$HPR_{t+1} = \frac{\Delta_t S_t}{\Delta S_t - C_t} \cdot \frac{S_{t+1}}{S_t} - \frac{C_t}{\Delta S_t - C_t} \cdot \frac{C_{t+1}}{C_t} - 1 \tag{IA2}$$

$$= w_t \cdot \frac{S_{t+1}}{S_t} - (1 - w_t) \cdot \frac{C_{t+1}}{C_t} - 1, \tag{IA3}$$

with

$$w_t = \frac{\Delta_t S_t}{\Delta_t S_t - C_t} \tag{IA4}$$

Thus, the expected return on delta-hedged option position is defined as:

$$\mathbb{E}_t \left[HPR_{t+1}\right] = w_t \, \mathbb{E}_t \left[R_{t+1}\right] - (1 - w_t) \, \mathbb{E}_t \left[C_{t+1}/C_t\right] - 1 \tag{IA5}$$

Given that the expected return to selling a delta-neutral call is the weighted average of the expected return on underlying stock, $\mathbb{E}_t \left[R_{t+1}\right]$, and the expected return on call option, $\mathbb{E}_t \left[C_{t+1}/C_t\right]$, we argue that both stock and option characteristics can be viewed as potential determinants of the cross-sectional differences in delta-hedged option returns. Since the literature on option pricing does not provide clear theoretical guidance on what delta-hedged expected returns should look like, we conjecture that $\mathbb{E}_t \left[HPR_{t+1}\right]$ can be a highly nonlinear function of stock and option characteristics.

Appendix IA3. 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):

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{i,s,t+1} - g(z_{i,s,t}))^2 \tag{IA6}$$

Appendix IA4. 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 (IA6)) 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).[1] 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 IA4.1 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.

[1] We use the implementation in Ray Tune (Liaw, Liang, Nishihara, Moritz, Gonzalez, and Stoica (2018)). We carry out the 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 IA4.1: 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.

Appendix IA5. Option-Based Characteristics

This section describes a broad set of the 80 option-based characteristics, motivated by earlier studies on the cross-section of option and/or stock returns. Of the 80 we compute, 43 characteristics operate on the level of the underlying stock, 20 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 5.3), and 17 on the level of individual option contracts.

IA5.1. 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 (*rns*τ)**. 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 (*rnk*τ)**. 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 and Li (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, Nguyen, and Whitby (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. **Option demand pressure (*demand_pressure*).** We follow Cao et al. (2019) and measure the option demand pressure by the ratio of option market value (total option open interest times mid price of the contract) and market capitalization for the stock at hand.
14. **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.
15. **Dollar trading volume (*dvol*).** Following Cao and Wei (2010), we include the dollar trading volume across all options,

$$\sum_j VOL_j \times (ask_j + bid_j)/2.$$

16. **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.

17. **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.

18. **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.

19. **Number of traded options (*nopt*).** The average number of options per underlying stock per month.

20. **Total open interest (*toi*).** Open interest across all options on an underlying.

21. **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.

22. **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.

23. **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. Goyal and Saretto (2009) study a related measure.

24. **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.

25. **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.

26. **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.

27. **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.

28. **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.

29. **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.

30. **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***.

31. **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.

32. **Change in weighted put-call spread (*vs_change*)**. Following Cremers and Weinbaum (2010), we compute changes in ***vs_level***.

33. **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.

34. **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.

35. **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).

IA5.2. 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.

1. **At-the-money implied volatility (*atm_iv*).** The maturity-specific at-the-money implied volatility, as the linearly-interpolated average implied volatility of the two options closest to the underlying's current price. Maturity-specific.
2. **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,

$$\text{illiq} = -COV_t(\Delta p_{itd}, \Delta p_{itd+1}).$$

3. **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.

4. **Illiquidity measure based on zero returns (*pzeros*)** As in Lesmond et al. (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.

5. **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.

6. **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.

7. **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*.

8. **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.

9. **Std.dev of the Amihud measure (*stdamihud*)** The standard deviation of the daily Amihud (2002) measure within a month.

10. **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 $gammaps = -\gamma$.

11. **Volatility (*hvol*).** Historical volatility is estimated using daily data over the last month.
12. **Skewness (*hskew*).** Historical skewness is estimated using daily data over the last month.
13. **Kurtosis (*hkurt*).** Historical kurtosis is estimated using daily data over the last month.
14. **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

$$\text{ocgo} = -\frac{O_{t-2} - R_{t-1}}{O_{t-2}}.$$

15. **Open interest vs. stock volume (*oistock*).** As a measure of demand, we compute the ratio of open interest to underlying stock volume.
16. **Volume (*bucket_vol*).** The options volume as the sum of the volume of all options contracts within the bucket.
17. **Dollar volume (*bucket_dvol*).** The options dollar volume as the sum of the dollar volume of all options contracts within the bucket.
18. **Relative volume (*bucket_vol_share*).** ***bucket_vol*** divided by the options volume of all options contracts for the same underlying.
19. **Turnover (*turnover*).** The ratio of options volume to options open interest.
20. **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.

IA5.3. Contract-Level

Data on contract level are obtained from OptionMetrics.

1. **Call indicator (*C*).** Indicator equalling 1 if option is a call option, 0 otherwise.
2. **Put indicator (*P*).** Indicator equalling 1 if option is a put option, 0 otherwise.
3. **Expiration flag (expiration_month).** Indicator equalling 1 if the option expires within the observation month, 0 otherwise.
4. **Time-to-maturity (*ttm*).** The number of calendar years to maturity.
5. **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.

6. **Standardized moneyness (*mdegree*).** Moneyness standardized by the maturity-specific at-the-money implied volatility:

$$mdegree = \frac{\log(K/S_t)}{\sqrt{\tau} \times IV^{\text{atm}}_{t,\tau}}$$

7. **Implied volatility (*iv*).** Following Buchner and Kelly (2020), the Black and Scholes (1973) implied volatility of the option contract.
8. **Delta (*delta*).** Following Buchner and Kelly (2020), 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.
9. **Gamma (*gamma*).** Following Buchner and Kelly (2020), 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.
10. **Theta (*theta*).** Following Buchner and Kelly (2020), 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.
11. **Vega (*vega*).** Following Buchner and Kelly (2020), the Black and Scholes (1973) vega, i.e., the sensitivity of the option with respect to changes in the implied volatil-

ity. We scale *vega* by the price of the underlying stock to make it comparable in the cross-section.

12. **Volga (*volga*).** Following Buchner and Kelly (2020), 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.
13. **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.

14. **Open interest (*oi*).** The open interest of the option contract.
15. **Dollar open interest (*doi*).** The dollar open interest of the option contract.
16. **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.

17. **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.

Appendix IA6. Classification of Characteristics

In this section we provide a detailed summary of the characteristics used in our analyses. We consider a total of 273 characteristics, of which 80 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 5. 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.

Furthermore, we additionally partition the characteristics into different information sets. This information set is provided in the table. We consider four different sets. Characteristics belonging to set "S" relate to stock-based information. "O" contains option-based information. "B" refers to option-based information on the instrument level "Bucket". "I" denotes option-based information on the instrument level "Contract". The union of "B" and "I" is a proper subset of "O".

As a last set of information we also group the characteristics into 12 groups based on an economic meaning. Group "Accruals" contains five characteristics, "Contract" seven, "Frictions" contains four characteristics, "Illiquidity" 29, "Industry" 90, "Informed Trading" 18, "Investment" 11, "Past Prices" 13, "Profitability" 16, "Quality" 29, "Risk" 41, and "Value" 10. The grouping for stock-based characteristics follows the intuition formed by Green et al. (2017) and Jensen et al. (2022). We group the remaining option-based characteristics accordingly.

Table IA6.1: Classification of Characteristics

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
absacc	Absolute accruals	Green et al. (2017)	Stock	Underlying	S	Accruals
acc	Working capital accruals	Green et al. (2017)	Stock	Underlying	S	Accruals
aeavol	Abnormal earnings announcement volume	Green et al. (2017)	Stock	Underlying	S	Profitability
age	# years since first Compustat coverage	Green et al. (2017)	Stock	Underlying	S	Quality
agr	Asset growth	Green et al. (2017)	Stock	Underlying	S	Investment
ailliq	Absolute illiquidity	Cao and Wei (2010)	Option	Underlying	O	Illiquidity
amihud	Amihud illiquidity per bucket	Amihud (2002)	Option	Bucket	B	Illiquidity
atm_civpiv	At-the-money put vs. call implied volatility		Option	Underlying	O	Informed Trading
atm_dcivpiv	Change in atm put vs. call implied volatility	An et al. (2014)	Option	Underlying	O	Informed Trading
atm_iv	At-the-money implied volatility (maturity-specific)		Option	Bucket	B	Risk
baspread	Bid-ask spread	Green et al. (2017)	Stock	Underlying	S	Illiquidity
bear_beta	Bear beta	Lu and Murray (2019)	Stock	Underlying	S	Risk
beta	Beta	Green et al. (2017)	Stock	Underlying	S	Risk
betasq	Beta squared	Green et al. (2017)	Stock	Underlying	S	Risk
bm	Book-to-market	Green et al. (2017)	Stock	Underlying	S	Value
bm_ia	Industry-adjusted book-to-market	Green et al. (2017)	Stock	Underlying	S	Value
bucket_dvol	Option bucket dollar volume		Option	Bucket	B	Illiquidity
bucket_vol	Option bucket volume		Option	Bucket	B	Illiquidity
bucket_vol_share	Relative option bucket volume		Option	Bucket	B	Illiquidity
C	Call indicator		Option	Contract	I	Contract
cash	Cash holdings	Green et al. (2017)	Stock	Underlying	S	Quality
cashdebt	Cash flow to debt	Green et al. (2017)	Stock	Underlying	S	Value
cashpr	Cash productivity	Green et al. (2017)	Stock	Underlying	S	Profitability
cfp	Cash-flow-to-price ratio	Green et al. (2017)	Stock	Underlying	S	Risk
cfp_ia	Industry-adjusted cash-flow-to-price ratio	Green et al. (2017)	Stock	Underlying	S	Risk
chatoia	Industry-adjusted change in asset turnover	Green et al. (2017)	Stock	Underlying	S	Quality
chcsho	Change in shares outstanding	Green et al. (2017)	Stock	Underlying	S	Investment
chempia	Industry-adjusted change in employees	Green et al. (2017)	Stock	Underlying	S	Investment
chinv	Change in inventory	Green et al. (2017)	Stock	Underlying	S	Investment
chmom	Change in 6-month momentum	Green et al. (2017)	Stock	Underlying	S	Past Prices
chpmia	Industry-adjusted change in profit margin	Green et al. (2017)	Stock	Underlying	S	Profitability
chtx	Change in tax expense	Green et al. (2017)	Stock	Underlying	S	Quality

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
cinvest	Corporate investment	Green et al. (2017)	Stock	Underlying	S	Investment
civpiv	Near-the-money put vs. call implied volatility	Bali and Hovakimian (2009)	Option	Underlying	O	Informed Trading
close	Close price	Eisdorfer et al. (2022)	Stock	Underlying	S	Informed Trading
convind	Convertible debt indicator	Green et al. (2017)	Stock	Underlying	S	Risk
currat	Current ratio	Green et al. (2017)	Stock	Underlying	S	Accruals
dciv	Change in atm call implied volatility	An et al. (2014)	Option	Underlying	O	Informed Trading
defrisk	Default risk	Vasquez and Xiao (2021)	Stock	Underlying	S	Risk
delta	Delta	Buchner and Kelly (2020)	Option	Contract	I	Risk
demand_pressure	Option Demand Pressure		Option	Underlying	O	Informed Trading
depr	Depreciation / PP&E	Green et al. (2017)	Stock	Underlying	S	Investment
divi	Dividend initiation	Green et al. (2017)	Stock	Underlying	S	Value
divo	Dividend omission	Green et al. (2017)	Stock	Underlying	S	Value
doi	Dollar open interest		Option	Contract	I	Illiquidity
dolvol	Dollar trading volume	Green et al. (2017)	Stock	Underlying	S	Profitability
dpiv	Change in atm put implied volatility	An et al. (2014)	Option	Underlying	O	Informed Trading
dso	Stock vs. option volume in USD	Roll et al. (2010)	Option	Underlying	O	Informed Trading
dvol	Dollar trading volume	Cao and Wei (2010)	Option	Underlying	O	Illiquidity
dy	Dividend to price	Green et al. (2017)	Stock	Underlying	S	Value
ear	Earnings announcement return	Green et al. (2017)	Stock	Underlying	S	Profitability
egr	Growth in common shareholder equity	Green et al. (2017)	Stock	Underlying	S	Investment
embedlev	Embedded Leverage	Karakaya (2014)	Option	Contract	I	Risk
ep	Earnings to price	Green et al. (2017)	Stock	Underlying	S	Value
expiration_month	Expiration month indicator		Option	Contract	I	Informed Trading
fric	Contribution of market frictions to expected returns	Hiraki and Skiadopoulos (2020)	Option	Underlying	O	Frictions
gamma	Gamma	Buchner and Kelly (2020)	Option	Contract	I	Risk
gammaps	Pastor and Stambaugh liquidity measure	Pástor and Stambaugh (2003)	Option	Bucket	B	Illiquidity
gma	Gross profitability	Green et al. (2017)	Stock	Underlying	S	Quality
grcapx	Growth in capital expenditures	Green et al. (2017)	Stock	Underlying	S	Profitability
grltnoa	Growth in long-term net operating assets	Green et al. (2017)	Stock	Underlying	S	Profitability
herf	Industry sales concentration	Green et al. (2017)	Stock	Underlying	S	Quality
hire	Employee growth rate	Green et al. (2017)	Stock	Underlying	S	Profitability
hkurt	Historic kurtosis		Option	Bucket	B	Risk

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
hskew	Historic skewness		Option	Bucket	B	Risk
hvol	Historic Volatility		Option	Bucket	B	Risk
idiovol	Idiosyncratic return volatility	Green et al. (2017)	Stock	Underlying	S	Risk
ill	Amihud Illiquidity	Green et al. (2017)	Stock	Underlying	S	Illiquidity
illiq	Illiquidity	Bao et al. (2011)	Option	Bucket	B	Illiquidity
ind_10	Industry code		Stock	Underlying	S	Industry
ind_11	Industry code		Stock	Underlying	S	Industry
ind_12	Industry code		Stock	Underlying	S	Industry
ind_13	Industry code		Stock	Underlying	S	Industry
ind_14	Industry code		Stock	Underlying	S	Industry
ind_15	Industry code		Stock	Underlying	S	Industry
ind_16	Industry code		Stock	Underlying	S	Industry
ind_17	Industry code		Stock	Underlying	S	Industry
ind_18	Industry code		Stock	Underlying	S	Industry
ind_19	Industry code		Stock	Underlying	S	Industry
ind_20	Industry code		Stock	Underlying	S	Industry
ind_21	Industry code		Stock	Underlying	S	Industry
ind_22	Industry code		Stock	Underlying	S	Industry
ind_23	Industry code		Stock	Underlying	S	Industry
ind_24	Industry code		Stock	Underlying	S	Industry
ind_25	Industry code		Stock	Underlying	S	Industry
ind_26	Industry code		Stock	Underlying	S	Industry
ind_27	Industry code		Stock	Underlying	S	Industry
ind_28	Industry code		Stock	Underlying	S	Industry
ind_29	Industry code		Stock	Underlying	S	Industry
ind_30	Industry code		Stock	Underlying	S	Industry
ind_31	Industry code		Stock	Underlying	S	Industry
ind_32	Industry code		Stock	Underlying	S	Industry
ind_33	Industry code		Stock	Underlying	S	Industry
ind_34	Industry code		Stock	Underlying	S	Industry
ind_35	Industry code		Stock	Underlying	S	Industry
ind_36	Industry code		Stock	Underlying	S	Industry
ind_37	Industry code		Stock	Underlying	S	Industry

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
ind_38	Industry code		Stock	Underlying	S	Industry
ind_39	Industry code		Stock	Underlying	S	Industry
ind_40	Industry code		Stock	Underlying	S	Industry
ind_41	Industry code		Stock	Underlying	S	Industry
ind_42	Industry code		Stock	Underlying	S	Industry
ind_43	Industry code		Stock	Underlying	S	Industry
ind_44	Industry code		Stock	Underlying	S	Industry
ind_45	Industry code		Stock	Underlying	S	Industry
ind_46	Industry code		Stock	Underlying	S	Industry
ind_47	Industry code		Stock	Underlying	S	Industry
ind_48	Industry code		Stock	Underlying	S	Industry
ind_49	Industry code		Stock	Underlying	S	Industry
ind_50	Industry code		Stock	Underlying	S	Industry
ind_51	Industry code		Stock	Underlying	S	Industry
ind_52	Industry code		Stock	Underlying	S	Industry
ind_53	Industry code		Stock	Underlying	S	Industry
ind_54	Industry code		Stock	Underlying	S	Industry
ind_55	Industry code		Stock	Underlying	S	Industry
ind_56	Industry code		Stock	Underlying	S	Industry
ind_57	Industry code		Stock	Underlying	S	Industry
ind_58	Industry code		Stock	Underlying	S	Industry
ind_59	Industry code		Stock	Underlying	S	Industry
ind_60	Industry code		Stock	Underlying	S	Industry
ind_61	Industry code		Stock	Underlying	S	Industry
ind_62	Industry code		Stock	Underlying	S	Industry
ind_63	Industry code		Stock	Underlying	S	Industry
ind_64	Industry code		Stock	Underlying	S	Industry
ind_65	Industry code		Stock	Underlying	S	Industry
ind_66	Industry code		Stock	Underlying	S	Industry
ind_67	Industry code		Stock	Underlying	S	Industry
ind_68	Industry code		Stock	Underlying	S	Industry
ind_69	Industry code		Stock	Underlying	S	Industry
ind_70	Industry code		Stock	Underlying	S	Industry

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
ind_71	Industry code		Stock	Underlying	S	Industry
ind_72	Industry code		Stock	Underlying	S	Industry
ind_73	Industry code		Stock	Underlying	S	Industry
ind_74	Industry code		Stock	Underlying	S	Industry
ind_75	Industry code		Stock	Underlying	S	Industry
ind_76	Industry code		Stock	Underlying	S	Industry
ind_77	Industry code		Stock	Underlying	S	Industry
ind_78	Industry code		Stock	Underlying	S	Industry
ind_79	Industry code		Stock	Underlying	S	Industry
ind_80	Industry code		Stock	Underlying	S	Industry
ind_81	Industry code		Stock	Underlying	S	Industry
ind_82	Industry code		Stock	Underlying	S	Industry
ind_83	Industry code		Stock	Underlying	S	Industry
ind_84	Industry code		Stock	Underlying	S	Industry
ind_85	Industry code		Stock	Underlying	S	Industry
ind_86	Industry code		Stock	Underlying	S	Industry
ind_87	Industry code		Stock	Underlying	S	Industry
ind_88	Industry code		Stock	Underlying	S	Industry
ind_89	Industry code		Stock	Underlying	S	Industry
ind_90	Industry code		Stock	Underlying	S	Industry
ind_91	Industry code		Stock	Underlying	S	Industry
ind_92	Industry code		Stock	Underlying	S	Industry
ind_93	Industry code		Stock	Underlying	S	Industry
ind_94	Industry code		Stock	Underlying	S	Industry
ind_95	Industry code		Stock	Underlying	S	Industry
ind_96	Industry code		Stock	Underlying	S	Industry
ind_97	Industry code		Stock	Underlying	S	Industry
ind_98	Industry code		Stock	Underlying	S	Industry
ind_99	Industry code		Stock	Underlying	S	Industry
indmom	Industry momentum	Green et al. (2017)	Stock	Underlying	S	Past Prices
invest	Capital expenditures and inventory	Green et al. (2017)	Stock	Underlying	S	Investment
iv	Implied volatility	Buchner and Kelly (2020)	Option	Contract	I	Contract
iv_rank	Implied volatility rank vs. last year		Option	Bucket	B	Past Prices

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
ivarud30	Option implied variance asymmetry	Huang and Li (2019)	Option	Underlying	O	Risk
ivd	Implied volatility duration	Schlag et al. (2020)	Option	Underlying	O	Risk
ivrv	Implied volatility minus realized volatility	Bali and Hovakimian (2009)	Option	Underlying	O	Risk
ivrv_ratio	Implied volatility minus realized volatility ratio		Option	Underlying	O	Risk
ivslope	Implied volatility slope	Vasquez (2017)	Option	Underlying	O	Risk
ivvol	Volatility of atm volatility	Baltussen et al. (2018)	Option	Underlying	O	Risk
ldso	Log changes in the stock to option volume	Roll et al. (2010)	Option	Underlying	O	Informed Trading
lev	Leverage	Green et al. (2017)	Stock	Underlying	S	Quality
lgr	Growth in long-term debt	Green et al. (2017)	Stock	Underlying	S	Quality
lso	Log of stock vs. option volume	Roll et al. (2010)	Option	Underlying	O	Informed Trading
m_degree	Standardized strike		Option	Contract	I	Contract
maxret	Maximum daily return	Green et al. (2017)	Stock	Underlying	S	Risk
mid	Option mid price		Option	Contract	I	Contract
modos	Modified stock vs. option volume	Johnson and So (2012)	Option	Underlying	O	Informed Trading
mom12m	12-month momentum	Green et al. (2017)	Stock	Underlying	S	Past Prices
mom1m	1-month momentum	Green et al. (2017)	Stock	Underlying	S	Past Prices
mom36m	36-month momentum	Green et al. (2017)	Stock	Underlying	S	Past Prices
mom6m	6-month momentum	Green et al. (2017)	Stock	Underlying	S	Past Prices
moneyness	Moneyness		Option	Contract	I	Contract
ms	Financial statement score	Green et al. (2017)	Stock	Underlying	S	Quality
mve	Size	Green et al. (2017)	Stock	Underlying	S	Quality
mve_ia	Industry-adjusted size	Green et al. (2017)	Stock	Underlying	S	Quality
nincr	Number of earnings increases	Green et al. (2017)	Stock	Underlying	S	Quality
nopt	Number of options trading		Option	Underlying	O	Illiquidity
ocgo	Disposition Effect	Bergsma et al. (2020)	Option	Bucket	B	Past Prices
oi	Open interest		Option	Contract	I	Illiquidity
oistock	Open interest vs. stock volume		Option	Bucket	B	Informed Trading
operprof	Operating profitability	Green et al. (2017)	Stock	Underlying	S	Quality
optspread	Option bid-ask spread		Option	Contract	I	Illiquidity
orgcap	Organization capital	Green et al. (2017)	Stock	Underlying	S	Quality
P	Put-flag		Option	Contract	I	Contract
pba	Proportional bid-ask spread	Cao and Wei (2010)	Option	Underlying	O	Illiquidity
pchcapx_ia	Industry-adjusted % change in capital expenditures	Green et al. (2017)	Stock	Underlying	S	Investment

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
pchcurrat	% change in current ratio	Green et al. (2017)	Stock	Underlying	S	Quality
pchdepr	% change in depreciation	Green et al. (2017)	Stock	Underlying	S	Quality
pchgm_pchsale	% change in gross margin - % change in sales	Green et al. (2017)	Stock	Underlying	S	Quality
pchquick	% change in quick ratio	Green et al. (2017)	Stock	Underlying	S	Quality
pchsale_pchinvt	% change in sales - % change in inventory	Green et al. (2017)	Stock	Underlying	S	Profitability
pchsale_pchrect	% change in sales - % change in A/R	Green et al. (2017)	Stock	Underlying	S	Profitability
pchsale_pchxsga	% change in sales - % change in SG&A	Green et al. (2017)	Stock	Underlying	S	Profitability
pchsaleinv	% change in sales-to-inventory	Green et al. (2017)	Stock	Underlying	S	Profitability
pcpv	Put-call parity deviations	Ofek et al. (2004)	Option	Underlying	O	Frictions
pcratio	Put-call ratio	Blau et al. (2014)	Option	Underlying	O	Informed Trading
pctacc	Percent accruals	Green et al. (2017)	Stock	Underlying	S	Accruals
pfht	Modified illiquidity measure based on zero returns	Fong et al. (2017)	Option	Bucket	B	Illiquidity
pifht	An extended FHT measured based on zero returns		Option	Bucket	B	Illiquidity
pilliq	Percentage illiquidity	Cao and Wei (2010)	Option	Underlying	O	Illiquidity
piroll	Extended Roll's measure	Goyenko et al. (2009)	Option	Bucket	B	Illiquidity
pricedelay	Price delay	Green et al. (2017)	Stock	Underlying	S	Illiquidity
ps	Financial statements score (Piotroski)	Green et al. (2017)	Stock	Underlying	S	Quality
pzeros	Illiquidity measure based on zero returns	Lesmond et al. (1999)	Option	Bucket	B	Illiquidity
quick	Quick ratio	Green et al. (2017)	Stock	Underlying	S	Quality
rd	R&D increase	Green et al. (2017)	Stock	Underlying	S	Investment
rd_mve	R&D to market capitalization	Green et al. (2017)	Stock	Underlying	S	Quality
rd_sale	R&D to sales	Green et al. (2017)	Stock	Underlying	S	Quality
realestate	Real estate holdings	Green et al. (2017)	Stock	Underlying	S	Quality
retvol	Return volatility	Green et al. (2017)	Stock	Underlying	S	Risk
rnk182	182-day risk-neutral kurtosis		Option	Underlying	O	Risk
rnk273	273-day risk-neutral kurtosis		Option	Underlying	O	Risk
rnk30	30-day risk-neutral kurtosis		Option	Underlying	O	Risk
rnk365	365-day risk-neutral kurtosis		Option	Underlying	O	Risk
rnk91	91-day risk-neutral kurtosis		Option	Underlying	O	Risk
rns182	182-day risk-neutral skewness	Borochin et al. (2020)	Option	Underlying	O	Risk
rns273	273-day risk-neutral skewness	Borochin et al. (2020)	Option	Underlying	O	Risk
rns30	30-day risk-neutral skewness	Borochin et al. (2020)	Option	Underlying	O	Risk
rns365	365-day risk-neutral skewness	Borochin et al. (2020)	Option	Underlying	O	Risk

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
rns91	91-day risk-neutral skewness	Borochin et al. (2020)	Option	Underlying	O	Risk
roaq	Return on assets	Green et al. (2017)	Stock	Underlying	S	Profitability
roavol	Earnings volatility	Green et al. (2017)	Stock	Underlying	S	Quality
roeq	Return on equity	Green et al. (2017)	Stock	Underlying	S	Profitability
roic	Return on invested capital	Green et al. (2017)	Stock	Underlying	S	Profitability
roll	Roll's measure of illiquidity	Roll (1984)	Option	Bucket	B	Illiquidity
rsup	Revenue surprise	Green et al. (2017)	Stock	Underlying	S	Profitability
rv	Realized variance	Cao et al. (2019)	Stock	Underlying	S	Risk
salecash	Sales to cash	Green et al. (2017)	Stock	Underlying	S	Value
saleinv	Sales to inventory	Green et al. (2017)	Stock	Underlying	S	Value
salerec	Sales to receivables	Green et al. (2017)	Stock	Underlying	S	Value
season1	Seasonal return - 1 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Stock	Underlying	S	Past Prices
season2	Seasonal return - 2 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Stock	Underlying	S	Past Prices
season3	Seasonal return - 3 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Stock	Underlying	S	Past Prices
season4	Seasonal return - 4 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Stock	Underlying	S	Past Prices
secured	Secured debt	Green et al. (2017)	Stock	Underlying	S	Quality
securedind	Secured debt indicator	Green et al. (2017)	Stock	Underlying	S	Quality
sgr	Sales growth	Green et al. (2017)	Stock	Underlying	S	Investment
shrtfee	Implied shorting fees	Muravyev et al. (2021)	Option	Underlying	O	Frictions
sin	Sin stocks	Green et al. (2017)	Stock	Underlying	S	Quality
skewiv	IV skew	Xing et al. (2010)	Option	Underlying	O	Informed Trading
so	Stock vs. option volume	Roll et al. (2010)	Option	Underlying	O	Informed Trading
sp	Sales to price	Green et al. (2017)	Stock	Underlying	S	Quality
std_dolvol	Volatility of liquidity (dollar trading volume)	Green et al. (2017)	Stock	Underlying	S	Illiquidity
std_turn	Volatility of liquidity (share turnover)	Green et al. (2017)	Stock	Underlying	S	Illiquidity
stdacc	Accrual volatility	Green et al. (2017)	Stock	Underlying	S	Accruals
stdamihud	Standard deviation of Amihud's illiquidity measure		Option	Bucket	B	Illiquidity
stdcf	Cash flow volatility	Green et al. (2017)	Stock	Underlying	S	Risk
tang	Dept capacity/firm tangibility	Green et al. (2017)	Stock	Underlying	S	Quality

Continued on Next Page

Table IA6.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Information Set	Group
tb	Tax income to book income	Green et al. (2017)	Stock	Underlying	S	Past Prices
theta	Theta	Buchner and Kelly (2020)	Option	Contract	I	Risk
tlm30	Tail loss measure	Vilkov and Xiao (2012)	Option	Underlying	O	Risk
toi	Total option open interest		Option	Underlying	O	Illiquidity
ttm	Time-to-maturity		Option	Contract	I	Contract
turn	Share turnover	Green et al. (2017)	Stock	Underlying	S	Risk
turnover	Option turnover		Option	Bucket	B	Illiquidity
underlying_return	Return of the underlying		Stock	Underlying	S	Frictions
vega	Vega	Buchner and Kelly (2020)	Option	Contract	I	Risk
vol	Trading volume in options		Option	Underlying	O	Illiquidity
volga	Volga	Buchner and Kelly (2020)	Option	Contract	I	Risk
volunc	Volatility uncertainty	Cao et al. (2019)	Option	Underlying	O	Risk
vs_change	Change in weighted put-call spread	Cremers and Weinbaum (2010)	Option	Underlying	O	Informed Trading
vs_level	Weighted put-call spread	Cremers and Weinbaum (2010)	Option	Underlying	O	Informed Trading
zerotrade	Zero trading days	Green et al. (2017)	Stock	Underlying	S	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 reference, 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), the information set (Information Set) and to which characteristic group (Group) it belongs to.

Appendix IA7. Additional Summary Statistics

IA7.1. 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	1705.95	139.84	1531.0	1619.25	1704.5	1807.5	1886.6
Stock coverage of stock universe (EW)	1705.95	139.84	1531.0	1619.25	1704.5	1807.5	1886.6
Stock coverage of stock universe (VW)	76.29	5.49	69.11	71.44	76.59	81.29	83.68
Stock traded at NYSE or AMEX	51.04	2.46	48.12	49.84	51.21	52.63	53.67
Stock already included in previous month	83.38	6.23	78.56	80.88	83.67	86.76	88.58
Panel B: Time-Series Average of Cross-Sectional Distributions							
Firm size in million	7367	25376	232	531	1453	4508	14409
Firm size CSRP percentile	71	18	44	58	74	87	93
Firm volatility CSRP percentile	45	25	11	24	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	4.43%	4.94%	Telecom	3.48%	2.56%
Consumer durables	0.52%	0.48%	Utilities	2.68%	2.26%
Manufacturing	9.23%	8.26%	Wholesale	11.01%	8.63%
Energy	4.85%	2.77%	Healthcare	11.59%	9.14%
Chemicals	2.4%	1.65%	Finance	9.93%	23.91%
Business Equipment	20.42%	14.43%	Other	19.48%	20.96%

Table IA7.1: 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.

IA7.2. Delta-Hedged Option Return per Bucket

Table IA7.2: Delta-Hedged Option Return per Bucket

	Mean	Sd	10-Pctl	Q1	Q2	Q3	90-Pctl	Skew	Kurt	JB
Panel A: Long Term (N=6,933,006)										
Delta-Hedged Return	0.39	797.48	-4.27	-1.88	-0.29	1.46	4.75	1.17	10.48	0.0
Days to Maturity	269.25	185.62	112.0	141.0	200.0	324.0	571.0	1.52	1.35	
Moneyness	1.05	0.44	0.72	0.86	1.0	1.16	1.38	2.56	15.3	
Implied volatility	46.07	23.48	23.56	29.91	40.22	55.86	76.05	1.63	3.84	
Absolute Delta	0.45	0.24	0.13	0.25	0.43	0.63	0.79	0.21	-0.92	
Panel B: Long Term Atm (N=5,633,672)										
Delta-Hedged Return	0.49	884.66	-4.03	-1.8	-0.27	1.5	4.69	1.05	10.78	0.0
Days to Maturity	273.38	190.14	113.0	141.0	201.0	326.0	597.0	1.5	1.23	
Moneyness	1.04	0.24	0.8	0.9	1.01	1.14	1.29	1.49	5.57	
Implied volatility	45.96	23.22	23.41	29.74	40.17	55.99	76.11	1.54	3.31	
Absolute Delta	0.47	0.19	0.22	0.31	0.46	0.62	0.75	0.2	-0.97	
Panel C: Long Term Itm Call (N=224,363)										
Delta-Hedged Return	-0.0	29.74	-1.97	-0.89	-0.18	0.53	1.68	1.02	31.22	0.0
Days to Maturity	245.15	163.81	110.0	138.0	173.0	295.0	506.0	1.59	1.81	
Moneyness	0.66	0.14	0.47	0.58	0.68	0.77	0.82	-0.62	-0.02	
Implied volatility	49.19	20.76	28.2	34.87	44.72	58.57	75.72	1.41	2.99	
Absolute Delta	0.91	0.04	0.86	0.88	0.91	0.93	0.96	-0.1	-0.25	
Panel D: Long Term Itm Put (N=159,969)										
Delta-Hedged Return	-0.12	1.49	-1.18	-0.56	-0.14	0.28	1.0	0.17	12.76	0.0
Days to Maturity	231.01	148.96	110.0	137.0	172.0	266.0	472.0	1.7	2.45	
Moneyness	1.98	1.86	1.24	1.35	1.57	2.01	2.86	6.39	61.77	
Implied volatility	57.94	39.75	22.49	30.71	46.11	72.72	108.29	1.89	4.69	
Absolute Delta	0.87	0.1	0.76	0.84	0.89	0.94	0.96	-2.01	6.01	
Panel E: Long Term Otm Call (N=341,037)										
Delta-Hedged Return	0.33	9.91	-7.91	-3.7	-0.46	3.41	9.41	0.95	5.75	0.0
Days to Maturity	249.26	159.35	112.0	141.0	198.0	296.0	505.0	1.53	1.67	
Moneyness	1.62	0.59	1.22	1.3	1.45	1.72	2.18	3.42	17.94	
Implied volatility	41.6	22.07	20.95	26.32	35.54	50.27	70.6	1.63	3.37	
Absolute Delta	0.13	0.06	0.06	0.09	0.12	0.16	0.2	1.67	9.18	

Continued on Next Page

Table IA7.2 from previous page

	Mean	Sd	10-Pctl	Q1	Q2	Q3	90-Pctl	Skew	Kurt	JB
Panel F: Long Term Otm Put (N=573,965)										
Delta-Hedged Return	-0.27	5.76	-5.63	-3.07	-0.91	1.51	5.31	1.72	7.4	0.0
Days to Maturity	260.64	168.7	112.0	141.0	201.0	324.0	536.0	1.46	1.34	
Moneyness	0.66	0.14	0.48	0.58	0.68	0.76	0.82	-0.73	0.36	
Implied volatility	45.28	20.3	26.25	31.66	40.29	53.07	70.51	1.76	4.55	
Absolute Delta	0.09	0.04	0.03	0.05	0.09	0.12	0.14	0.21	-0.56	
Panel G: Short Term (N=5,203,395)										
Delta-Hedged Return	-0.26	7.77	-4.69	-2.11	-0.47	0.99	4.03	1.51	11.66	0.0
Days to Maturity	44.48	23.05	17.0	21.0	49.0	52.0	80.0	0.26	-1.16	
Moneyness	1.01	0.24	0.84	0.92	1.0	1.07	1.17	2.29	15.22	
Implied volatility	50.33	29.19	23.35	30.73	42.79	61.49	86.4	1.92	5.86	
Absolute Delta	0.48	0.26	0.14	0.26	0.47	0.69	0.84	0.1	-1.1	
Panel H: Short Term Atm (N=3,972,640)										
Delta-Hedged Return	-0.2	5.36	-4.52	-2.12	-0.49	1.14	4.16	1.35	9.73	0.0
Days to Maturity	44.46	22.92	17.0	21.0	49.0	52.0	80.0	0.25	-1.14	
Moneyness	1.01	0.1	0.9	0.95	1.0	1.06	1.12	0.84	3.14	
Implied volatility	49.32	27.46	23.01	30.36	42.4	60.75	84.46	1.65	4.06	
Absolute Delta	0.49	0.19	0.23	0.32	0.48	0.65	0.76	0.1	-1.13	
Panel I: Short Term Itm Call (N=308,902)										
Delta-Hedged Return	0.16	8.4	-1.6	-0.77	-0.23	0.22	1.13	9.61	110.18	0.0
Days to Maturity	41.73	23.23	17.0	19.0	47.0	52.0	79.0	0.43	-1.13	
Moneyness	0.82	0.1	0.69	0.77	0.84	0.89	0.92	-1.3	2.18	
Implied volatility	53.8	29.74	26.23	33.93	46.28	65.06	90.36	1.93	5.92	
Absolute Delta	0.89	0.04	0.84	0.86	0.89	0.93	0.95	-0.13	0.16	
Panel J: Short Term Itm Put (N=240,799)										
Delta-Hedged Return	-0.54	3.46	-1.62	-0.8	-0.32	0.1	0.77	-12.3	205.47	0.0
Days to Maturity	41.31	23.07	17.0	19.0	46.0	52.0	79.0	0.48	-1.09	
Moneyness	1.4	0.79	1.09	1.14	1.23	1.41	1.76	5.95	54.28	
Implied volatility	66.32	49.42	24.11	33.79	50.62	81.57	128.13	2.0	4.69	
Absolute Delta	0.89	0.07	0.81	0.85	0.89	0.94	0.96	-1.35	4.07	

Continued on Next Page

Table IA7.2 from previous page

	Mean	Sd	10-Pctl	Q1	Q2	Q3	90-Pctl	Skew	Kurt	JB
Panel K: Short Term Otm Call (N=255,247)										
Delta-Hedged Return	-0.47	24.96	-8.75	-4.36	-0.85	2.72	8.44	0.85	5.42	0.0
Days to Maturity	47.35	23.42	17.0	22.0	50.0	77.0	80.0	0.11	-1.25	
Moneyness	1.27	0.24	1.09	1.13	1.2	1.32	1.51	2.76	11.49	
Implied volatility	48.39	29.28	22.05	28.63	39.78	58.87	87.27	1.78	4.16	
Absolute Delta	0.12	0.06	0.05	0.08	0.12	0.15	0.17	1.95	16.3	
Panel L: Short Term Otm Put (N=425,807)										
Delta-Hedged Return	-0.91	6.05	-6.86	-3.63	-1.24	0.95	4.62	1.46	6.98	0.0
Days to Maturity	46.75	23.34	17.0	21.0	50.0	74.0	80.0	0.12	-1.22	
Moneyness	0.82	0.1	0.68	0.77	0.84	0.89	0.92	-1.25	2.0	
Implied volatility	49.26	26.01	25.33	31.83	42.45	58.73	82.15	1.81	4.62	
Absolute Delta	0.1	0.04	0.04	0.06	0.1	0.13	0.15	-0.01	-0.74	

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 5. Specifically, we separately consider predictability for short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (Itm: $m^{stand} > 1$ for puts, $m^{stand} < 1$ for calls), out-of-the-money (Otm: $m^{stand} < 1$ for puts, $m^{stand} > 1$ for calls) calls and puts, and at-the-money options (Atm: $-1 \leq m^{stand} \leq 1$), as well as time-to-maturity and moneyness combinations. The moneyness buckets are based on standardized moneyness, i.e., $m^{stand} = \log \frac{K}{S} / (\sigma^{atm} \sqrt{\tau})$, where σ^{atm} is the at-the-money implied volatility for time to maturity τ. Skew denotes skewness. Kurt denotes excess kurtosis. JB denotes the p-value in percent of testing if delta-hedged option returns follow a normal distribution via the Jarque-Bera test. Each panel shows statistics for pooled options belonging to one bucket.

IA7.3. Number of Options per Bucket

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Long Term	15.4	26.36	1.1	2.49	6.05	17.16	40.9
Long Term Atm	12.55	18.83	1.08	2.36	5.65	15.02	32.98
Long Term Itm Call	2.3	3.09	1.0	1.0	1.14	2.39	4.51
Long Term Itm Put	2.33	2.61	1.0	1.0	1.15	2.56	4.87
Long Term Otm Call	3.03	4.3	1.0	1.0	1.7	3.37	6.24
Long Term Otm Put	4.24	6.38	1.0	1.0	2.1	4.88	9.8
Short Term	11.53	16.51	1.03	2.57	6.17	14.15	27.91
Short Term Atm	8.83	10.41	1.01	2.45	5.47	11.43	20.58
Short Term Itm Call	2.19	2.64	1.0	1.0	1.21	2.44	4.26
Short Term Itm Put	2.22	2.52	1.0	1.0	1.11	2.37	4.49
Short Term Otm Call	2.35	3.14	1.0	1.0	1.34	2.51	4.48
Short Term Otm Put	3.08	4.27	1.0	1.0	1.63	3.5	6.41

Table IA7.3: 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 5. 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: $m^{stand} > 1$ for puts, $m^{stand} < 1$ for calls), out-of-the-money (Otm: $m^{stand} < 1$ for puts, $m^{stand} > 1$ for calls) calls and puts, and at-the-money options (Atm: $-1 \leq m^{stand} \leq 1$), as well as time-to-maturity and moneyness combinations. The moneyness buckets are based on standardized moneyness, i.e., $m^{stand} = \log \frac{K}{S}/(\sigma^{atm}\sqrt{\tau})$, where σ^{atm} is the at-the-money implied volatility for time to maturity τ. 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.

Appendix IA8. Model Comparison

Additional Analyses

IA8.1. Cross-sectional Diebold and Mariano (1995) Tests

Panel A: Diebold and Mariano (1995) Cross-Sectional Forecast Comparison

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	1.57	1.73	−0.51	**-4.28**	**3.13**	**8.17**	**6.07**	**5.37**	**10.49**	**9.02**
Lasso		1.03	**-2.46**	**-5.30**	**4.28**	**4.63**	**4.28**	**3.09**	**5.65**	**6.10**
ENet			**-3.12**	**-5.26**	**4.14**	**4.50**	**4.19**	**2.92**	**5.18**	**5.98**
PCR				**-2.78**	**5.71**	**5.77**	**6.38**	**4.05**	**5.80**	**7.39**
PLS					**6.63**	**10.43**	**9.40**	**8.00**	**11.80**	**11.35**
L-En						**3.46**	**3.16**	**2.19**	**3.89**	**5.58**
GBR							**-2.89**	−0.18	−0.68	**4.09**
RF								1.16	1.69	**7.89**
Dart									−0.21	**2.04**
FFN										**3.73**

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).

IA8.2. Cross-sectional Comparison between N-En and L-En

N-En beats L-En in 86.1% of the months when tasked with predicting future cross-sectional return spreads for delta-hedged single-equity options:

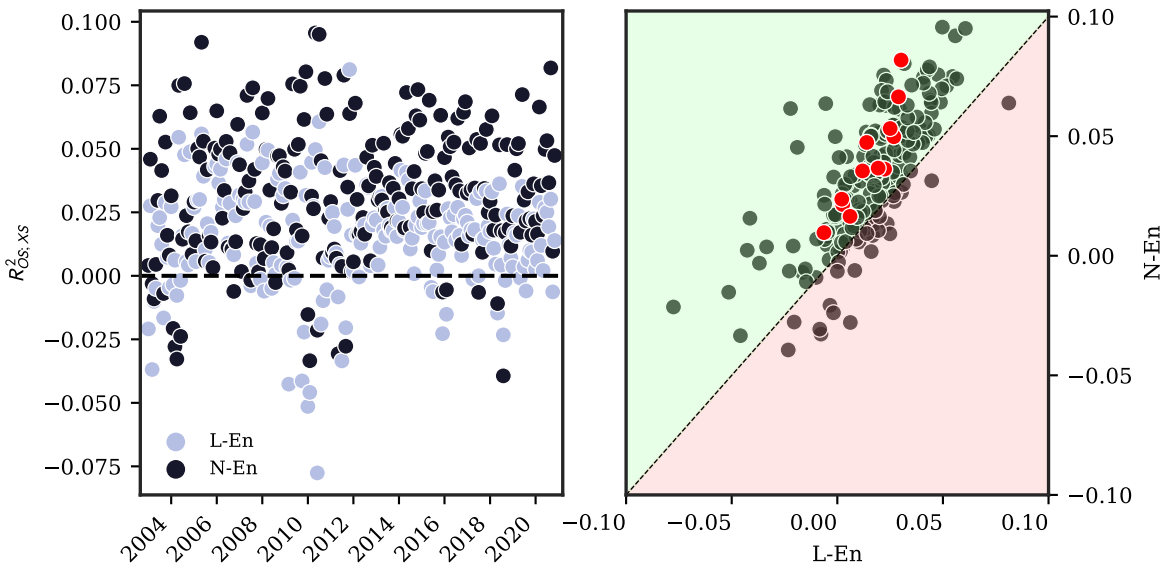

Fig. IA8.1. Comparing Linear and Nonlinear Ensembles – $R^2_{OS;XS}$

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 red circles represent the Coronavirus selloff and subsequent recovery from December 2019 through December 2020.

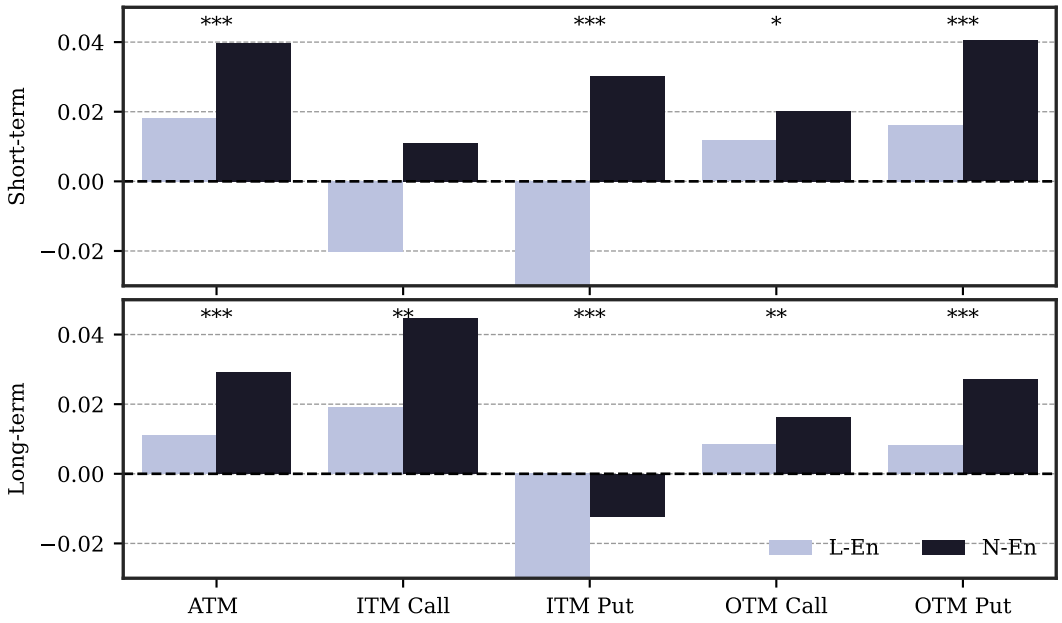

Fig. IA8.2. Comparing Linear and Nonlinear Ensembles – $R^2_{OS;XS}$

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 5. The plots are cut at −0.03 for better visibility. Predicting in-the-money put returns is a difficult task for the linear ensemble.

Appendix IA9. Consistency of Expected Returns for N-En

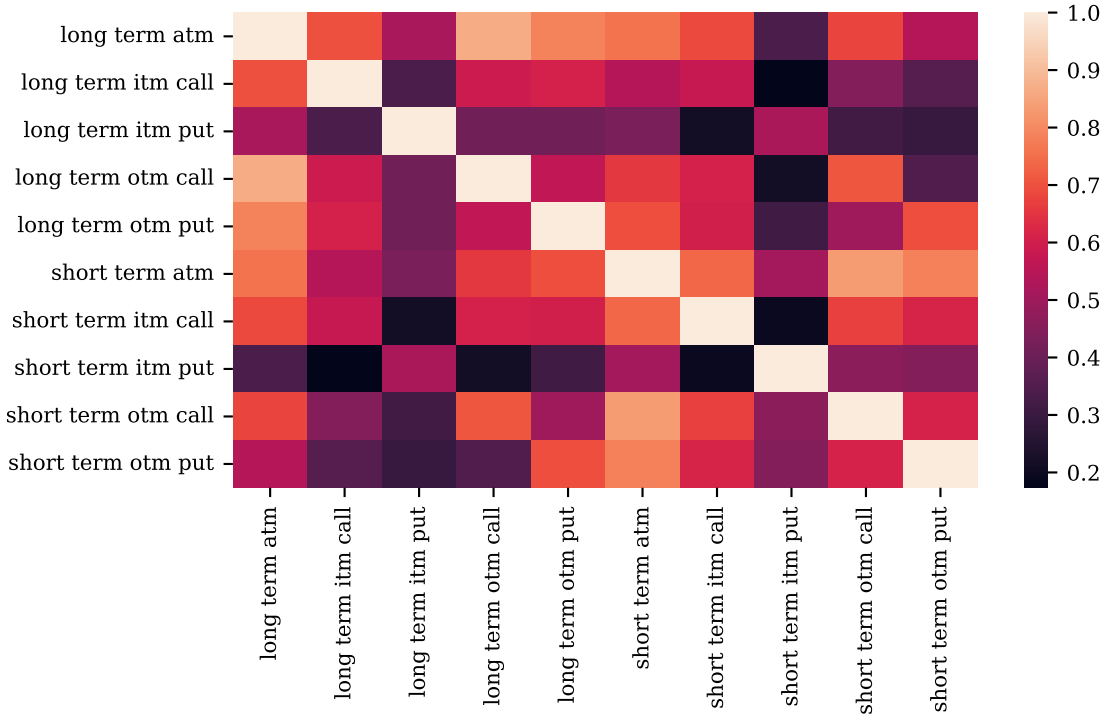

Fig. IA9.1. Average Correlation of Expected Returns by Option Bucket

The figure shows the average correlation of expected returns by option buckets. For each underlying in the sample, we each moneyness-maturity bucket we calculate the sample correlation of expected returns, requiring at least 10 valid monthly observations per bucket. Here, we plot the resulting average across all underlyings.

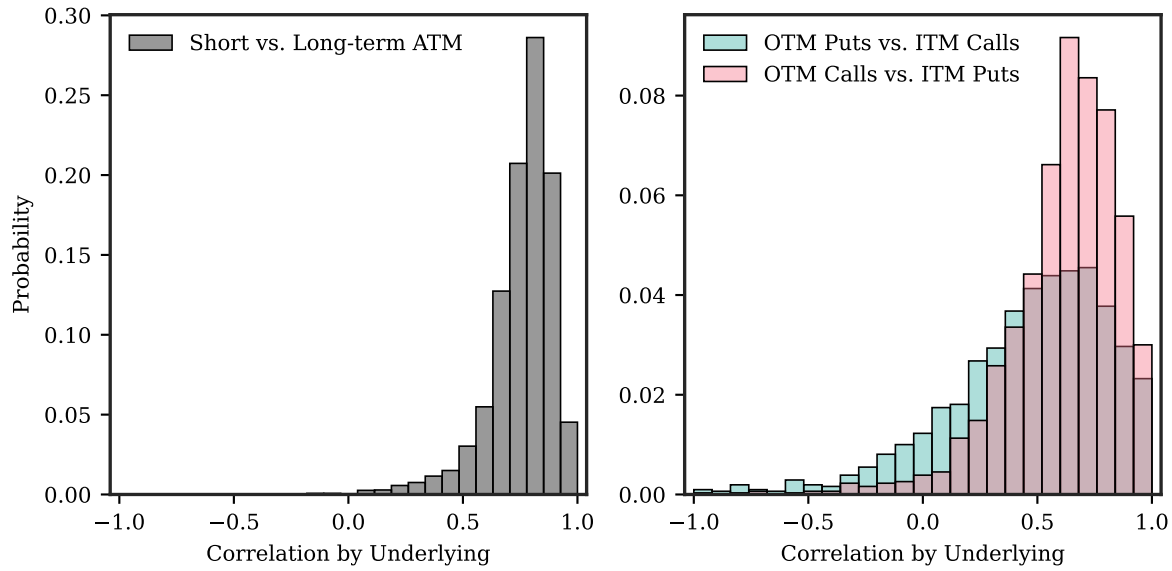

Fig. IA9.2. Histogram of the Correlation of Expected Returns by Option Bucket

The figure shows histogram of how the return predictions by the nonlinear ensemble N-En of options in various buckets, but on the same underlying, correlate with one another. The left panel shows the correlations for short and long-term at-the-money options, the right compares expected returns for in-the-money vs. out-of-the-money short-term options.

Appendix IA10. Machine Learning Portfolios

Additional Analyses

IA10.1. Performance for ML Portfolios using Equally-weighted Returns

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.351	−1.087	1.398	−0.778	−1.730	−1.649	1.942	−0.849	***
2	−0.775	−0.528	1.535	−0.344	−0.781	−0.700	1.529	−0.458	
3	−0.542	−0.365	1.434	−0.255	−0.460	−0.415	1.368	−0.303	
4	−0.369	−0.259	1.466	−0.177	−0.280	−0.266	1.255	−0.212	
5	−0.224	−0.196	1.497	−0.131	−0.155	−0.166	1.295	−0.128	
6	−0.092	−0.122	1.509	−0.081	−0.050	−0.119	1.325	−0.090	
7	0.038	−0.061	1.494	−0.041	0.052	−0.075	1.425	−0.053	
8	0.174	−0.027	1.510	−0.018	0.166	−0.031	1.491	−0.021	
9	0.337	0.046	1.436	0.032	0.324	0.090	1.555	0.058	
Hi	0.637	0.216	1.485	0.146	0.791	0.391	1.835	0.213	
H-L	1.988	1.303	1.270	1.026	2.521	2.040	1.598	1.277	***
		(13.27)		(8.95)		(13.27)		(8.83)	
call	1.864	1.400	1.614	0.867	2.596	2.290	1.941	1.180	***
put	1.943	1.232	1.274	0.967	2.264	1.971	1.663	1.185	***

Table IA10.1: 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.

IA10.2. Trading Strategy Performance Over Time

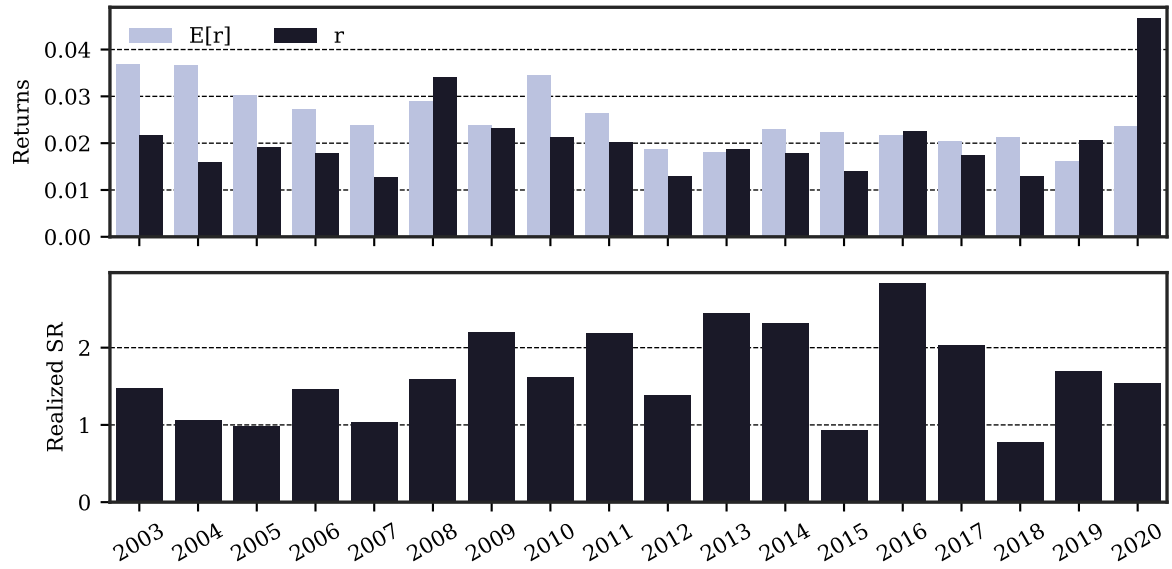

Fig. IA10.1. Trading Strategy Performance Over Time

The figure shows the year-by-year performance of the trading strategy following predictions by the nonlinear ensemble N-En. The upper panel shows expected returns in light blue and realized returns in dark blue. The lower panel depicts the monthly realized Sharpe ratios for each year.

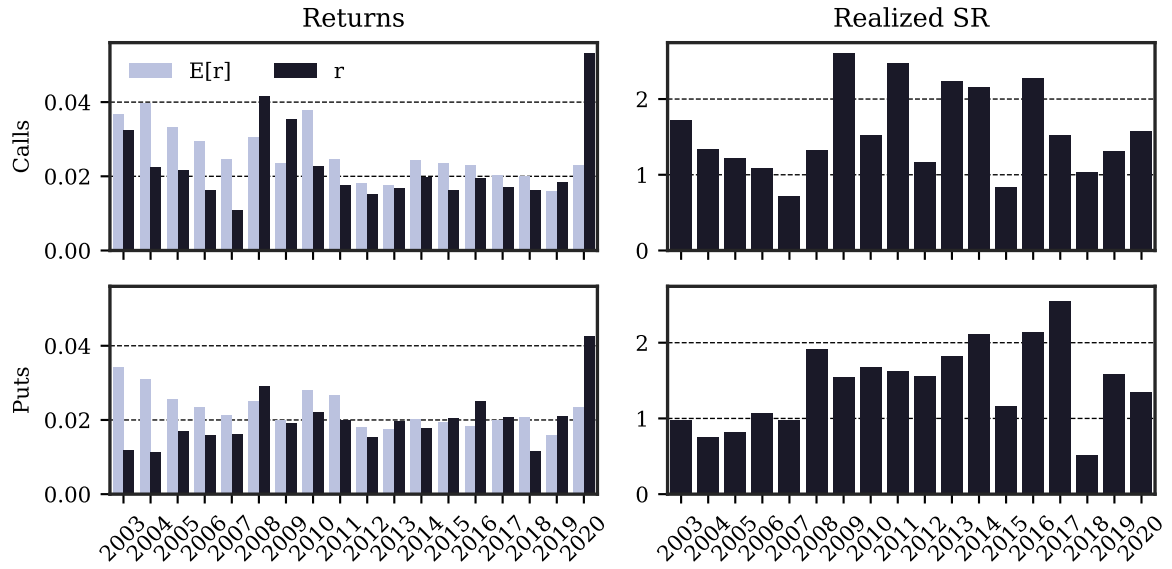

Fig. IA10.2. Trading Strategy Performance Over Time – Put and Call Options

The figure shows the year-by-year performance of the trading strategy following predictions by the nonlinear ensemble N-En. The left panels show expected returns in light blue and realized returns in dark blue. The right panels depict the monthly realized Sharpe ratios for each year. The upper panel shows results for all call options, the lower for all put options.

IA10.3. Performance for ML Portfolios Conditional on Earnings Announcements and News Days

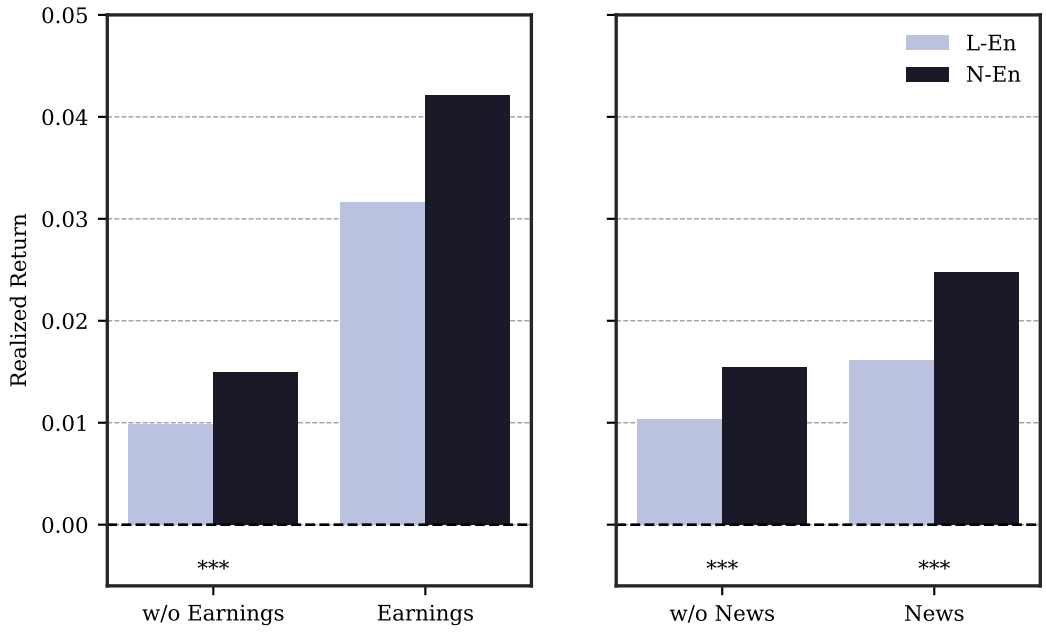

Fig. IA10.3. The Impact of Earnings Announcements and News on the High-Minus-Low Spreads

The figure shows returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods conditional on the occurrence on earnings announcements and news during the investment period. w/o Earnings (w/o News) denotes results when trading is implemented on options whose underlying stocks experience no earnings announcements (news) during the investment period. Earnings (News) denotes results when trading is implemented on options whose underlying stocks experience earnings announcements (news announcements) during the investment period. The analysis is restricted to short-term at-the-money options with weekly investment horizons. News occurrences are identified using the Dow Jones version of Ravenpack News Analytics. News are only recorded if the relevance score is 100 and if they are highly positive (sentiment score above 0.75) or hihgly negative (sentiment score below 0.25). All values are given per month. ***, **, * correspond to N-En outperforming L-En significantly at the 1%, 5%, 10% level, respectively.

IA10.4. Summary Statistics for ML Portfolios – Put and Call Composition

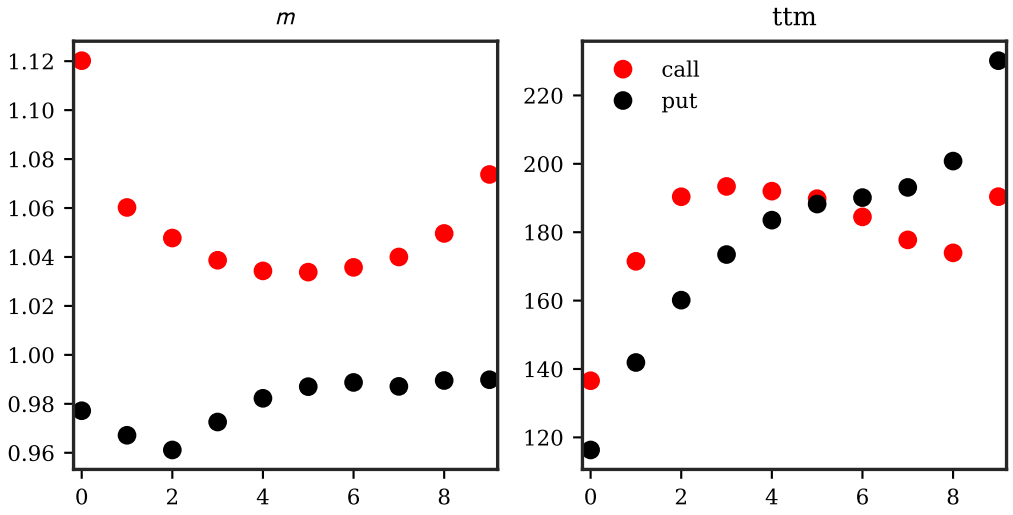

Fig. IA10.4. Machine Learning Portfolios – Moneyness and Days-to-maturity

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.

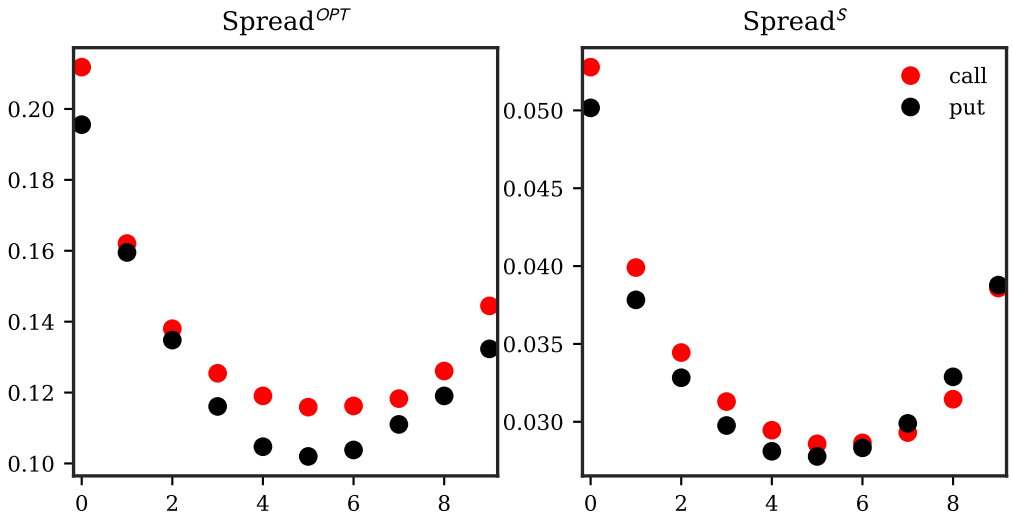

Fig. IA10.5. Machine Learning Portfolios – Spreads of the Options and the Underlyings

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.

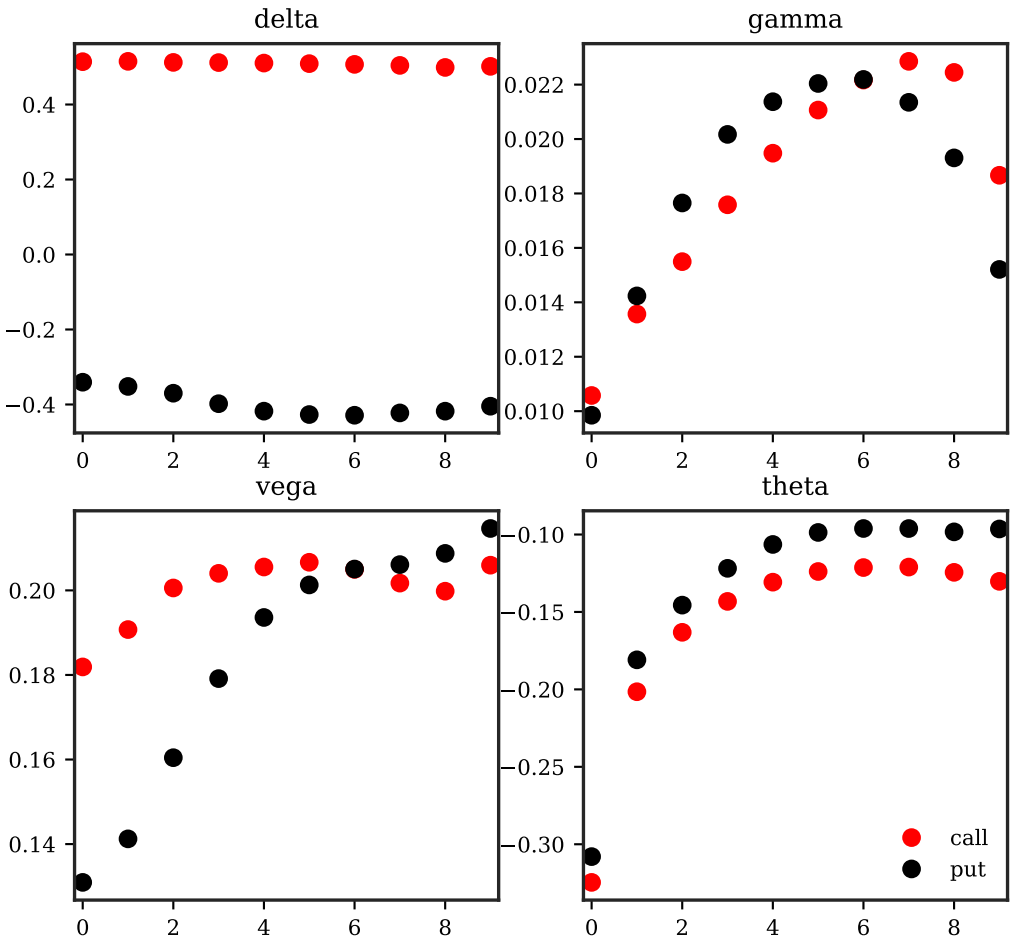

Fig. IA10.6. Machine Learning Portfolios – Greeks

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. IA10.7. 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.

IA10.5. Machine Learning Portfolios – Underlying's Concentration in Decile Portfolios

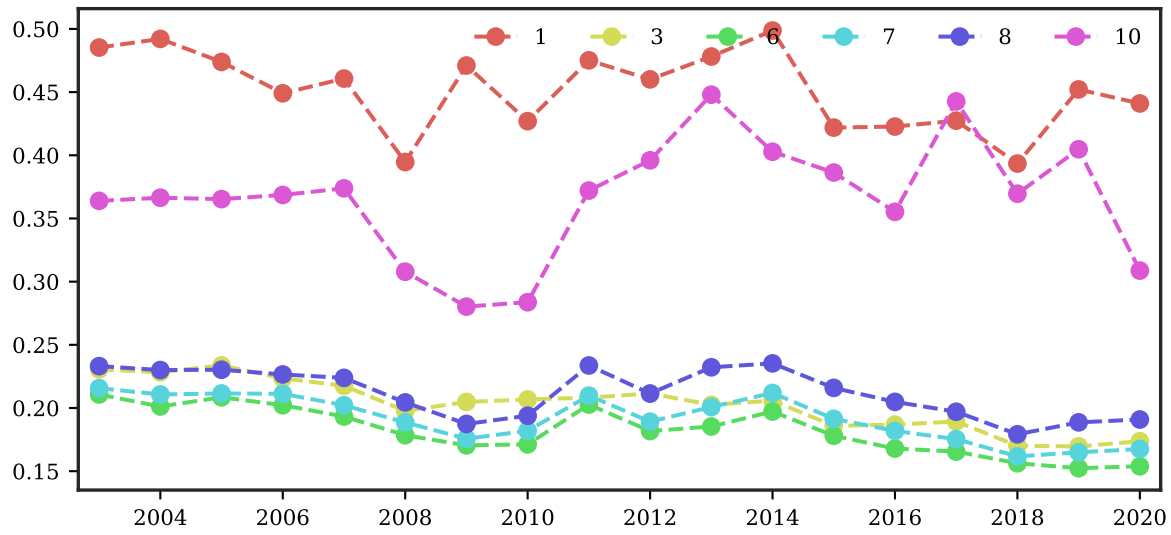

Fig. IA10.8. Machine Learning Portfolio – Underlying-Option Concentration

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.

IA10.6. Machine Learning Portfolios – Return Differentials Within and Across Underlyings

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.509	−1.413	2.098	−0.674	−2.016	−1.710	2.185	−0.783	
2	−1.007	−0.838	1.941	−0.432	−1.040	−1.021	2.008	−0.509	
3	−0.743	−0.549	1.805	−0.304	−0.600	−0.402	1.777	−0.226	
4	−0.552	−0.394	1.629	−0.242	−0.350	−0.276	1.521	−0.182	
5	−0.376	−0.350	1.526	−0.229	−0.204	−0.232	1.350	−0.172	
6	−0.239	−0.073	1.646	−0.044	−0.086	−0.123	1.328	−0.093	
7	−0.103	−0.162	1.432	−0.113	0.013	−0.050	1.405	−0.035	
8	0.027	−0.065	1.346	−0.048	0.112	−0.132	1.313	−0.101	
9	0.194	−0.026	1.321	−0.020	0.250	−0.039	1.385	−0.028	
Hi	0.440	0.084	1.323	0.064	0.574	0.223	1.699	0.131	
H-L	1.949	1.498	1.863	0.804	2.590	1.933	1.888	1.024	**
		(11.07)		(6.92)		(10.76)		(8.75)	
call	1.711	1.546	2.200	0.703	2.248	1.593	2.229	0.715	
put	1.744	1.223	1.815	0.674	1.677	1.373	1.951	0.704	

Table IA10.2: Trading on Machine Learning Predictions – Restriction to All Options Per Underlying

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method, when restricting the portfolio construction to include *all* options on a respective underlying. For this, we assign underlyings into portfolios by the average expected return on all options trading on it. 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.

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−0.921	−0.656	1.001	−0.655	−0.799	−0.812	1.227	−0.662	
2	−0.540	−0.450	1.286	−0.350	−0.476	−0.582	1.297	−0.449	
3	−0.370	−0.379	1.344	−0.282	−0.336	−0.451	1.419	−0.318	
4	−0.251	−0.306	1.411	−0.217	−0.238	−0.357	1.392	−0.256	
5	−0.156	−0.244	1.493	−0.164	−0.159	−0.263	1.432	−0.184	
6	−0.056	−0.153	1.508	−0.102	−0.069	−0.179	1.462	−0.122	
7	0.032	−0.118	1.552	−0.076	−0.001	−0.083	1.531	−0.054	
8	0.105	−0.047	1.657	−0.028	0.057	−0.038	1.611	−0.024	
9	0.191	0.030	1.792	0.017	0.122	0.091	1.730	0.053	
Hi	0.318	0.228	1.828	0.125	0.376	0.208	1.336	0.155	
H-L	1.239	0.885	1.273	0.695	1.175	1.020	0.877	1.163	
		(6.06)		(4.59)		(11.27)		(8.39)	
call	1.864	1.400	1.614	0.867	2.596	2.290	1.941	1.180	***
put	1.943	1.232	1.274	0.967	2.264	1.971	1.663	1.185	***

Table IA10.3: Trading on Machine Learning Predictions – Splitting All Options Per Underlying Into Deciles

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method, when restricting the portfolio construction to include all options on a respective underlying. For this, we assign options into decile portfolios for each underlying. Subsequently, the final decile portfolios are obtained by averaging over decile portfolios on the underlying-level. Weighting is done by the 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.

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.932	−1.264	1.608	−0.786	−3.373	−4.374	3.080	−1.420	***
2	−1.399	−1.180	1.388	−0.850	−2.011	−2.392	2.400	−0.997	***
3	−1.168	−1.002	1.313	−0.763	−1.437	−1.532	2.397	−0.639	***
4	−0.981	−0.808	1.176	−0.688	−1.082	−1.220	1.968	−0.620	***
5	−0.788	−0.643	1.355	−0.474	−0.802	−1.011	1.397	−0.724	***
6	−0.593	−0.428	1.741	−0.246	−0.545	−0.533	1.350	−0.395	
7	−0.323	−0.323	1.679	−0.192	−0.264	−0.258	1.243	−0.208	
8	0.007	0.046	1.504	0.031	0.089	−0.065	1.555	−0.042	
9	0.396	0.396	1.756	0.225	0.498	0.363	1.350	0.268	
Hi	1.019	0.744	1.978	0.376	1.428	0.723	1.704	0.424	
H-L	2.951	2.008	2.483	0.808	4.801	5.096	3.109	1.639	***
		(7.96)		(6.73)		(16.19)		(9.20)	
call	2.641	2.713	2.816	0.964	4.472	4.640	3.571	1.300	***
put	2.608	1.548	1.782	0.869	3.717	4.301	3.043	1.413	***

Table IA10.4: Trading on Machine Learning Predictions – Restriction to One Option Per Underlying

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method, when restricting the portfolio construction to include *only one* option per underlying and decile portfolio. 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.

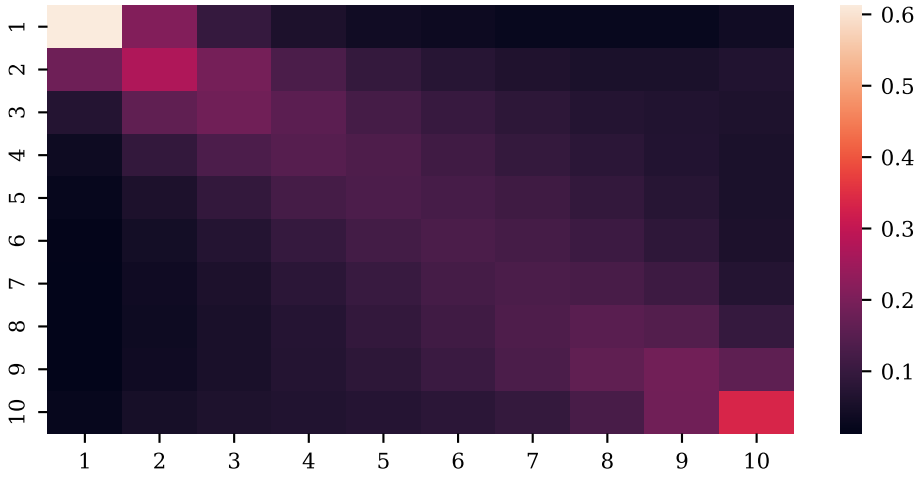

Fig. IA10.9. ML Portfolio Transition Matrix by Option Bucket

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 5.3.

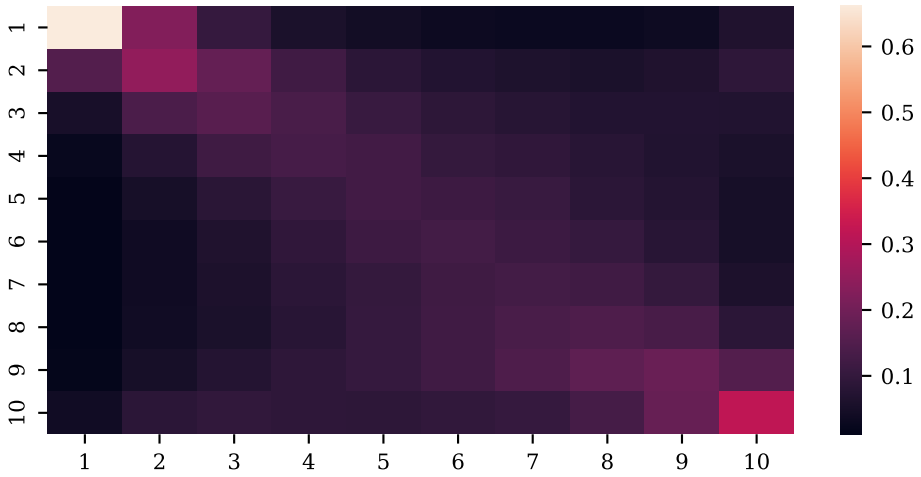

Fig. IA10.10. ML Portfolio Transition Matrix by Underlying

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.

IA10.8. 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	1.711	1.152	1.001	1.151	2.369	1.760	1.271	1.385	***
		(13.66)		(5.27)		(15.75)		(6.33)	
High VIX	2.268	1.455	1.483	0.981	2.674	2.323	1.834	1.266	***
		(8.87)		(6.98)		(9.69)		(7.15)	
Low EPU	2.081	1.222	1.109	1.102	2.661	1.839	1.295	1.420	***
		(13.44)		(6.06)		(14.78)		(6.03)	
High EPU	1.895	1.384	1.415	0.978	2.380	2.243	1.838	1.220	***
		(8.92)		(7.32)		(8.62)		(7.64)	
Neg. CFNAI	1.956	1.233	1.218	1.012	2.432	1.971	1.467	1.344	***
		(9.84)		(7.16)		(10.47)		(6.96)	
Pos. CFNAI	2.024	1.382	1.328	1.041	2.621	2.118	1.738	1.218	***
		(11.32)		(6.18)		(9.86)		(6.20)	
Low FED Stress	1.876	1.218	1.169	1.042	2.468	1.893	1.524	1.242	***
		(14.31)		(6.39)		(11.91)		(6.53)	
High FED Stress	2.235	1.490	1.462	1.019	2.637	2.366	1.718	1.377	***
		(7.93)		(7.65)		(9.39)		(5.48)	
Low SENT	2.108	1.195	1.072	1.114	2.579	1.893	1.378	1.373	***
		(13.13)		(6.29)		(16.80)		(6.67)	
High SENT	1.965	1.405	1.358	1.035	2.617	1.869	1.343	1.392	
		(8.28)		(5.29)		(9.32)		(5.19)	

Table IA10.5: 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.

IA10.9. Risk Attribution (N-En)

			CAPM		FF6		FF6+PS	
Full Sample			2.067	(12.47)	2.028	(15.54)	2.025	(14.49)
					Buckets			
$\tau \leq 90$	atm		2.326	(12.00)	2.273	(14.70)	2.269	(14.06)
	itm	C	1.017	(13.61)	0.992	(14.56)	0.992	(14.43)
	itm	P	1.334	(7.83)	1.315	(8.48)	1.315	(8.53)
	otm	C	3.382	(7.63)	3.295	(8.89)	3.291	(8.82)
	otm	P	4.832	(14.66)	4.749	(14.13)	4.752	(14.15)
$\tau > 90$	atm		2.302	(11.75)	2.256	(13.63)	2.252	(12.99)
	itm	C	1.213	(10.33)	1.152	(10.55)	1.150	(10.29)
	itm	P	0.820	(6.04)	0.834	(6.42)	0.833	(6.24)
	otm	C	2.971	(6.07)	2.813	(6.58)	2.804	(6.25)
	otm	P	2.964	(9.51)	2.968	(9.76)	2.962	(9.45)

			AN		BM		LBC	
Full Sample			1.972	(13.74)	1.792	(20.53)	1.932	(13.96)
					Buckets			
$\tau \leq 90$	atm		2.244	(12.98)	2.206	(11.54)	2.224	(12.73)
	itm	C	0.980	(13.08)	0.958	(14.76)	0.996	(14.71)
	itm	P	1.298	(8.32)	1.232	(6.24)	1.207	(7.69)
	otm	C	3.157	(7.82)	3.265	(7.48)	3.130	(6.98)
	otm	P	5.022	(13.57)	4.508	(15.90)	4.891	(14.42)
$\tau > 90$	atm		2.190	(12.15)	1.965	(15.63)	2.168	(12.29)
	itm	C	1.126	(10.30)	1.082	(8.69)	1.187	(10.28)
	itm	P	0.810	(6.38)	0.811	(4.38)	0.784	(5.55)
	otm	C	2.997	(4.90)	2.843	(7.82)	2.854	(5.29)
	otm	P	2.985	(8.15)	2.810	(10.92)	3.122	(9.93)

Table IA10.6: 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 et al. (2022), 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. (2022), LBC. We also provide the information for option buckets defined in Section 5.

Appendix IA11. Which Characteristics Matter?

Additional Analyses

IA11.1. Feature Group Importance Per Bucket (N-En)

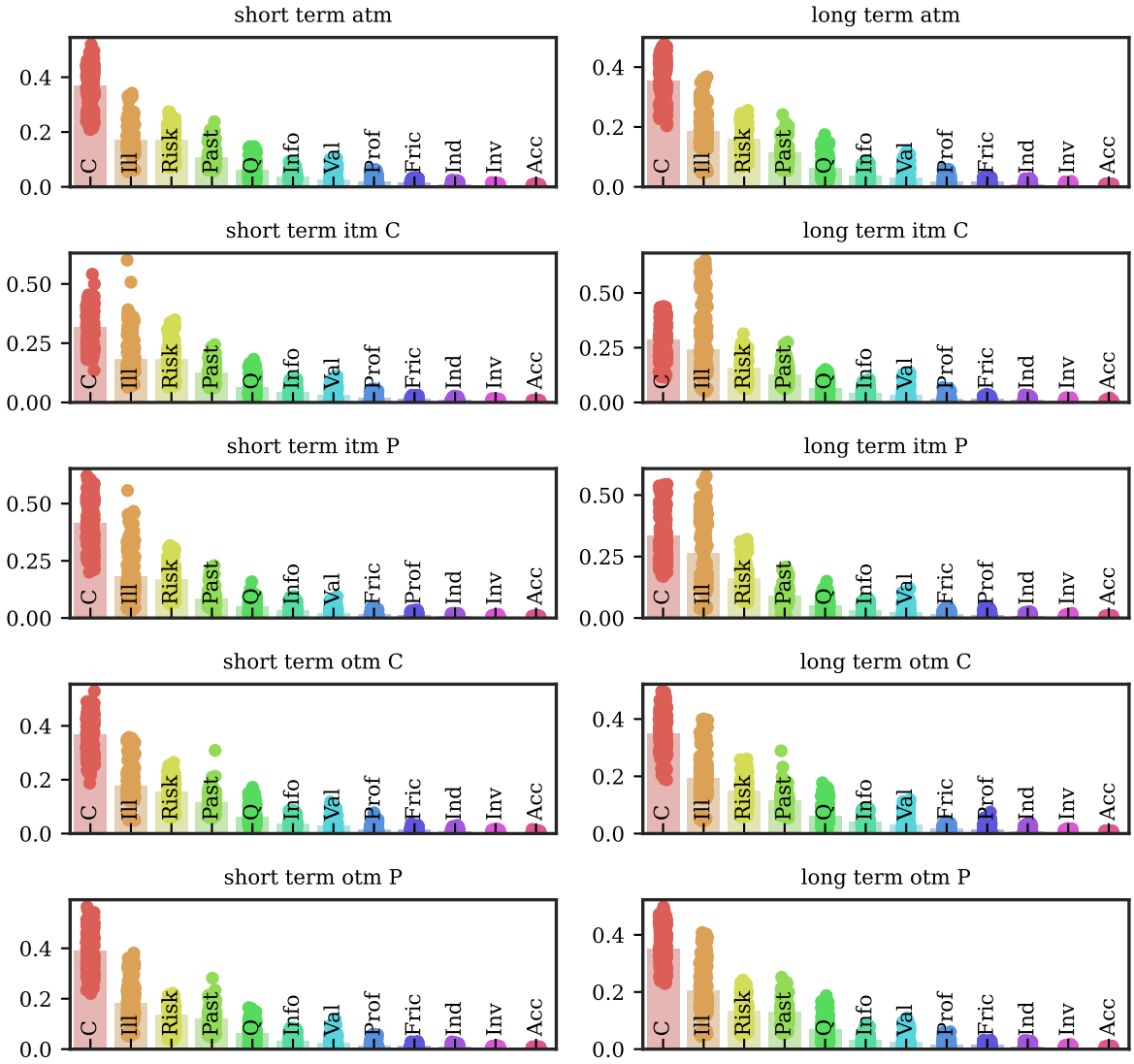

Fig. IA11.1. Feature Group Importance for N-En Per Bucket

The figure shows the feature group importance for the twelve feature groups defined in Appendix IA6 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.

IA11.2. Feature Group Importance Over Time (N-En)

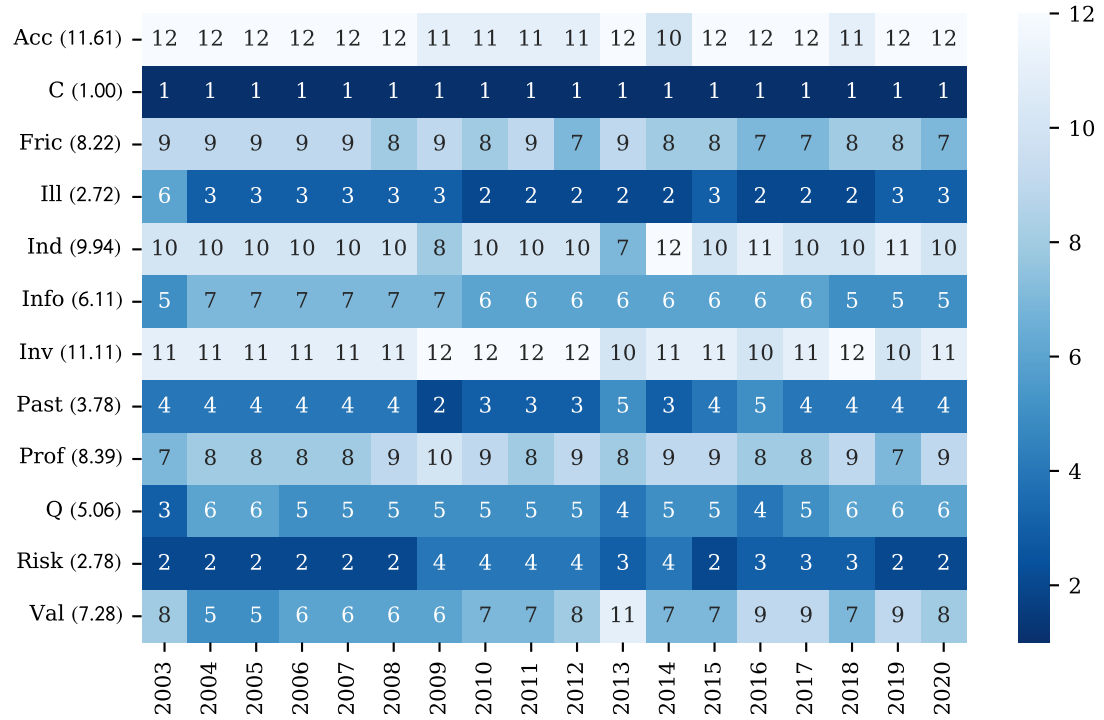

Fig. IA11.2. Feature Group Importance for N-En Over Time

The figure shows the time evolution of ranking the twelve feature groups defined in Appendix IA6 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.

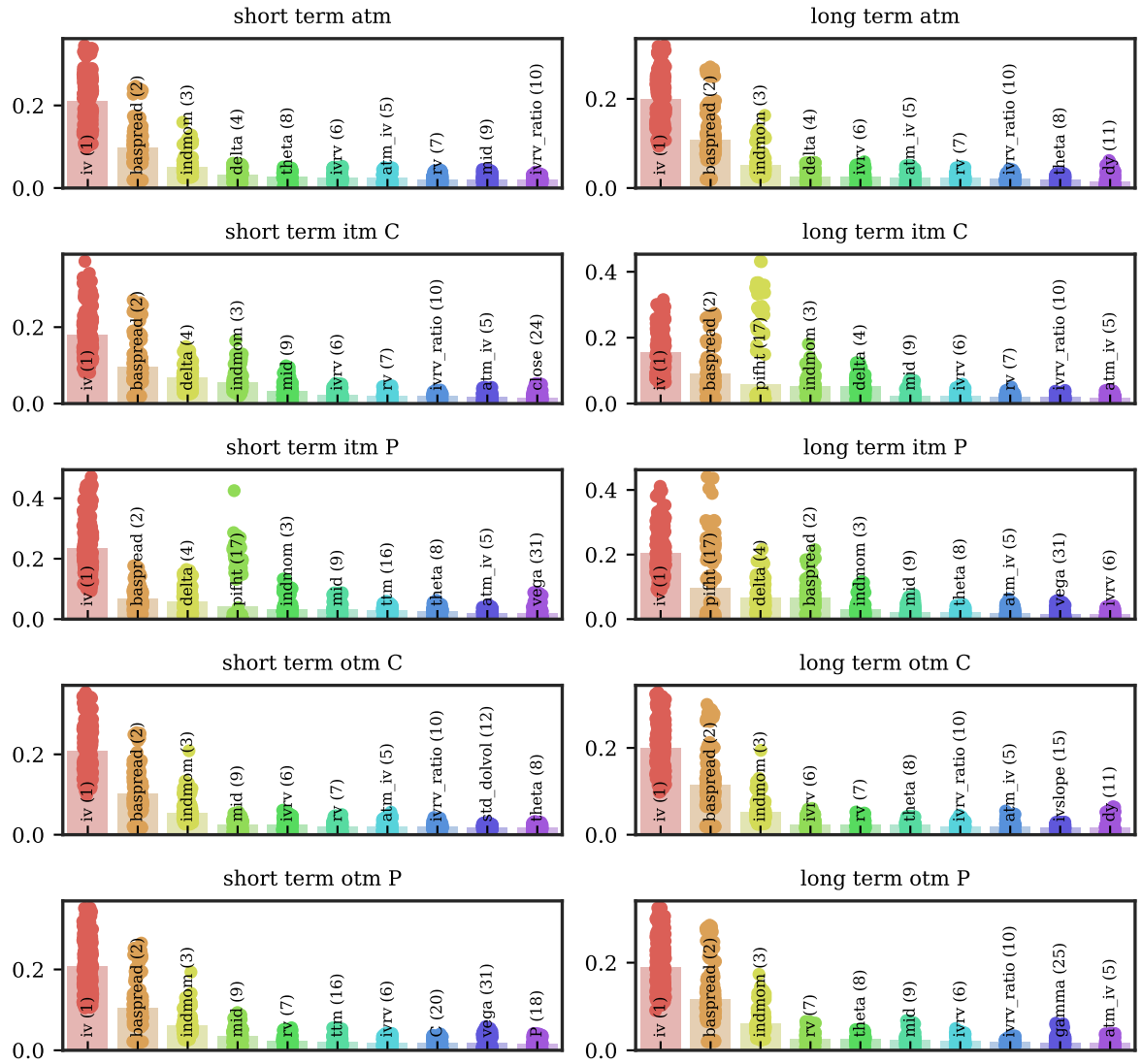

Fig. IA11.3. Importance of Most Influential Features for N-En Per Bucket

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 273 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.

IA11.4. Impact of Most Important Features Over Time (N-En)

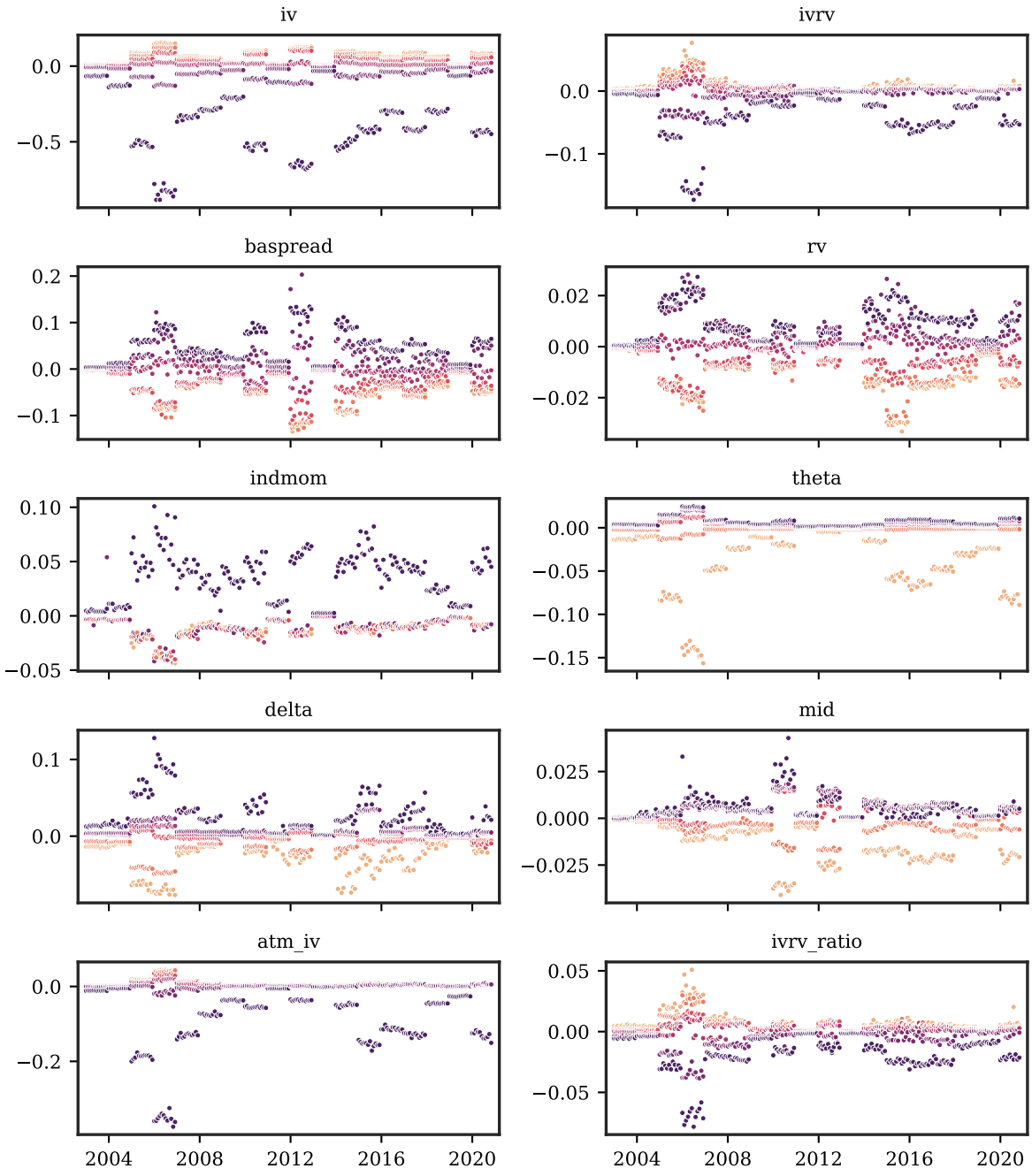

Fig. IA11.4. Impact of Most Important Features Over Time

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.

IA11.5. Functional Form of Impact of Most Important Features (N-En)

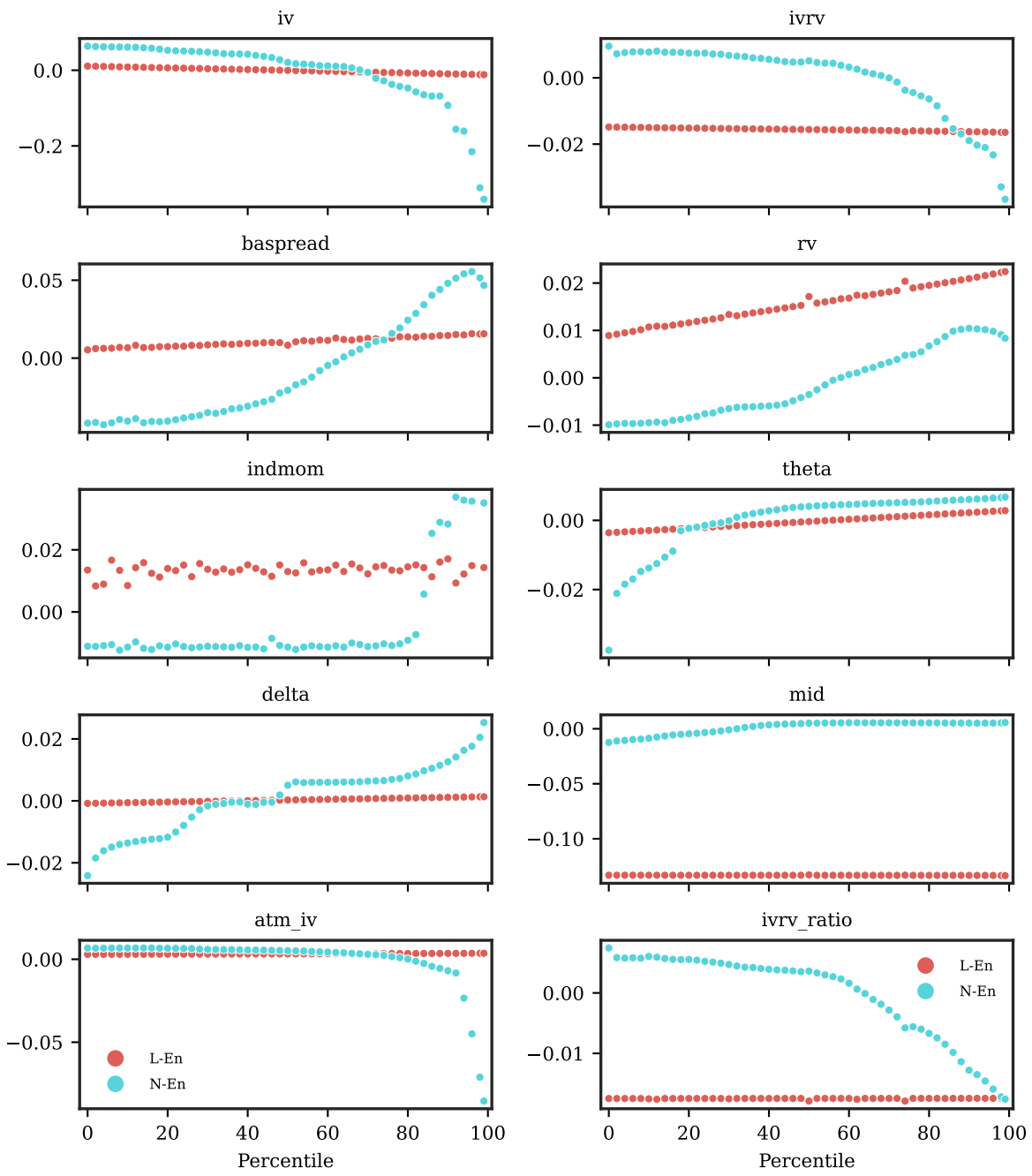

Fig. IA11.5. Functional Form of Impact of Most Important Features

The figure shows the functional form of the impact of the ten most influential features on the predictions of the linear (L-En) and nonlinear ensemble (N-En). We measure the impact using SHAP values following Lundberg and Lee (2017).

IA11.6. Importance of Volatility and Jump Risk (N-En)

			Volatility risk				Jump risk				
			ivvol	volunc	vega	ivrv	tlm30	skewiv	rns30	rnk30	gamma
	all Sample		72	151	31	6	28	145	98	101	25
						Buckets					
	atm		−1	−5	5	1	−3	0	0	0	−3
	itm	C	7	−1	−1	0	4	1	−1	0	6
$\tau \leq 90$	itm	P	2	6	−21	5	16	5	7	−4	−2
	otm	C	1	−2	−9	−1	−4	−13	−3	2	1
	otm	P	−6	−5	−24	4	−1	−12	−34	−10	−8
	atm		−1	0	9	−1	0	−1	0	0	8
	itm	C	10	5	−19	0	7	3	5	−1	−2
$\tau > 90$	itm	P	5	6	−22	4	17	5	1	12	2
	otm	C	0	1	13	−2	−3	−7	0	2	−6
	otm	P	−5	−3	−5	1	1	−11	−15	−11	−15

Table IA11.1: 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).

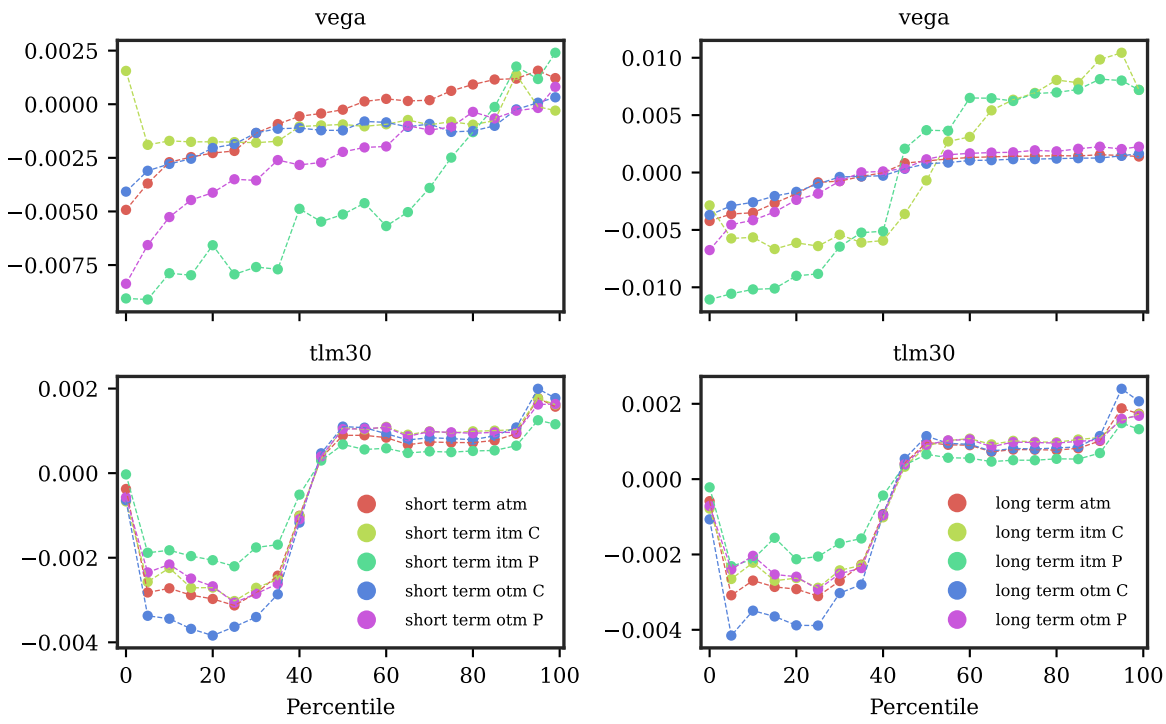

Fig. IA11.6. Relative Impact of Volatility and Jump Risk Premia Per Bucket

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).

Appendix IA12. Impact of the Information Set

Additional Analyses

IA12.1. Restricting the Information Set (N-En)

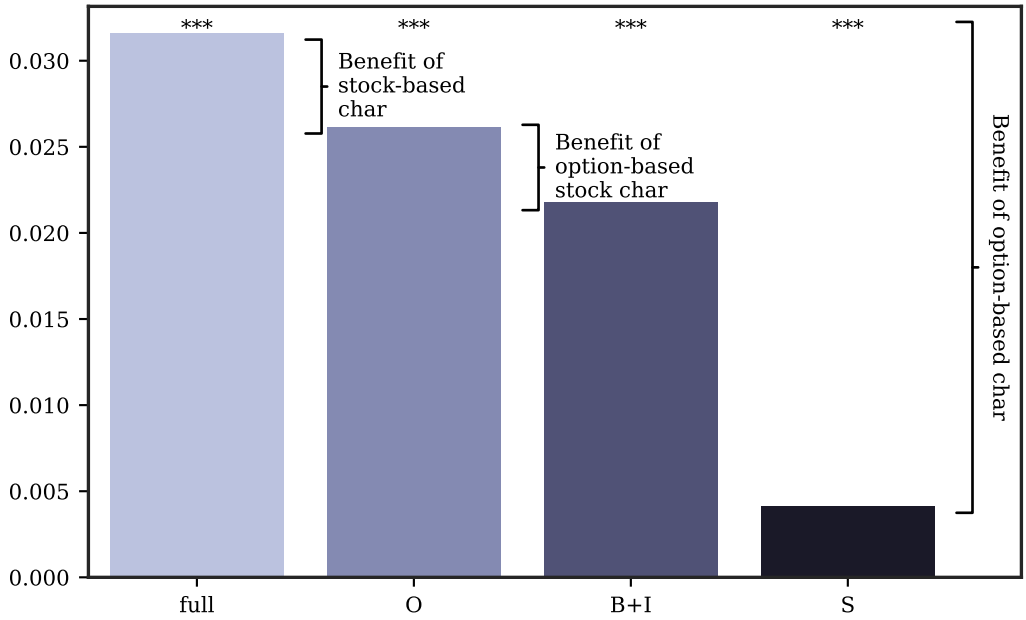

Fig. IA12.1. Restricting the Information Set for N-En – $R^2_{OS;XS}$

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 individual contract-based information (I+C) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA6. ***, **, * 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.

IA12.2. Option-bucket Performance for Different Information Sets

			L-En				N-En			
TTM	Mon.		B+I	O	S	full	B+I	O	S	full
	atm		0.008*	0.010**	0.004	0.013***	0.027***	0.031***	0.009***	0.033***
	itm	C	0.005	-0.003	-0.002	-0.012	-0.095	-0.094	0.006	0.006
$\tau \leq 90$	itm	P	-0.001	0.015	0.025	-0.036	0.004	0.012	0.041***	0.014
	otm	C	0.007	0.008	0.002	0.008*	0.016***	0.016***	0.003	0.019***
	otm	P	0.026***	0.023***	0.004	0.025***	0.046***	0.039***	0.003	0.041***
	atm		-0.007	-0.001	-0.009	0.003	0.010	0.016***	-0.004	0.021***
	itm	C	0.010	0.006	-0.010	0.012	-0.003	-0.012	-0.002	0.008
$\tau > 90$	itm	P	-0.091	-0.092	-0.045	-0.118	0.001	-0.008	-0.028	-0.042
	otm	C	-0.007	-0.004	-0.005	-0.003	-0.003	0.002	-0.003	0.006
	otm	P	0.001	0.002	-0.005	0.006	0.017*	0.013*	-0.001	0.021***

Table IA12.1: 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 5. The full model is compared with models using all option-based information (O), models using only bucket- and individual contract-based information (B+I) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA6. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

Appendix IA13. Alternating the Estimation Window

In this subsection, we investigate the robustness of our results regarding the training scheme of the machine learning models. In our paper, we follow Gu et al. (2020) and use an expanding training window, refitting the model once a year in January, and increasing the size of the training sample by one year after each iteration. Instead, we now consider a rolling-window estimation approach, with a training sample of a fixed size of ten years, as well as a training scheme that explicitly excludes the two years of the great financial crisis (2008 and 2009) to understand how important this information is for the overall efficacy of the models.

IA13.1. Fixed-Length Training Window

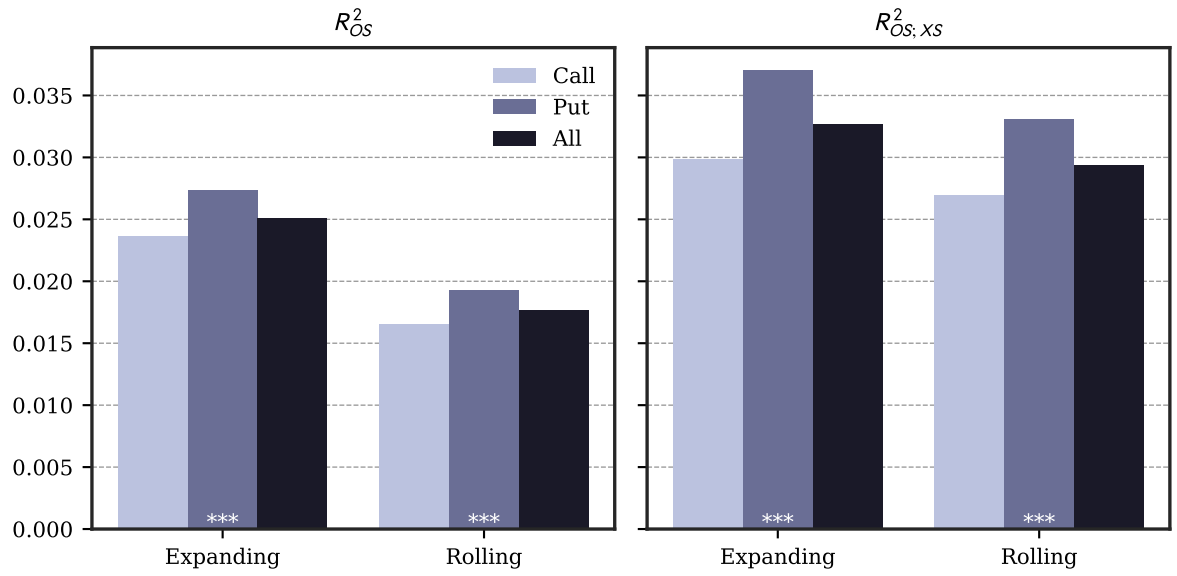

Fig. IA13.1. Predictability using a 10-year Fixed-Length Training Window

The figure shows the impact on predictability measures R^2_{OS} in the left and $R^2_{OS;XS}$ in the right plot, when estimating the N-En model using a rolling training window of a fixed length of 10 years. We separately show the testing-sample predictability for all options and only calls and puts.

	Expanding				Rolling				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	Exp. vs. Roll
Lo	−1.589	−1.678	2.021	−0.830	−1.503	−1.502	1.865	−0.805	
2	−0.713	−0.673	1.651	−0.408	−0.714	−0.575	1.679	−0.343	
3	−0.435	−0.378	1.497	−0.252	−0.469	−0.360	1.535	−0.234	
4	−0.280	−0.218	1.366	−0.159	−0.331	−0.212	1.492	−0.142	
5	−0.173	−0.115	1.412	−0.081	−0.234	−0.094	1.471	−0.064	
6	−0.082	−0.044	1.451	−0.030	−0.151	−0.092	1.405	−0.065	
7	0.010	−0.008	1.569	−0.005	−0.070	−0.010	1.494	−0.007	
8	0.114	0.028	1.654	0.017	0.020	0.049	1.544	0.031	
9	0.261	0.157	1.718	0.091	0.144	0.125	1.631	0.077	
Hi	0.709	0.477	2.060	0.231	0.544	0.480	2.023	0.237	
H-L	2.298	2.155	1.634	1.319	2.047	1.982	1.397	1.419	
		(10.57)		(7.32)		(13.71)		(7.07)	
call	2.331	2.370	2.000	1.185	2.030	2.089	1.861	1.123	
put	2.091	2.176	1.654	1.316	1.937	2.015	1.432	1.408	

Table IA13.1: Trading on Machine Learning Predictions – 10-year Fixed-Length Training Window

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method, when estimating the models using a rolling training window of a fixed length of 10 years. 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.

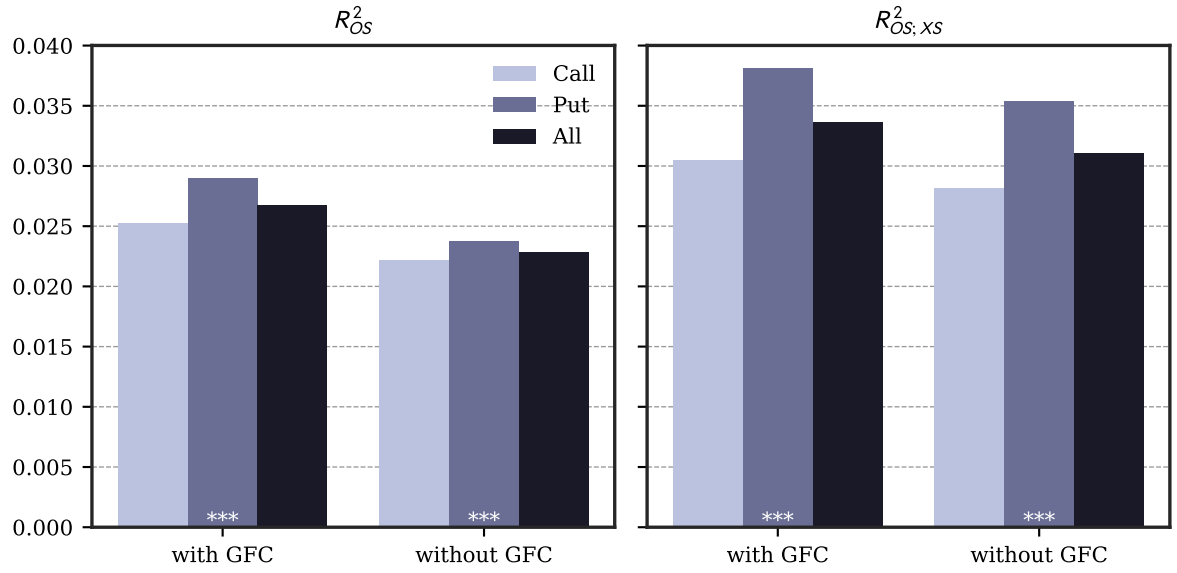

Fig. IA13.2. Predictability Including vs. Excluding the Impact of the 2008/2009 Financial Crisis

The figure shows the impact on predictability measures R^2_{OS} in the left and $R^2_{OS;XS}$ in the right plot including the information on option returns obtained during the financial crisis in the training and validation set ("with GFC") versus excluding it ("without GFC") for the nonlinear ensemble (N-En). We separately show the testing-sample predictability for all options and only calls and puts.

	with GFC				without GFC				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	GFC
Lo	−1.466	−1.655	1.984	−0.834	−1.603	−1.571	1.834	−0.857	
2	−0.641	−0.668	1.359	−0.491	−0.786	−0.673	1.516	−0.444	
3	−0.399	−0.341	1.242	−0.274	−0.524	−0.310	1.208	−0.256	
4	−0.271	−0.177	1.153	−0.153	−0.374	−0.166	1.226	−0.135	
5	−0.183	−0.074	1.214	−0.061	−0.270	−0.062	1.228	−0.050	
6	−0.108	−0.007	1.255	−0.005	−0.181	−0.039	1.284	−0.031	
7	−0.034	0.040	1.373	0.029	−0.098	0.009	1.377	0.007	
8	0.052	0.054	1.447	0.037	−0.006	0.092	1.384	0.067	
9	0.178	0.132	1.403	0.094	0.114	0.082	1.482	0.056	
Hi	0.588	0.360	1.771	0.203	0.466	0.320	1.709	0.187	
H-L	2.054	2.015	1.679	1.200	2.069	1.891	1.499	1.262	
		(7.60)		(5.99)		(7.47)		(6.11)	
call	2.071	2.115	1.923	1.100	2.051	2.070	1.878	1.102	
put	1.914	2.139	1.779	1.203	1.910	1.949	1.725	1.130	

Table IA13.2: Trading on Machine Learning Predictions – The Impact of the Financial Crisis

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method including the information on option returns obtained during the financial crisis in the training and validation set ("with GFC") versus excluding it ("without GFC"). 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 (GFC) gives the significance of comparing the mean realized returns for the two model types. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

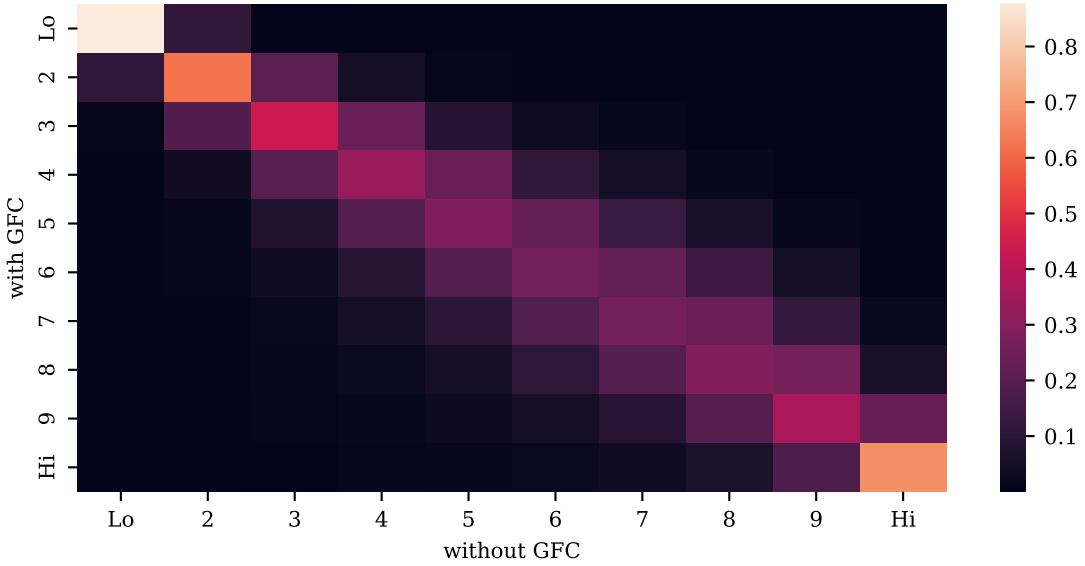

Fig. IA13.3. Expected Return Portfolio Migration Including vs. Excluding the Impact of the 2008/2009 Financial Crisis

The figure shows changes in the portfolio assignment for the models including the impact of the financial crisis in the training and validation set ("with GFC") versus excluding it ("without GFC"). The rows are normalized to one.

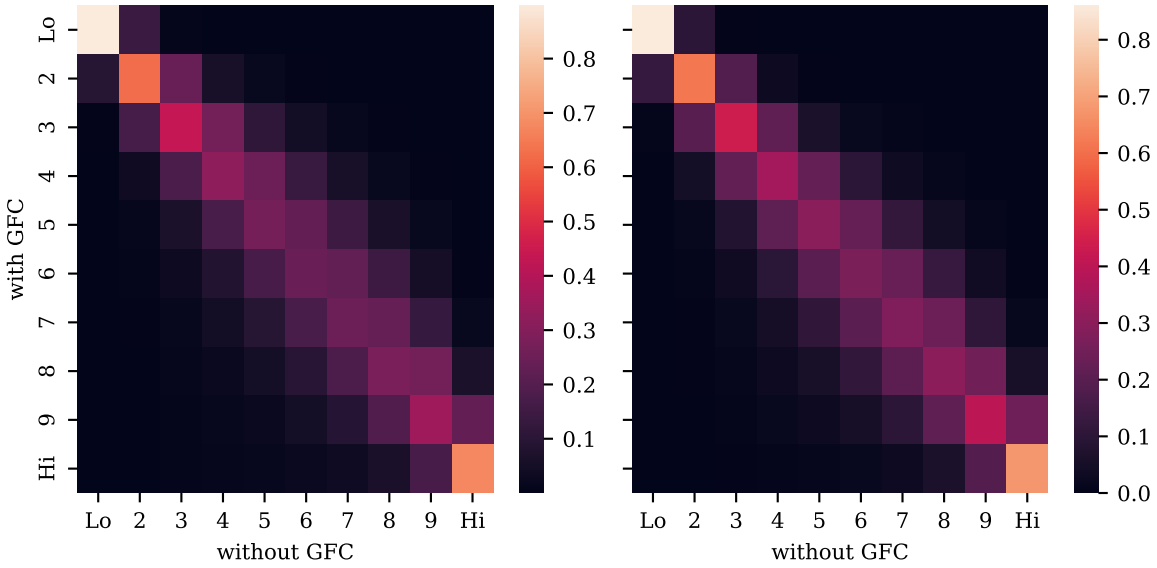

Fig. IA13.4. Expected Return Portfolio Migration Including vs. Excluding the Impact of the 2008/2009 Financial Crisis – Puts vs. Calls

The figure shows changes in the portfolio assignment for the models including the impact of the financial crisis in the training and validation set ("with GFC") versus excluding it ("without GFC"), separately for put and call options. The rows are normalized to one.

Appendix IA14. Predictability for Options on the 500 Largest CRSP Stocks

In this section we investigate return predictability of the most liquid options, by focusing on those written on the 500 largest stocks, as measured by the CRSP universe.

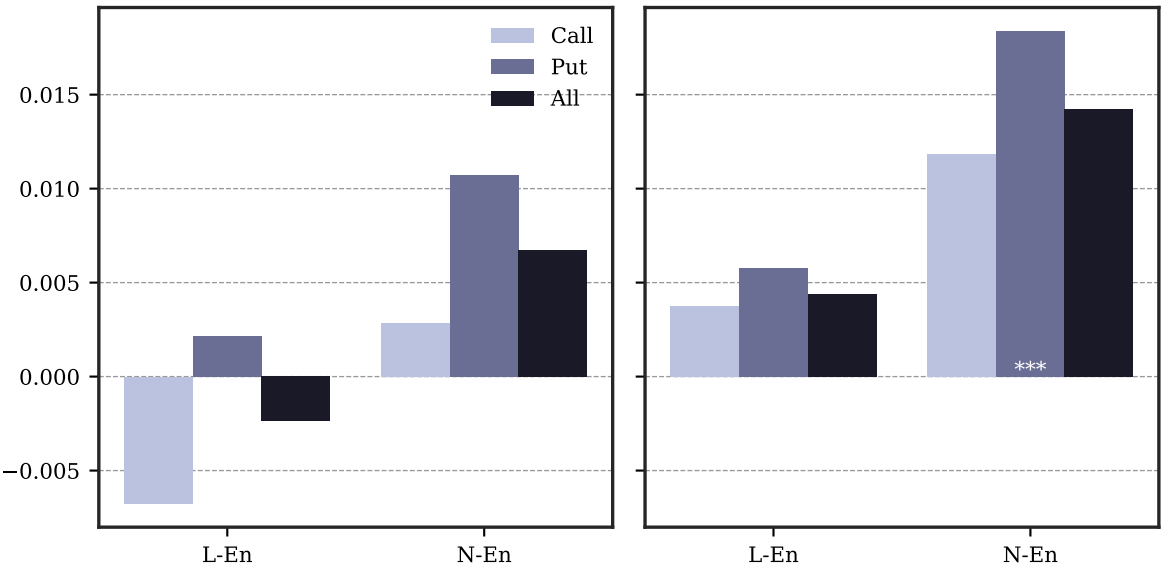

Fig. IA14.1. Predictability for Options on the 500 Largest CRSP-Universe Stocks

The figure shows the predictability measures R^2_{OS} in the left and $R^2_{OS;XS}$ in the right plot for options written on the 500 largest companies, as measured by the CRSP universe. We separately show the testing-sample predictability for all options and only calls and puts.

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.322	−0.878	1.592	−0.551	−1.476	−1.470	2.905	−0.506	***
2	−0.769	−0.485	1.632	−0.297	−0.765	−0.759	1.706	−0.445	*
3	−0.540	−0.390	1.409	−0.277	−0.456	−0.470	1.400	−0.336	
4	−0.368	−0.288	1.454	−0.198	−0.279	−0.290	1.254	−0.231	
5	−0.224	−0.230	1.510	−0.152	−0.154	−0.182	1.289	−0.141	
6	−0.091	−0.146	1.520	−0.096	−0.050	−0.134	1.326	−0.101	
7	0.039	−0.088	1.499	−0.058	0.052	−0.100	1.432	−0.070	
8	0.174	−0.055	1.521	−0.036	0.166	−0.060	1.496	−0.040	
9	0.337	0.011	1.431	0.007	0.323	0.044	1.572	0.028	
Hi	0.630	0.144	1.472	0.098	0.778	0.265	1.822	0.145	
H-L	1.952	1.022	1.472	0.694	2.254	1.734	2.273	0.763	***
		(10.83)		(6.49)		(13.17)		(6.16)	
call	1.509	0.760	1.436	0.529	1.458	1.115	1.515	0.736	**
put	1.700	0.821	1.689	0.486	1.548	1.230	1.763	0.698	**

Table IA14.1: Trading on Machine Learning Predictions – 500 Largest CRSP-Universe Stocks

The table shows the returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) method for options written on the 500 largest companies, as measured by the CRSP universe. 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 week. The last column (GFC) gives the significance of comparing the mean realized returns for the two model types. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

Appendix IA15. Weekly Investment Period

In this robustness analysis, we are considering option return predictability over a shorter horizon of one week. This investment choice naturally increases the sample size substantially, such that we focus on short-term and at-the-money options to be able to fit the models in a reasonable time frame.

Ex-ante it is unclear whether we should expect higher or lower levels of return predictability for the shorter horizon. On the one hand, a shorter horizon suggests a tighter temporal link between today's option characteristics and returns over the next week, suggesting higher return predictability. At the same time, measuring option returns over long horizons potentially introduces additional fluctuation, as the option migrates to shorter maturities. Options with shorter lifespans vary much more than their long-term counterpart. On the other hand, measuring returns over longer horizons potentially averages out some of the associated noise, which in turn would lead to higher levels of predictability for the monthly horizons.

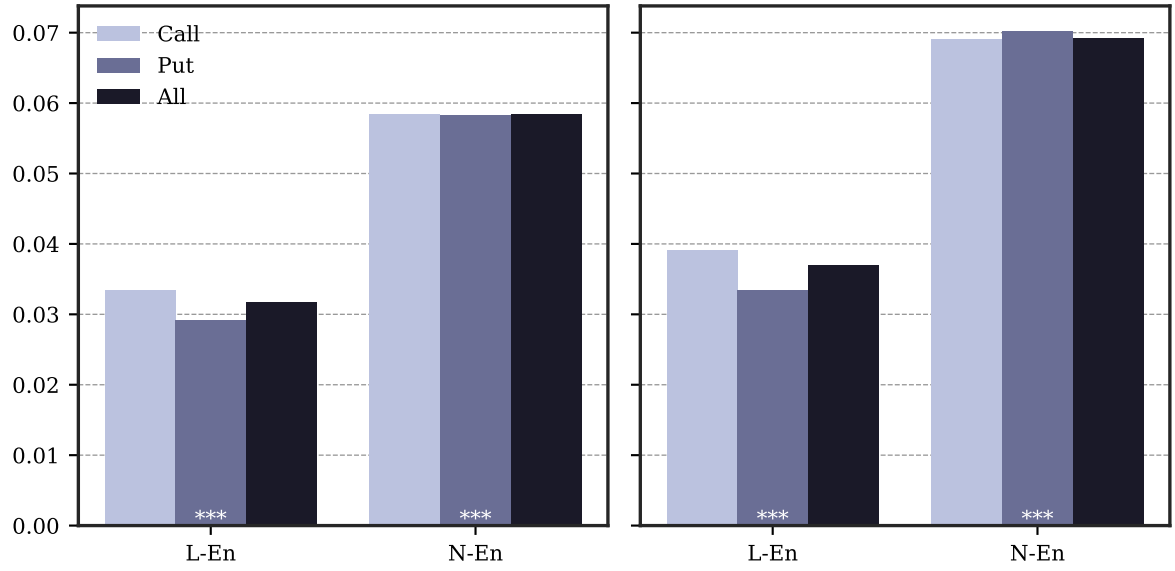

Fig. IA15.1. Predictability for Weekly Option Returns

The figure shows the impact on predictability measures R^2_{OS} in the left and $R^2_{OS;XS}$ in the right plot for the linear (L-En) and nonlinear ensembles (N-En) for weekly option returns, instead of the monthly return horizon in the main analyses. We focus our analysis on short-term at-the-money options, following the bucket definition in the paper. We separately show the testing-sample predictability for all options and calls and puts.

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−0.807	−0.768	0.792	−0.969	−1.181	−1.137	1.104	−1.030	***
2	−0.496	−0.468	0.791	−0.591	−0.645	−0.537	0.851	−0.631	*
3	−0.362	−0.330	0.737	−0.447	−0.447	−0.332	0.747	−0.445	
4	−0.263	−0.205	0.721	−0.285	−0.319	−0.214	0.652	−0.328	
5	−0.178	−0.118	0.705	−0.167	−0.221	−0.147	0.604	−0.244	
6	−0.100	−0.053	0.713	−0.074	−0.135	−0.078	0.599	−0.131	
7	−0.021	0.004	0.692	0.006	−0.049	−0.008	0.605	−0.013	
8	0.067	0.088	0.717	0.122	0.055	0.090	0.670	0.135	
9	0.181	0.203	0.748	0.271	0.211	0.236	0.786	0.301	
Hi	0.442	0.421	0.860	0.489	0.578	0.643	1.114	0.577	***
H-L	1.249	1.188	0.905	1.313	1.759	1.780	1.236	1.440	***
		(27.99)		(16.32)		(30.79)		(17.73)	
call	1.248	1.310	1.116	1.174	1.854	1.916	1.454	1.318	***
put	1.244	1.061	0.883	1.203	1.638	1.655	1.214	1.363	***

Table IA15.1: Trading on Machine Learning Predictions – Weekly Return Horizon

The table shows the returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) method for weekly option returns, instead of the monthly return horizon in the main analyses. We focus our analysis on short-term at-the-money options, following the bucket definition in the paper. 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 week. The last column (GFC) gives the significance of comparing the mean realized returns for the two model types. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

Appendix IA16. Alternating the Return Definition

IA16.1. Margin Requirements

In the main analyses we scale the delta-hedged portfolio gain by the cash requirement to enter into a delta-hedged option position. In the following, we change the return definition to incorporate margin requirements. Precisely, we change Equation (12) to

$$r_{t,t+\tau} = \frac{\Pi(t, t+\tau)}{M_t}, \tag{IA7}$$

where $M_t > 0$ denotes the margin requirement for sustaining a delta-hedged option position from t to $t+\tau$. For the exact margin requirements, we adopt the CBOE minimum margin for customer accounts.[2] Also assuming a 50% margin requirement for long and short positions in the underlying stock, the margin requirement is given as

$$M_t = \begin{cases} V_t + 0.5|\Delta_t|S_t, & \text{for hedged long positions} \\ V_t + \max(0.1S_t, 0.2S_t - \max(0, K - S_t)) + 0.5|\Delta_t|S_t, & \text{for hedged short calls} \\ 0.1K + 0.5|\Delta_t|S_t, & \text{for hedged short puts} \end{cases} \tag{IA8}$$

where K denotes the strike price, V_t the option price, and S_t is the price of the underlying stock.

We refit the models based on margin requirements for long delta-hedged positions. Figure IA16.1 reports that the statistical performance of the linear and non-linear ensemble in predicting margin adjusted returns remains largely unchanged. The resulting high-minus-low portfolio returns are given in Table IA16.1 and confirm the robustness of our results to using this alternative return specification.

Furthermore, in Table IA16.2 we adjust the short-leg of the high-minus-low portfolio to explicitly account for margin requirements of shorted options, which are typically much larger. We still base the portfolio selection using the returns which are adjusted by long margin requirements, as in Table IA16.1. Please note the difference to the third panel in Table 6. Return predictions in Table 6 are based on models which are fitted to the option return definition in Equation (12), whereas Table IA16.1 reports results for models fitted to the return definition in Equation (IA7).

[2] See https://www.cboe.com/us/options/strategy_based_margin.

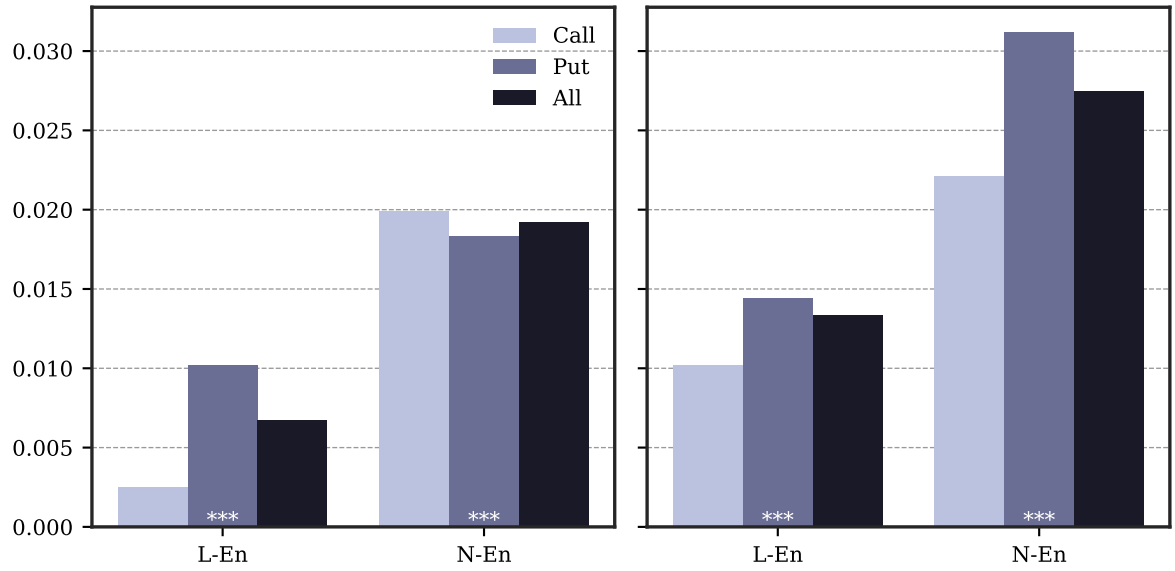

Fig. IA16.1. Predictability using Margin Requirement-Adjusted Returns

The figure shows the predictability measures R^2_{OS} in the left and $R^2_{OS;XS}$ in the right plot using margin requirement-adjusted returns as the target variable. We separately show the testing-sample predictability for all options and only calls and puts.

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.856	−1.325	1.757	−0.754	−2.308	−1.797	2.051	−0.876	***
2	−1.071	−0.618	1.971	−0.313	−1.093	−0.812	1.690	−0.480	
3	−0.764	−0.415	1.937	−0.214	−0.657	−0.496	1.546	−0.321	
4	−0.537	−0.291	1.951	−0.149	−0.410	−0.314	1.414	−0.222	
5	−0.347	−0.199	1.946	−0.102	−0.233	−0.191	1.420	−0.134	
6	−0.171	−0.150	1.852	−0.081	−0.086	−0.123	1.429	−0.086	
7	−0.002	−0.094	1.865	−0.050	0.056	−0.081	1.490	−0.054	
8	0.178	−0.057	1.837	−0.031	0.211	−0.010	1.562	−0.007	
9	0.391	0.018	1.751	0.010	0.416	0.108	1.616	0.067	
Hi	0.780	0.206	1.717	0.120	0.973	0.355	1.808	0.196	
H-L	2.636	1.531	1.439	1.064	3.281	2.152	1.571	1.370	***
		(17.06)		(7.94)		(15.92)		(8.62)	
call	2.389	1.202	1.530	0.785	2.903	2.263	1.978	1.144	***
put	2.583	1.840	1.832	1.004	3.314	2.268	1.783	1.272	**

Table IA16.1: Trading on Machine Learning Predictions – Margin Requirement-Adjusted Returns

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method using margin requirement-adjusted returns. 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.

Eff. Spread	L-En			N-En			N vs. L
	H-L	t	SR	H-L	t	SR	
			No Transaction Costs				
0%	1.856	(12.42)	0.834	2.541	(11.29)	1.090	***
			Option Costs				
25%	0.827	(7.52)	0.418	2.130	(8.41)	0.747	***
50%	0.374	(3.23)	0.183	1.646	(6.19)	0.539	***
75%	0.092	(0.74)	0.044	1.265	(4.18)	0.408	***
100%	−0.170	(−1.26)	−0.078	1.082	(3.64)	0.339	***
			Option And Delta-Hedging Costs				
25%	0.743	(6.89)	0.381	2.069	(8.50)	0.730	***
50%	0.301	(2.54)	0.149	1.403	(6.80)	0.487	***
75%	−0.022	(−0.17)	−0.011	0.951	(4.62)	0.334	***
100%	−0.275	(−1.87)	−0.129	0.654	(3.19)	0.238	***

Table IA16.2: Trading on Machine Learning Predictions – Different Margin Requirement for the Long and Short Portfolio

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method using margin requirement-adjusted returns. 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.

IA16.2. Delevered Returns

In this subsection, we account for the (time-varying) differences in leverage associated with the decile portfolios based on the predictions by the nonlinear ensemble N-En. While attaining different levels of leverage adds to the high performance of the resulting high-minus-low portfolio, the results are robust to accounting for it, highlighting that N-En manages to uncover trading opportunities in the options market.

	Lev_p		$\text{Lev}_{t,p}$		$\text{Lev}_{t,o}$	
	Avg	SR	Avg	SR	Avg	SR
Lo	−0.486	−0.849	−0.554	−0.718	−1.054	−0.537
2	−0.159	−0.458	−0.165	−0.354	−0.302	−0.329
3	−0.080	−0.303	−0.083	−0.266	−0.124	−0.218
4	−0.048	−0.212	−0.053	−0.200	−0.077	−0.172
5	−0.028	−0.128	−0.033	−0.126	−0.051	−0.139
6	−0.019	−0.090	−0.021	−0.076	−0.033	−0.087
7	−0.011	−0.053	−0.013	−0.043	−0.016	−0.040
8	−0.005	−0.021	−0.001	−0.002	0.008	0.019
9	0.013	0.058	0.024	0.067	0.050	0.109
Hi	0.067	0.213	0.091	0.176	0.157	0.196
H-L	0.553	1.252	0.646	0.950	1.211	0.645
	(15.63)	(7.79)	(7.68)	(9.70)	(6.97)	(5.50)
call	0.517	1.140	0.584	0.900	0.768	0.907
put	0.619	1.242	0.759	0.921	1.586	0.627

Table IA16.3: Adjusting for Time-Varying Leverage in the Machine Learning Portfolios

The table shows the returns to option portfolios sorted by the predictions made by the nonlinear ensemble (N-En) method when we account for the leverage of the traded options. Columns Lev_p scale the realized excess returns by portfolio p's average leverage. Columns $\text{Lev}_{t,p}$ do so on a time-varying basis and columns $\text{Lev}_{t,o}$ explicitly adjust returns of each traded option by the same option's leverage in month t. Each contract is weighted by its dollar open interest at the time of investment. Avg denotes the average realized return and SR the realized Sharpe ratio. All values are given per month.

Appendix IA17. Statistical And Economic Performance for Options Buckets

IA17.1. Comparison between N-En and L-En for Different Buckets

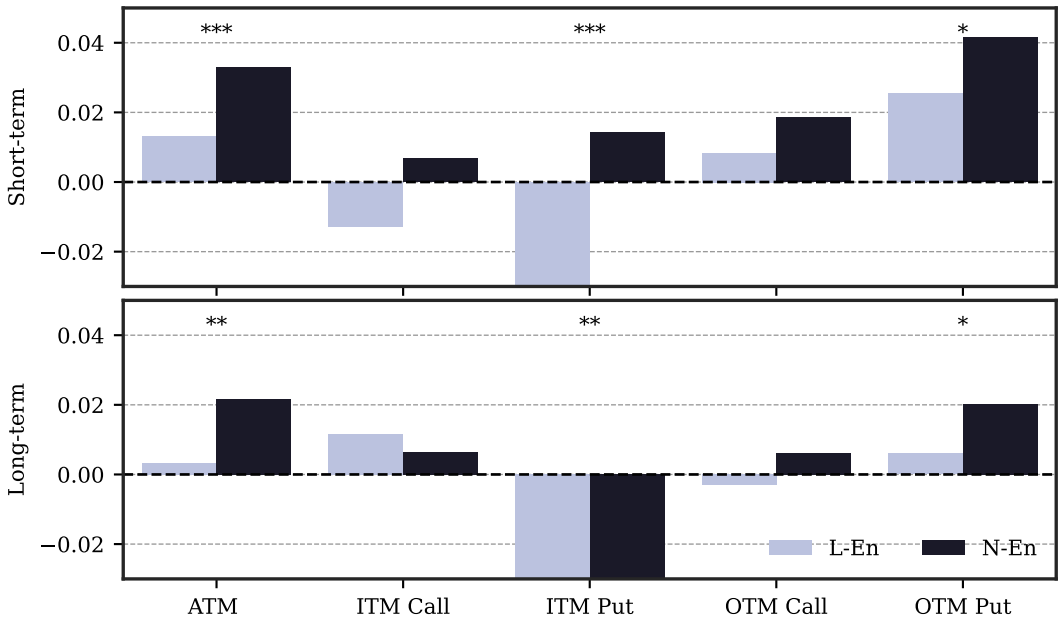

Fig. IA17.1. Predictive Power and Impact of Nonlinear Models for Option Buckets

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 5. ***, **, * below the bars denotes a statistically significant outperformance of N-En at the 0.1%, 1% and 5% level as defined in Equation (8).

IA17.2. 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		1.900	1.783	1.667	1.069	2.706	2.337	2.036	1.148	***
	itm	C	1.953	0.709	1.057	0.671	2.227	1.028	1.190	0.864	***
$\tau \leq 90$	itm	P	2.345	0.965	1.021	0.946	2.169	1.388	1.435	0.967	***
	otm	C	1.983	3.164	3.889	0.814	3.024	3.406	4.168	0.817	
	otm	P	1.697	3.396	3.121	1.088	2.995	4.489	3.859	1.163	***
	atm		1.849	1.297	1.726	0.751	2.235	2.245	2.084	1.078	***
	itm	C	1.572	0.842	1.152	0.731	1.759	1.216	1.558	0.780	***
$\tau > 90$	itm	P	2.152	0.628	1.541	0.407	1.722	0.879	1.729	0.508	
	otm	C	1.646	2.040	4.829	0.422	2.271	2.982	5.351	0.557	*
	otm	P	1.534	1.636	4.137	0.395	2.281	2.797	4.023	0.695	***

Table IA17.1: 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 5 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.

IA17.3. Predicting Individual Contracts vs. Option Portfolios

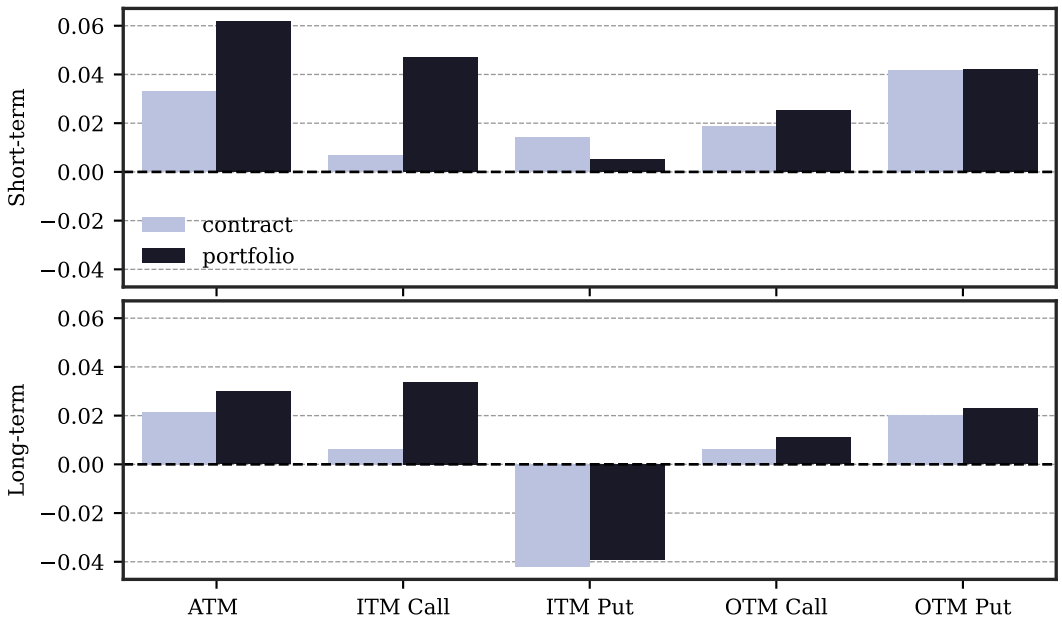

Fig. IA17.2. Predictive Power and Impact of Nonlinear Models for Option Portfolios

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 5) and predict future returns using N-En. This approach closely mimics recent studies on option return predictability (Cao and Han, 2013; Zhan et al., 2022; Goyenko and Zhang, 2021). We compare the resulting predictability levels with those obtained when predicting the individual contracts included in each bucket. Figure IA17.3 repeats this exercise using the cross-sectional out-of-sample $R^2_{OS;XS}$.

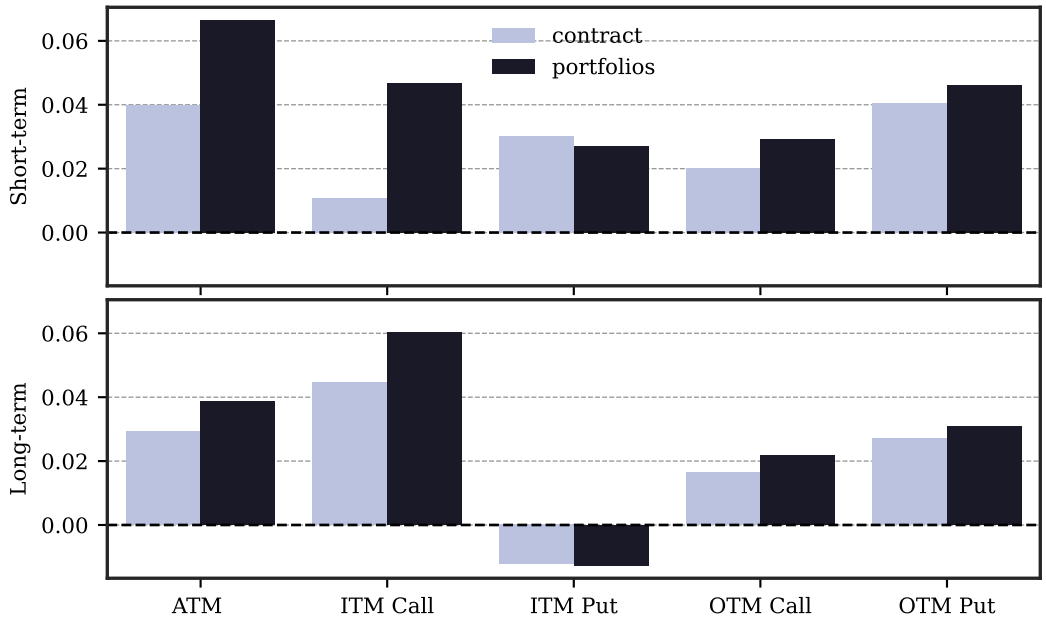

Fig. IA17.3. Predictive Power and Impact of Nonlinear Models for Option Portfolios

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 5) and predict future returns using N-En. This approach closely mimics recent studies on option return predictability (Cao and Han, 2013; Zhan et al., 2022; Goyenko and Zhang, 2021). We compare the resulting predictability levels with those obtained when predicting the individual contracts included in each bucket.

Appendix IA18. 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 5. 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 a slightly better predictability for the "ATM" model when applied to the sample of only short-term at-the-money options, but importantly a very similar performance of the full model when applied to only those options. The slight outperformance of the "ATM" model 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 all nonlinear models and most impressively for the nonlinear ensemble.

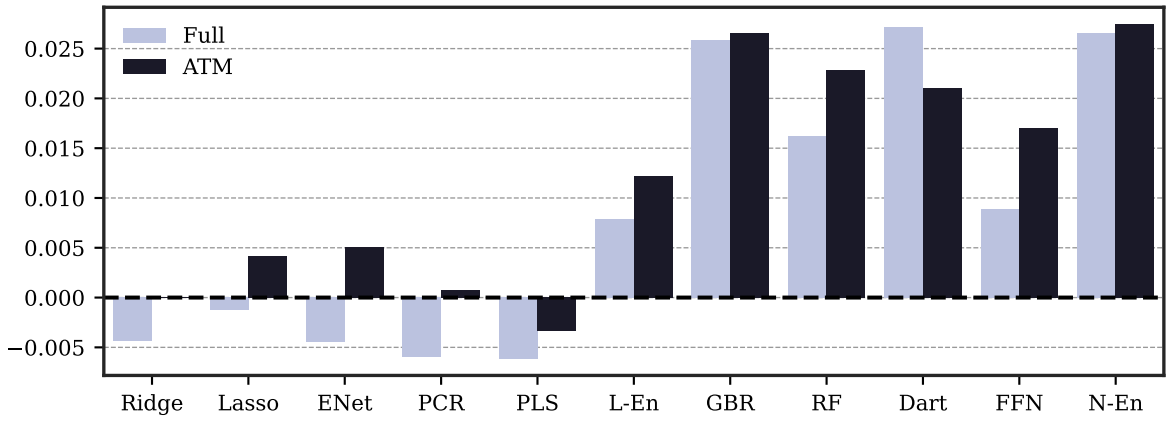

Fig. IA18.1. R^2_{OS} Model Comparison

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.

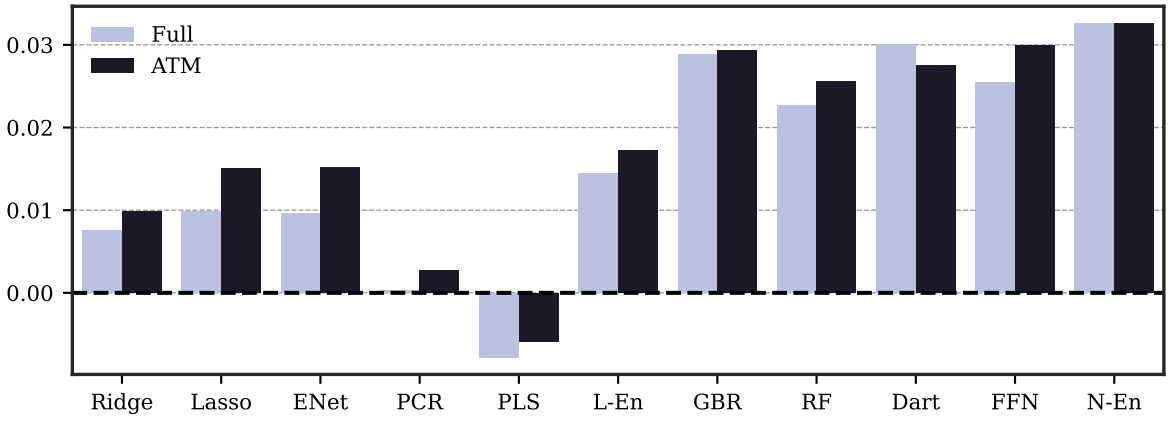

Fig. IA18.2. $R^2_{OS;XS}$ Model Comparison

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.

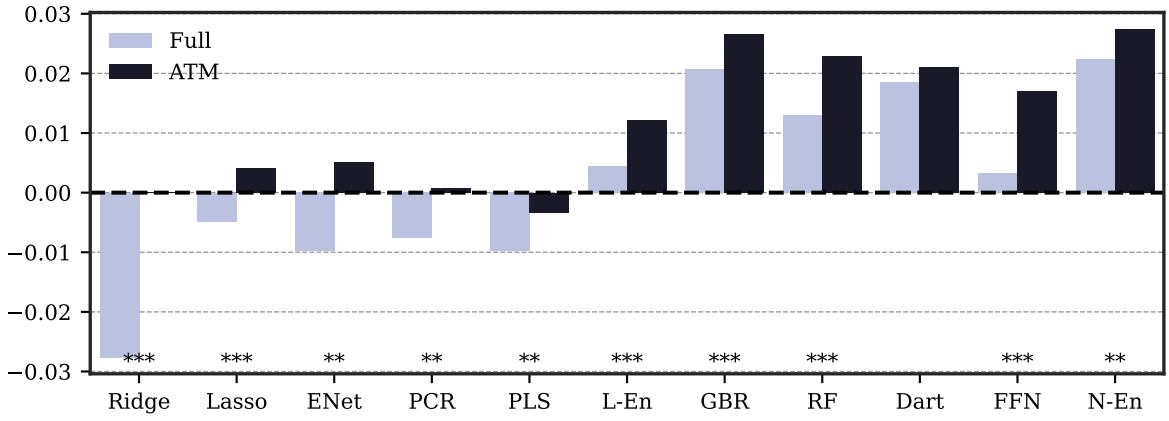

Fig. IA18.3. R^2_{OS} Model Comparison using only Short-Term At-The-Money Options

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 models trained using all data, but applied to short-term at-the-money options only. The testing sample spans the years 2003 through 2020.

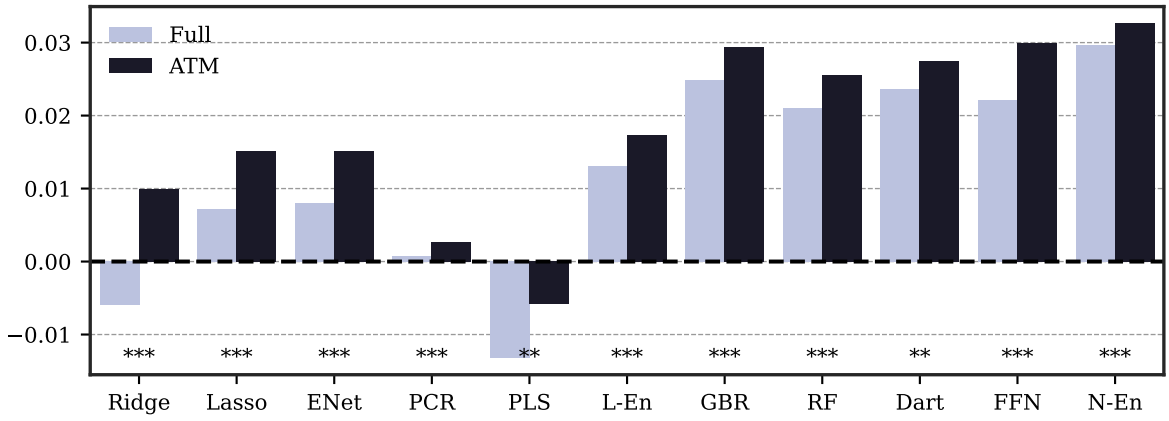

Fig. IA18.4. $R^2_{OS;XS}$ Model Comparison using only Short-Term At-The-Money Options

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 models trained using all data, but applied to short-term at-the-money options only. The testing sample spans the years 2003 through 2020.

Appendix IA19. Sources of Option Return Predictability

Additional Analyses

IA19.1. Informational Frictions

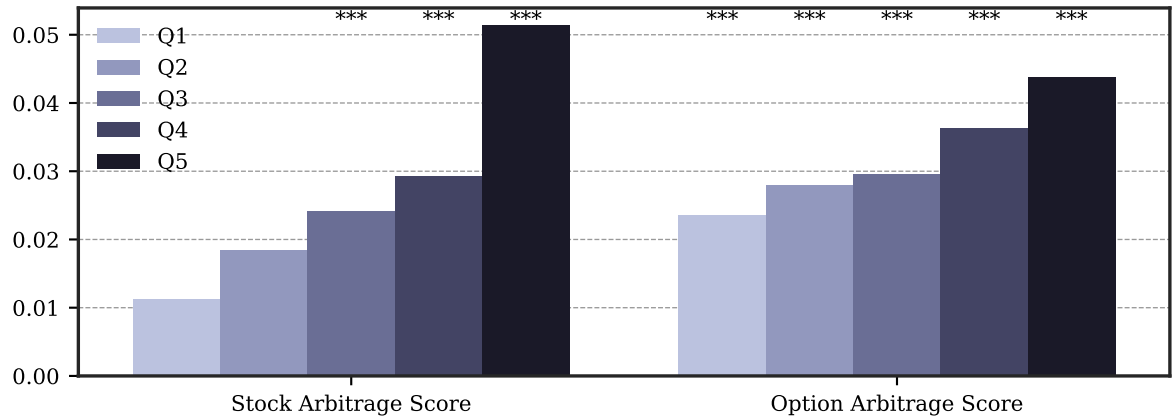

Fig. IA19.1. Predictability Conditional on Information Frictions – $R^2_{OS;XS}$

The figure shows the cross-sectional predictability using $R^2_{OS;XS}$ of the nonlinear ensemble for options sorted into quintiles by an index of informational frictions on the underlying-level and an index of informational frictions on the option-level. Index constructions follows Atilgan et al. (2020). Firm size, firm age, idiosyncratic volatility, institutional ownership, and analyst coverage are used to construct the index on the stock-level. The option contract's bid-ask spread, margin-requirement, dollar open interest, bucket volume, the volatility of the implied volatility, the historical excess-kurtosis of the underlying, and the underlying's bid-ask spread are taken for the index construction on the option-level. Since the level of institutional ownership, analyst coverage, firm age, firm size, dollar open interest, and bucket volume are inversely related to informational frictions, these characteristics are inversely sorted in the indices construction.

IA19.2. Option Mispricing

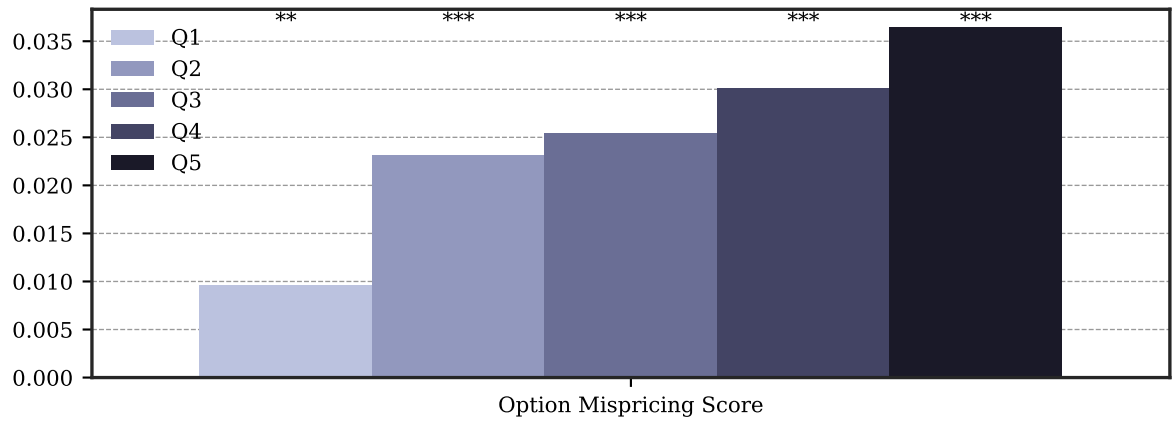

Fig. IA19.2. Predictability and Profitability Conditional on Option Mispricing – $R^2_{OS;XS}$

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 absolute option mispricing using a composite mispricing score. As inputs, we use $iv-rv$ (Goyal and Saretto, 2009; Carr and Wu, 2009), the mispricing measure by Eisdorfer et al. (2022), as well as the absolute return prediction of the nonlinear ensemble. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7).

References

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., Chabi-Yo, F., Murray, S., 2022. A factor model for stock returns based on option prices. Working paper.

Bali, T. G., Hovakimian, A., 2009. Volatility spreads and expected stock returns. Management Science 55, 1797–1812.

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. Working paper.

Cao, J., Vasquez, A., Xiao, X., Zhan, X., 2019. Volatility Uncertainty and the Cross-Section of Option Returns. Working paper.

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.

Eisdorfer, A., Goyal, A., Zhdanov, A., 2022. Limited attention and option prices. Working paper.

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 .

Goyal, A., Saretto, A., 2009. Cross-section of option returns and volatility. Journal of Financial Economics 94, 310–326.

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., 2022. The leverage bearing capacity: A new tool for intermediary asset pricing. Working paper.

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. Available at SSRN 3705573 .

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. Working paper.

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., 2022. Is there a replication crisis in finance? Journal of Finance, forthcoming.

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. Available at SSRN 3705500 .

Karakaya, M. M., 2014. Characteristics and expected returns in individual equity options. Working paper.

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.

Keloharju, M., Linnainmaa, J. T., Nyberg, P., 2016. Return seasonalities. Journal of Finance 71, 1557–1590.

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 .

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.

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.

Vasquez, A., Xiao, X., 2021. Default risk and option returns. Working paper.

Vilkov, G., Xiao, Y., 2012. Option-implied information and predictability of extreme returns. Working paper.

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.