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. Besides statistical significance, the nonlinear machine learning models generate economically sizeable profits in the long-short portfolios of equity options even after accounting for transaction costs. Although option-based characteristics are the most important standalone predictors, stock-based measures offer substantial incremental predictive power when considered alongside option-based characteristics. Finally, we provide compelling evidence that option return predictability is driven by informational frictions, costly arbitrage, and option mispricing.

JEL classification: G10, G12, G13, G14

Keywords: Machine learning, big data, option return predictability

*We thank Jay Cao, Victor DeMiguel, Ilias Filippou, Seth Pruitt, and Guofu Zhou for their insightful and constructive comments. We also thank Andrea Barbon for sharing open-close CBOE C1 exchange volume data with us.

†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 eight years. 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 percentage growth of more than 125%. Approximately 60% of these contracts are written on individual stocks and stock indices, making equity the most popular option underlying of financial market participants. Given the high popularity of options trading by investors, the question arises whether individual option returns are predictable and, if yes, which characteristics can give rise to such predictability. Our paper is devoted to answer these questions.

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

We study the cross-section of individual U.S. equity option returns using data from *OptionMetrics IvyDB* over the period from January 1996 to December 2020. To abstract from the directional exposure to the underlying, we follow Bakshi 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 270 variables composed of 77 option-based characteristics (e.g., option illiquidity, time-to maturity, or the implied shorting fee)

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

To obtain out-of-sample predictions from our different machine learning models, we proceed as follows: In a first step, we estimate the model parameters on a training sample of five years. Subsequently, we optimize hyperparameters of the models on a two-year validation sample. Lastly, the tuned models are used to forecast individual option returns over the next year. This procedure is repeated for each year, expanding the training sample by one year and consequently shifting the validation and out-of-sample period by one year. To assess the predictive power of the different models for individual option returns, we follow Gu et al. (2020) and Bali, Goyal, Huang, Jiang, and Wen (2021) and use the out-of-sample R^2-statistic, which benchmarks the R^2 against a forecast of zero excess returns.[2] To make pairwise comparisons of the forecast accuracy of different machine learning models, we utilize the model-free Diebold and Mariano (1995) test statistic.

Our empirical results advance the knowledge on predictability of the cross-section of individual option returns in various dimensions: First, we show that complexity of the prediction model matters. While none of the *linear* models manages to produce positive out-of-sample R^2s for the entire testing sample, all *nonlinear* models do. Our results

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

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

reveal that the best-performing models are gradient-boosted regression trees with and without dropout (*GBR* and *Dart*) that show out-of-sample R^2s of 2.58% and 2.72%.[3] Moreover, the equal-weighted ensemble of all nonlinear models beats the ensemble of all linear models by more than 1.8% 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 most other models significantly at the 5% level. The only exceptions to this rule are *GBR* and *Dart*, which both produce forecasts highly correlated with the nonlinear ensemble model (correlations amount to 0.95 and 0.94, respectively). The outperformance of nonlinear models compared to linear models is stable over time with a higher predictability for future option returns in 71.2% of the months in our sample.[4] Interestingly, we also find better predictions for the nonlinear models during December 2019 to April 2020, i.e., the period in which the COVID19 pandemic shook financial markets worldwide.[5]

Second, we inspect whether predictability of option returns through machine learning models can be exploited in an economically profitable trading strategy. To do so, in each month t, we first 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 one-month-ahead average realized returns 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 selling options with the lowest expected return forecast (decile 1). Our results indicate that trading strategies based on both ensembles generate economically significant return spreads of 2.63% and 1.92% per month, both of which are statistically significant at the 1% level. The respective monthly Sharpe ratios amount to 2.06 and 1.55. As before, our empirical results stress the importance for nonlinearities in the machine learning models. The long-minus-short return spread of the nonlinear ensemble outperforms the return spread of the linear ensemble by statistically significant 0.71% per month. We find that the short leg of the arbitrage portfolio, formed by the nonlinear ensemble, contains more

[3]Note that the magnitude of these R^2s is considerably higher than the corresponding numbers for the cross-section of stock returns (Gu et al. (2020) find out-sample R^2 of approximately 0.6% for nonlinear machine learning models), but slightly lower than for the cross-section of bond returns (Bali et al. (2021) document out-of-sample R^2s of approximately 3.5% for nonlinear machine learning models).

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

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

short-term and fewer call options, while the long leg includes more long-term call options.

Is the displayed long-minus-short return spread of the nonlinear ensemble robust to risk adjustments of established asset pricing models? We investigate this question by regressing the spread return on the market, the risk factors of the Fama and French (2015) five-factor model extended either by the Carhart (1997) momentum factor, or by the momentum and the Pástor and Stambaugh (2003) liquidity factor, a model using the Agarwal and Naik (2004) option-market factors, the leverage bearing capacity of intermediaries proposed in Grünthaler, Lorenz, and Meyerhof (2020), and the Bali and Murray (2020) factor model for optionable stocks. In all cases we find that the long-short return spread is not substantially reduced by systematic risk adjustments and remains economically and statistically significant. We conclude that the outperformance of a trading strategy based on the nonlinear ensemble is not explained by various asset pricing factors.

Ofek, Richardson, and Whitelaw (2004) observe that transaction costs in the options market are high and that these costs can substantially reduce economic profits of option-based trading strategies. Hence, to understand in how far the machine learning trading strategy based on the nonlinear ensemble is implementable, we examine its profitability after account for transaction costs. Since actual transaction costs of trades are not observable in the OptionMetrics IvyDB database, we assume that investors have to pay a percentage of the quoted bid and ask spread, which we denote as the effective spread (Eisdorfer et al., 2020). We consider the effective spreads of 15% and 25%. Our results show that the monthly returns of the machine learning trading strategy remain sizeable with monthly returns of 1.33% (0.47%) after accounting for 15% (25%) effective spreads. Moreover, we find that one-year rolling returns of the transaction cost-adjusted trading strategy stay above zero for 15% effective spreads in all months of our sample from January 2003 to December 2020.[6]

As our third main empirical result, we quantify the relative importance of different variables for the prediction of option returns. For this purpose, we classify our 270 option and stock predictor variables into twelve sub-groups: *Accruals*, *industry*, *investment*, *profitability*, *quality*, *value*, *contract*, *frictions*, *illiquidity*, *informed trading*, *past prices*, and *risk*. To uncover influential sub-groups for prediction, we follow recent advances in computer science and compute SHAP values (Lundberg and Lee, 2017), which ap-

[6]Trading strategies using only short-term options manage to survive 25% effective spreads for the entire sample, and in the case of out-of-the-money short-term puts, the strategies remain profitable even with 50% effective spreads.

proximate changes in the model predictions had we excluded certain characteristics in its estimation. 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 (e.g., an option's moneyness and time-to-maturity). *Illiquidity* (e.g. bid-ask spreads of the option and the underlying stock) and *risk* (e.g., risk-neural moments of the underlying) measures follow as the second and third most important variable group, respectively.

Our empirical setting also enables us to check whether option or stock characteristics are more powerful to accurately predict future option returns. Hence, we re-estimate the machine learning models (i) using only option-based characteristics, (ii) stock-based characteristics, as well as (iii) option-based characteristics that operate on the bucket- or contract level, and compare the out-of-sample forecasting results with the full information set of *all* option and stock characteristics.[7] We observe that the models only based on a subset of information show severely lower out-of-sample R^2s compared to the models that incorporate all option and stock characteristics. When comparing different subsets of information, our results indicate that restricting information to only option-based characteristics yields substantially higher predictive R^2s than information of only stock-based characteristics. The benefit of option-derived characteristics is large and should not be neglected when making informed forecasts on future option returns.

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

[7] In contrast to the feature importance described above, this approach has the benefit of correctly accounting for interaction effects between different feature groups.

return predictability is highest for options with little demand by professional traders, but substantial interest by public customers. We also find high levels of predictability when public customers are long the option and institutional investors take the short side of the contract, a phenomenon also reported in Garleanu et al. (2009) where institutional traders act as market makers in cases where they deem option prices sufficiently high to short them.[8] In addition, we consider costly arbitrage as another source of predictability. Our empirical results reveal a monotonically increasing relation between option return predictability and illiquidity of the underlying stock. For the highest (lowest) quintile of stock illiqudity, the out-of-sample R^2 for the nonlinear ensemble model equals close to 5% (0%). Lastly, we estimate the level of mispricing in a given underlying's options by using high-frequency realized volatility as an input to the Black and Scholes (1973) model. Consistent with our intuition that the machine learning models manage to identify both situations of over- *and* under-pricing, we find high levels of R^2_{OS}, exceeding 2.5% in both cases. Turning to the profitability of the machine learning portfolios conditional on the level of mispricing, we find that betting on the portfolios for the subset of mispriced options is significantly more profitable than following the predictions for the least mispriced options.

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

2. Related Literature

Our paper contributes to the literature on predicting and explaining the cross-section of individual option returns. Dennis and Mayhew (2002) document the importance of various factors, such as beta, size, and trading volume in explaining the risk-neutral volatility skew observed in stock option prices, whereas Bollen and Whaley (2004) investigate the relation between net buying pressure and the shape of the implied volatility

[8] A situation like this (i.e., public customers buy and professional traders sell) is potentially explained by different feature preferences for the option contract by the two investor groups: While risk-based measures are most important whenever both investor groups trade in the same direction, whenever professionals trade against public customers, our machine learning algorithms use more information about informed trading.

function of stock options. Garleanu et al. (2009) theoretically model and empirically investigate demand-pressure effects on option prices. Roll, Schwartz, and Subrahmanyam (2010) examine trading volume in option markets relative to the volume in underlying stocks and relate it to contemporaneous returns. Examining volatility risk in the options market, Goyal and Saretto (2009) find that options with a large positive difference between realized and implied volatility have low future delta-hedged returns. Cao and Han (2013) find that delta-hedged option returns decrease monotonically with an increase in the idiosyncratic volatility of the underlying stock. In a similar vein, Byun and Kim (2016) show that call options written on the most lottery-like stocks underperform otherwise similar call options on the least lottery-like stocks by more than 10% per month. An, Ang, Bali, and Cakici (2014) show, among others, that the cross-section of stock returns predicts option implied volatilities. Christoffersen, Goyenko, Jacobs, and Karoui (2018) include illiquidity premia in option valuation models and Kanne, Korn, and Uhrig-Homburg (2020) find that these premia are negative (positive) if there is net buying (selling) pressure. Ramachandran and Tayal (2021) report a monotonic relationship between various measures of short-sale constraints and delta-hedged returns of put options on overpriced stocks. Cao, Han, Tong, and Zhan (2021) uncover return predictability in the cross-section of delta-hedged equity options based on stock-based information, such as prices, profit margins, and firm profitability. Finally, in a contemporaneous working paper to ours, Goyenko and Zhang (2021) apply machine learning techniques to analyze which characteristics drive option and stock returns. Our paper differs from Goyenko and Zhang (2021) in several important aspects: First, our sample of option returns is substantially larger and comprises more than 33 times as many option-month observations. Instead of analyzing option portfolios as in Goyenko and Zhang (2021), we focus on single option contracts. Second, our paper focuses on the predictability of the cross-section of option returns, while Goyenko and Zhang (2021) also consider option determinants to predict the cross-section of stock returns. Finally, we provide a comprehensive investigation of the economic underpinnings of option return predictability and offer important insights on the cross-sectional pricing of equity options with machine learning and big data.

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:[9] Rapach, Strauss, and Zhou (2013) are the

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

first to use Lasso in predicting market returns across countries. Moritz and Zimmermann (2016) apply tree-based conditional portfolio sorts to examine the relation between past and future stock returns. Kelly, Pruitt, and Su (2019) apply instrumented principal component analysis, 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 asset risk premiums based on a large set of stock characteristics. Whereas Gu et al. (2020) use a broad set of stock characteristics (see Green et al., 2017), Murray, Xiao, and Xia (2021) focus solely on historical price data. The authors find that machine learning based forecasts on past cumulative returns have economically and statistically significant predictive power for the cross-section of future stock returns. Neuhierl, Tang, Varneskov, and Zhou (2021) examine the predictive power of option characteristics for the cross-section of stock returns by jointly using broad sets of firm and option characteristics. Kozak, Nagel, and Santosh (2020) 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 (2021) add to these insights, using deep neural networks to estimate an asset pricing model for individual stock returns. Martin and Nagel (2020) 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) 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 resting on instrumented principal component analysis. Bali et al. (2021) find that machine learning models substantially improve the out-of-sample performance of stock and bond charac-

teristics 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 various stock characteristics. Our article is — simultaneously with Goyenko and Zhang (2021) — the first to predict the cross-section of individual option returns using a large set of linear and nonlinear machine learning models.

3. Methodology and Performance Evaluation

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

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

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 $z_{i,s}$ 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; Bianchi et al., 2021; Bali et al., 2021), 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 IA1 provides a detailed description of these methods.

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

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

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

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

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

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

R^2_{OS} measures the reduction in the mean squared forecast error (MSFE) compared to a naive benchmark of zero excess returns for all options. We evaluate the predictive power on a testing sample, which is disjoint from the data used to estimate the model parameters and hyperparameters (such as the magnitude of the Lasso penalty). More specifically, we start by estimating model parameters on a training sample $\mathcal{T}_1$ of five years (January 1996 – December 2000). We then perform an extensive hyperparameter optimization validating the method's fit in the next two years $\mathcal{T}_2$ (2001 – 2002). Lastly, for each method, we use an equal-weighted ensemble of the eight models with the 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 IA2 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 $(r_{i,s,t+1} - \bar{r}_{i,s,t})$. $R^2_{OS;XS}$ focuses on relative expected returns across options, for which accurate predictions result in profitable long-short trading strategies.

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

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

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:

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

Here, $n_{\mathcal{T}_3}$ is the number of observations in the testing sample and $\hat{e}^{(j)}_{i,t+1}$ are the forecast errors on option i at time $t+1$ for method j. We use 12 lags for the standard errors, coinciding with the number of months we keep model parameters fixed for each slice of the testing sample.

Forecast comparison: To compare the forecasts of two methods, we use the modified Diebold and Mariano (1995) (DM) test, which accounts for potential cross-sectional dependence in equity option returns. The DM test-statistic for a comparison between methods 1 and 2 is defined as:

$$\mathrm{DM}^{(1,2)} = \frac{\bar{d}^{(1,2)}}{\hat{\sigma}_d^{(1,2)}}, \tag{8}$$

where $\bar{d}^{(1,2)}$ and $\hat{\sigma}_d^{(1,2)}$ denote the time-series average and Newey and West (1987) standard

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

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

We also use correlations as a secondary method to assess how similar the forecasts of different methods are. Formally, the forecast correlation is defined as:

$$\rho_{t+1}^{(1,2)} = \frac{Cov(\hat{e}_{t+1}^{(1)}, \hat{e}_{t+1}^{(2)})}{\sigma(\hat{e}_{t+1}^{(1)})\sigma(\hat{e}_{t+1}^{(2)})}, \tag{10}$$

where $\sigma(\hat{e}_{t+1}^{(1)})$ and $\sigma(\hat{e}_{t+1}^{(2)})$ denote the standard deviations of forecast errors for models 1 and 2, respectively, and $Cov(\hat{e}_{t+1}^{(1)}, \hat{e}_{t+1}^{(2)})$ denotes their covariance. Divergence in the forecasting power of two methods provides a high-level view on why some methods outperform.

4. Data and Variable Definitions

We first outline the data sources used and then provide summary statistics for the option sample and the sample of underlying optionable stocks. Our primary data source is OptionMetrics IvyDB, which provides historical prices for all U.S. single equity options. We also use the interpolated volatility surface data from OptionMetrics. Due to the starting date of this database, our sample covers the period from January 1996 through December 2020.

Historical prices and accounting data for underlying stocks are obtained from CRSP and Compustat. We retain only underlyings with share codes 10 or 11 and exchange codes 1, 2, 3, 31, 32, 33; i.e., stocks listed on the NYSE, NYSE American (formerly AMEX) or NASDAQ. Contrary to previous studies (see Cao et al., 2021), we purposely do not remove stocks with nominal prices below \$5 per share, as Eisdorfer et al. (2020) find that options trading on stocks with a low nominal price tend to be overpriced. Information on stock splits and dividend payments is taken from OptionMetrics and cross-checked with CRSP. We match these databases using the linking algorithm developed by WRDS. Daily risk-free rates are taken from Kenneth French's website.[10]

Option returns are notoriously noisy, especially for underlyings with few outstanding option contracts and less option trading activity. We therefore rely on a variety of stan-

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

dard filters established in the literature to assure the consistency of our analyses (Goyal and Saretto, 2009; Cao and Han, 2013; Muravyev, 2016; Cao et al., 2021).[11] 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. Finally, we check for the convexity of option prices per underlying following Bollerslev, Todorov, and Xu (2015). Specifically, we retain only those observations for which the difference between the prices of two neighboring call (put) options with strike price $K_1 < K_2$ is ≤ 0 (≥ 0).

4.1. Option Returns

Our main variable of interest is the excess return of buying an option that we delta-hedge on a daily rebalancing schedule. We consider delta-hedged option gains following Bakshi 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 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:

$$\begin{aligned}\Pi(t, t+\tau) = V_{t+\tau} - V_t - \sum_{n=0}^{N-1} \Delta_{V,t_n} \times [S(t_{n+1}) - S(t_n)] \\ - \sum_{n=0}^{N-1} \frac{a_n r_n}{365} \left[V(t_n) - \Delta_{V,t_n} S(t_n)\right], \end{aligned} \tag{11}$$

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

[11] 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.

month. When an individual stock exhibits large price movements over the investment horizon, establishing initial delta-neutrality may still expose the investor to substantial sensitivity to future movements in the underlying stock. Hence, we opt for delta-hedging at the end of each trading day. Tian and Wu (2021) estimate that one-time delta-hedging at initiation removes around 70% of the directional risks embedded in options, whereas daily delta-hedging manages to eliminate upwards of 90% of these risks. Many studies have found predictability for stock returns (Gu et al., 2020; Freyberger et al., 2020; Kozak et al., 2020). Instead, our objective is to understand how characteristics relate to the pricing of higher-order risks embedded in options. Daily delta-hedging enables us to do just that. Finally, we define option returns following Cao and Han (2013) and Cao et al. (2021) as:

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

4.2. Summary Statistics

After the data filters discussed earlier, our sample comprises 4,932,372 options on 7,046 unique underlyings, for a total of 12,129,844 option-month observations for the period January 1996 - December 2020. Our sample is made up of roughly 53% call and 47% put options. Panel A of Table 1 shows that the average monthly delta-hedged option return is -0.01%, whereas the median monthly delta-hedged option return is -0.34%. The average moneyness is 1.03 and the average implied volatility is 47.67%. The average (median) days to maturity are 193 (113), while every fourth option exhibits a time to maturity of less than 50 days. As Panels B and C depict, the median monthly delta-hedged option returns are slightly positive for call options at an average 0.1% per month, but strongly negative for put options, -0.13%. The median return, however, is negative for both puts and calls. Panel D in Table 1 shows summary statistics for the years of 1996 through 2002 which are used in the training step of the machine learning models, while Panel E gives the summary statistics only for the testing subsample from 2003 through 2020. Average monthly delta-hedged option returns tend to be slightly more negative for the more recent time period. Moreover, implied volatility and moneyness are lower and days to maturity higher for the testing period from 2003 to 2020.

Table IA5.1 in the Internet Appendix reports summary statistics of the 7,046 stocks in our sample. Our sample includes on average 1,747 optionable stocks per month, which comprise 83% of the total market capitalization of the U.S equity market. Moreover, our sample comprises large stocks with representative volatility, given that the average

size and volatility percentile within the total stock universe are 72 and 45, respectively. Finally, the Fama-French 12-industry distribution in our sample is comparable to the total stock sample, as evident from Panel C in Table IA5.1.

4.3. Option and Equity Characteristics

Throughout this paper, we differentiate between different parts of the time-to-maturity and moneyness domain of options, which we refer to as "buckets". Specifically, we separately consider predictability for short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (ITM: $K/S > 1.1$ for puts, $K/S < 0.9$ for calls), out-of-the-money (OTM: $K/S < 0.9$ for puts, $K/S < 1.1$ for calls) calls and puts, and at-the-money options (ATM: $0.9 \leq K/S \leq 1.1$), as well as time-to-maturity and moneyness combinations. The moneyness and time-to-maturity of option contracts change rapidly. Therefore, it is unreasonable to assume that flow measures, such as option volume, derived from a particular option contract over a historical period will be valid for the same contract in the next month. The defined buckets allow us to compute these flow measures for option contracts, as we abstract from the impact of changing moneyness and fleeting time-to-maturity. Table IA5.3 shows that on average, 15 long term options and 11 short term options belong to each underlying stock, with the bulk located at the current price of the underlying.

We build a comprehensive set of option-based characteristics, motivated by earlier studies on the cross-section of option and/or stock returns. Out of the 77 we compute, 42 characteristics operate on the level of the underlying (e.g., the implied shorting fee, Muravyev, Pearson, and Pollet, 2021), 19 on the level of option buckets (e.g., the Amihud, 2002, illiquidity measure) and 16 on the level of individual option contracts (e.g., the option's time-to-maturity). Appendix IA3 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),

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: All Options (N=12,129,844)							
Delta-Hedged Return	-0.01	5.17	-4.37	-1.94	-0.34	1.29	4.47
Days to Maturity	172.24	179.44	21.0	50.0	113.0	204.0	444.0
Moneyness	1.03	0.31	0.77	0.89	1.0	1.11	1.28
Implied volatility	47.67	25.62	23.46	30.2	41.18	58.0	80.07
Absolute Delta	0.46	0.25	0.13	0.25	0.44	0.66	0.81
Panel B: Call Options (N=6,553,753)							
Delta-Hedged Return	0.1	5.62	-4.72	-2.01	-0.29	1.53	5.14
Days to Maturity	176.0	181.95	21.0	50.0	113.0	204.0	448.0
Moneyness	1.06	0.27	0.82	0.93	1.03	1.15	1.32
Implied volatility	47.01	24.26	23.23	29.99	40.98	57.61	78.85
Absolute Delta	0.51	0.24	0.18	0.32	0.52	0.71	0.84
Panel C: Put Options (N=5,576,091)							
Delta-Hedged Return	-0.13	4.59	-4.0	-1.87	-0.39	1.04	3.7
Days to Maturity	167.82	176.34	21.0	50.0	112.0	204.0	417.0
Moneyness	0.98	0.35	0.72	0.85	0.96	1.06	1.21
Implied volatility	48.46	27.1	23.74	30.44	41.41	58.48	81.6
Absolute Delta	0.4	0.24	0.1	0.19	0.36	0.57	0.77
Panel D: All Options 1996-2002 (N=2,026,636)							
Delta-Hedged Return	0.48	7.45	-5.67	-2.28	-0.11	2.39	7.04
Days to Maturity	152.5	174.2	22.0	50.0	108.0	176.0	330.0
Moneyness	1.08	0.44	0.78	0.9	1.02	1.16	1.4
Implied volatility	65.16	29.18	32.17	42.9	60.6	82.14	103.56
Absolute Delta	0.49	0.23	0.18	0.3	0.48	0.68	0.82
Panel E: All Options 2003-2020 (N=10,103,208)							
Delta-Hedged Return	-0.1	4.58	-4.14	-1.89	-0.37	1.12	3.97
Days to Maturity	176.2	180.21	21.0	50.0	113.0	206.0	448.0
Moneyness	1.01	0.28	0.77	0.89	0.99	1.1	1.26
Implied volatility	44.16	23.31	22.69	28.88	38.71	52.8	71.22
Absolute Delta	0.45	0.25	0.12	0.24	0.44	0.65	0.81

Table 1: Delta-Hedged Option Return Summary Statistics

The table reports descriptive statistics for delta-hedged option returns. Panel A reports statistics of returns and option characteristics over the period from 1996 to 2020. Delta-hedged option returns are measured over a period of one calendar month, or until option maturity. Delta-hedging is performed daily. Days-to-maturity is defined as the number of calendar days until option expiration. Moneyness is the ratio between the underlying's stock price and the option's strike price. Option implied volatility is provided by OptionMetrics. Absolute delta follows the model of Black and Scholes (1973). Panels B and C depict statistics for call and put options, respectively, whereas panel D reports summary statistics for the period 1996 to 2002 used exclusively in the training step of the models. Panel E reports summary statistics for the entire out-of-sample period, comprising the years 2003 through 2020.

the underlying's close price (Eisdorfer et al., 2020), and realized volatility (Cao, Vasquez,

Xiao, and Zhan, 2019).[12]

We are left with a total of 270 characteristics, which can be broadly classified into 12 groups. *Accruals, Industry, Investment, Profitability, Quality* and *Value* exclusively include stock-based characteristics and loosely follow the classification proposed in Jensen, Kelly, and Pedersen (2021) and Green et al. (2017). Classification group *Contract* contains five option-based characteristics, the time-to-maturity, moneyness, implied volatility, and put and call identifier, and thus combines information about the location of the respective option on the underlying's implied volatility surface. We therefore assume that these characteristics play a pivotal role in predicting future option returns as a sort of reference point, much as the market return in traditional factor models (Fama and French, 1993). Groups *Frictions, Illiquidity, Informed Trading, Past Prices* and *Risk* contain both stock- and option-based information and highlight the predictions' dependence on informational frictions. Appendix IA4 provides a detailed list of all 270 characteristics, their origin from the literature, the primary information source (option- vs. stock-based) and the feature group we have assigned them to.

5. Predicting Option Returns

5.1. Predictability Comparison

Figure 1 shows the out-of-sample R^2 for the pooled testing sample from January 2003 through December 2020 using the nine machine learning methods and two ensembles outlined in Section 3. Nonlinear models routinely outperform the predictability uncovered by linear models for option returns. The R^2_{OS} for none of the linear models is significantly positive, ranging from -0.61% for PCR to -0.12% for Lasso regressions. The Clark and West (2007) test statistics indicate that none of these predictions outperforms a naive benchmark of predicting delta-hedged returns of zero.[13] Nonlinear models, in contrast, manage to uncover substantial predictability in single-equity option returns. GBR and Dart generate the highest out-of-sample R^2, above 2.5% for the pooled sample. All but FFN generate forecasts statistically better than the naive benchmark.

[12]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).

[13]We find that our results remain intact when the models' relative performance is evaluated against naive benchmarks predicting zero excess returns or the historical mean, either for the entire sample or subsamples by the respective bucket to which an option belongs at the time of the investment.

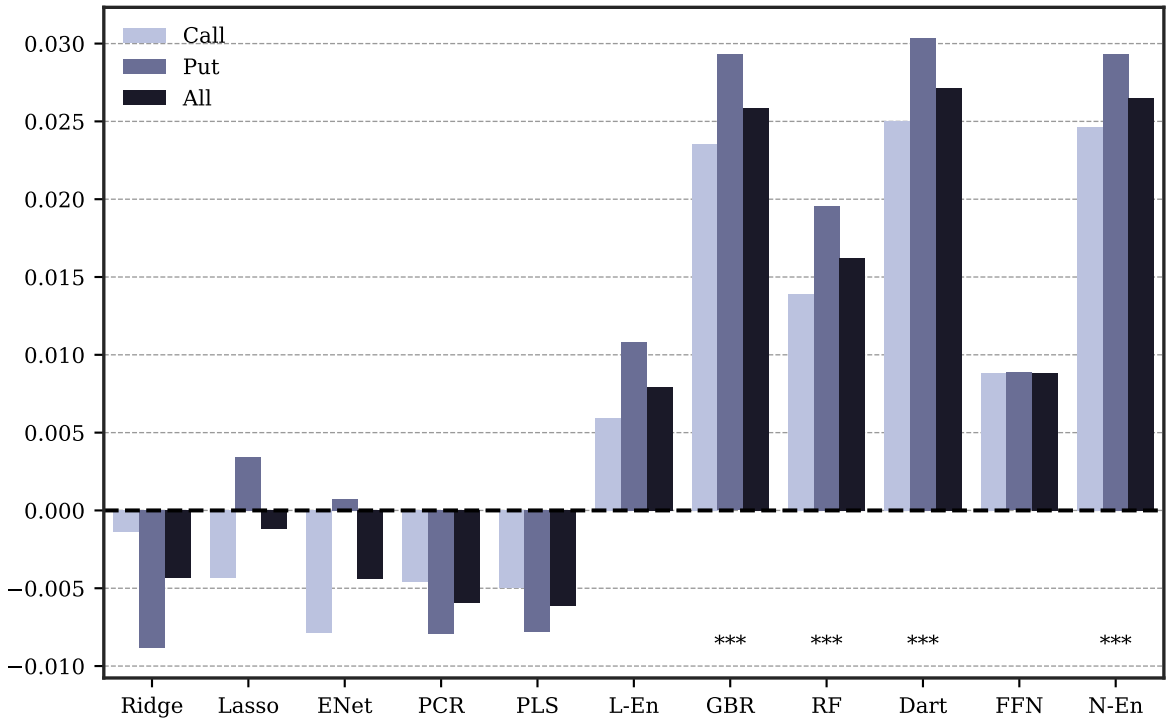

Fig. 1. R^2_{OS} Model Comparison

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.

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 Dart, with $R^2_{OS} > 2.5\%$, the worst call predictions are generated by PLS, and the worst put predictions are made by PCR. Linear dimensionality-reduction regressions do not appear to adequately uncover an adequate relationship between characteristics and future returns. While FFN was the most promising model in Gu et al. (2020) for stock returns, we find that it uncovers low predictability in the case of delta-hedged option returns. Given that FFN is the method with the highest potential complexity, this suggests that this complex structure does not generalize well to the testing sample of option returns. In contrast, tree-based methods tend to outperform, suggesting that histogram-based estimation including *nonlinear interactions* trumps model complexity in this market.

Ensembles have been shown to improve the accuracy and consistency of the predictions

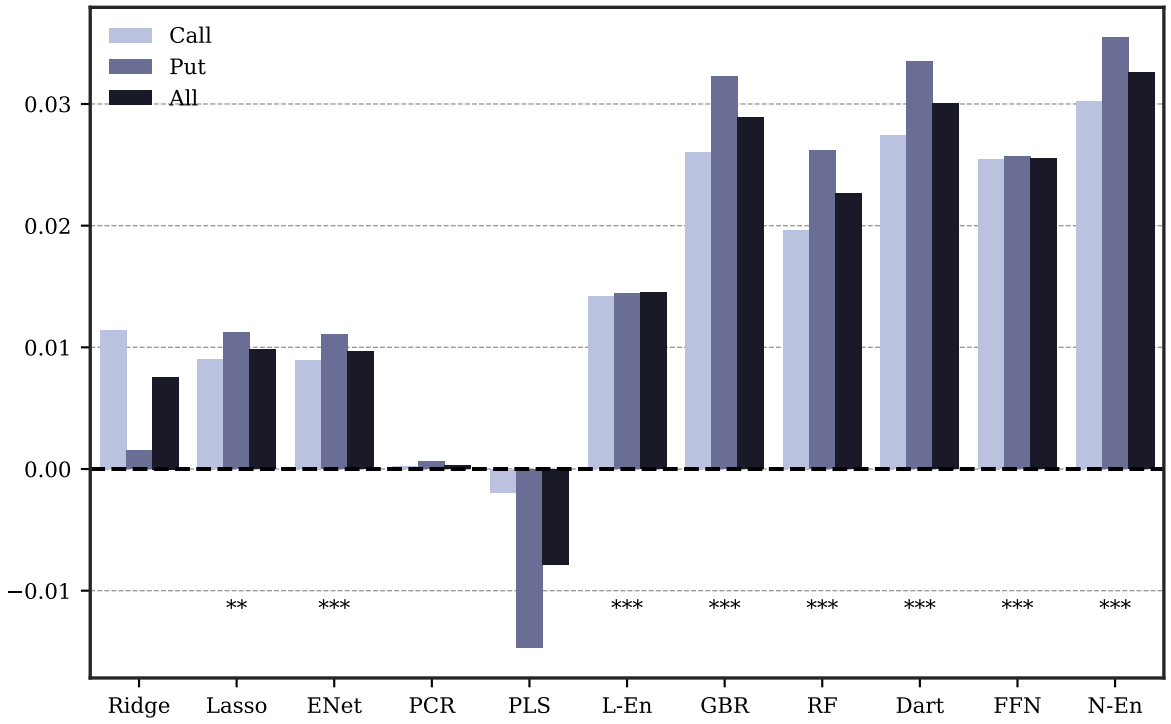

Fig. 2. $R^2_{OS;XS}$ Model Comparison

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.

and are a staple in modern machine learning estimation (e.g., Krizhevsky, Sutskever, and Hinton, 2012; Lakshminarayanan, Pritzel, and Blundell, 2016).[14] 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 also not statistically significant. The level of predictability at roughly 0.8% is about twice as high as the levels of predictability of stock returns found by Gu et al. (2020). Within the class of nonlinear methods, Dart tends to outperform N-En for the full sample, as well as separately for puts and calls. Just as all tree-based methods, the resulting predictions comfortably beat the naive benchmark of zero excess returns.

We are ultimately interested in how far the models uncover cross-sectional return spreads in our option sample. For this, Figure 2 compares the cross-sectional $R^2_{OS;XS}$ de-

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

fined in Equation (5). Interestingly, while none of the individual linear models generated predictability on the full sample, especially penalized regressions are able to generate realistic return spreads. For Lasso and ENet these are even statistically significant at the 5% and 1% level, respectively. The cross-sectional pricing power of all nonlinear models is highly significant, and surprisingly FFN manages to produce $R^2_{OS;XS}$ comparable to the tree-based methods, despite failing to adequately predict the average level of future returns. This exercise clearly highlights the benefits of using ensemble models, in that the cross-sectional predictive power increases both for L-En and N-En. The nonlinear ensemble generates the highest cross-sectional predictability for all subsamples considered with an $R^2_{OS;XS}$ of 3.3%.

Most studies up-to-date overlook the ability to zoom in on the predictive power of the models considered and provide an intuition of how stable the resulting predictions are. The focus mostly lies on pooled out-of-sample predictability as in Figure 1 and Figure 2. Instead, we also show the dispersion of R^2_{OS} and $R^2_{OS;XS}$ over time. While the pooled approach weights the predictions of each year by the number of option-month observations contained, we now provide estimates of the predictive power per year, which allows us to investigate the stability of the forecasts over time. Stable predictability is imperative for investors who wish to use the model forecasts in their investment decision-making process.

The upper panel of Figure 3 adds three main points to the pooled R^2_{OS} consideration above. First, the predictive power of all models fluctuates significantly over time. Linear models produce the largest dispersion, with the possibility of very large 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. A notable exception from this is Dart, which manages to produce the best predictability most consistently, with an interquartile range of $1\% - 4.4\%$. Leveraging the informational content uncovered by Dart may thus grant large benefits to investors. Following this intuition, we lastly find that ensemble models produce more stable forecasts. The minimum to maximum predictability for L-En ranges from -8% to $+3.9\%$ and for N-En between -2.7% and $+5.7\%$. The forecasts of N-En are the most stable over time, producing significant predictive power for all years in the testing sample. Cross-sectional predictability is much less disperse (lower panel of Figure 3) and is generally increasing in the model complexity and the ability to model nonlinear interaction effects between characteristics. N-En is once more the most stable model with the highest median and mean level of predictability over time.

Forecast comparison. We now turn to Diebold and Mariano (1995) tests to itera-

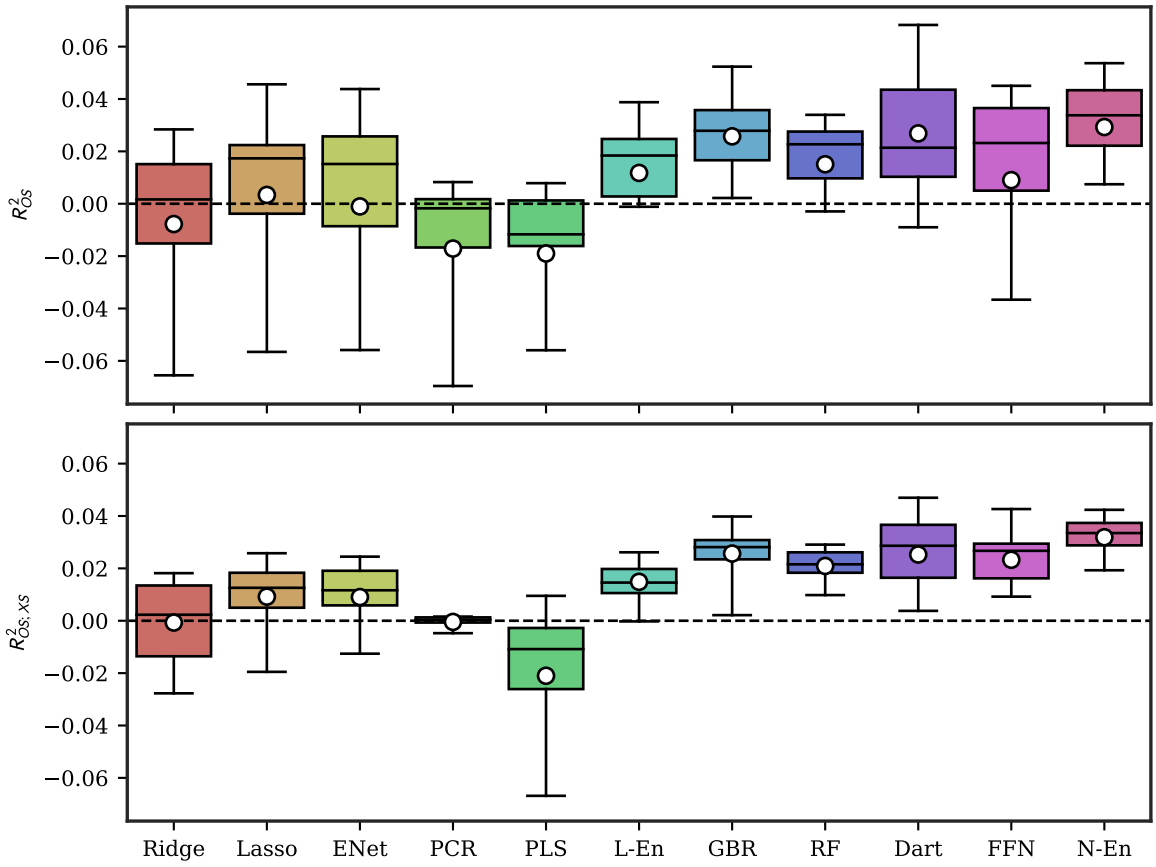

Fig. 3. Time-Series Dispersion of R^2_{OS} and $R^2_{OS;XS}$

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).

tively 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.

Within the class of nonlinear models, only GBR manages to outperform L-En comfortably with a t-statistic of 2.09. Interestingly, the forecasts of GBR and Dart, which have produced the highest single-model predictability, are indistinguishable from one an-

Panel A: Diebold and Mariano (1995) Forecast Comparison

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	1.76	0.27	−0.52	−0.56	**3.64**	**3.12**	**2.11**	**2.52**	**3.24**	**3.76**
Lasso		−1.24	−0.94	−0.98	**2.13**	**2.31**	1.42	1.93	1.85	**2.79**
ENet			−0.48	−0.51	1.94	**2.15**	1.43	1.84	**2.90**	**2.49**
PCR				−0.19	**2.60**	**7.17**	**7.03**	**5.49**	1.59	**8.08**
PLS					**2.61**	**7.78**	**8.02**	**5.71**	1.61	**8.64**
L-En						**2.09**	0.79	1.62	−0.42	**2.86**
GBR							**-4.40**	0.40	−1.60	1.04
RF								**2.52**	−0.72	**6.27**
Dart									−1.34	0.05
FFN										1.94

Panel B: Forecast Correlation

	Lasso	ENet	PCR	PLS	L-En	GBR	RF	Dart	FFN	N-En
Ridge	**0.84**	**0.83**	−0.10	0.53	**0.91**	0.51	0.44	0.50	**0.78**	0.64
Lasso		**0.95**	−0.14	0.51	**0.92**	0.45	0.41	0.44	**0.79**	0.59
ENet			−0.16	0.47	**0.90**	0.42	0.39	0.40	**0.79**	0.57
PCR				0.18	0.02	0.12	0.23	0.10	−0.06	0.10
PLS					**0.75**	0.52	0.53	0.51	0.57	0.60
L-En						0.56	0.53	0.55	**0.83**	0.70
GBR							**0.86**	**0.92**	0.61	**0.95**
RF								**0.79**	0.54	**0.88**
Dart									0.61	**0.94**
FFN										**0.79**

Table 2: Forecast Comparison

Panel A of the table shows Diebold and Mariano (1995) test statistics following Equation (8) for the nine models and two ensembles considered in the paper. A positive number indicates that the model in the column outperforms the row model. If it is highlighted in light blue (blue), this outperformance is statistically significant at the 1% level (5% level). Panel B shows forecast correlations defined in Equation (10). Here, highlighting in light blue (blue) denotes large values with a cutoff at 90% (70%).

other. FFN manages to beat Ridge and ENet, but none of the nonlinear models.[15] The nonlinear ensemble N-En produces forecasts that beat any of the other models but Dart and GBR, which are its most vital inputs. Comparing the performance of L-En and N-En, we find that the forecasts of the latter are more informative with a t-statistic of 2.86. While the forecasts of GBR and Dart do not manage to statistically beat those of FFN, N-En manages to do so at the 10% level (t-stat = 1.94).

The forecast correlation analysis in Panel B of Table 2 confirms these insights. First,

[15]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 IA6.1 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.

penalized linear regression methods yield similar predictions, which is expected given that Lasso and Ridge are special cases of ENet. Notably, L-En produces forecasts highly correlated to either of these methods ($\rho > 90\%$), as does FFN ($\rho > 70\%$), suggesting that the underlying process proposed by FFN is quite similar to a regularized linear function. Second, tree-based methods are a class of their own, showing correlations of $\rho > 70\%$ only among themselves. Given their unique setup of identifying quantile splits in the input characteristics to relate to option returns, this is not surprising. At the same time, these methods, especially boosted tree-based methods (GBR and Dart), outperform all other models. Consequently, predictions by N-En share many of their properties ($\rho > 90\%$). However, the ensemble predictions are also highly correlated to the two other nonlinear methods with $\rho^{\text{N-En,FFN}} = 79\%$ and $\rho^{\text{N-En,RF}} = 88\%$.

Impact of Nonlinearities. The results highlight the usefulness of forming forecast ensembles of many models. Therefore, from now on, we will compare the linear (L-En) and nonlinear (N-En) ensemble methods to understand the implications of nonlinear interaction effects.

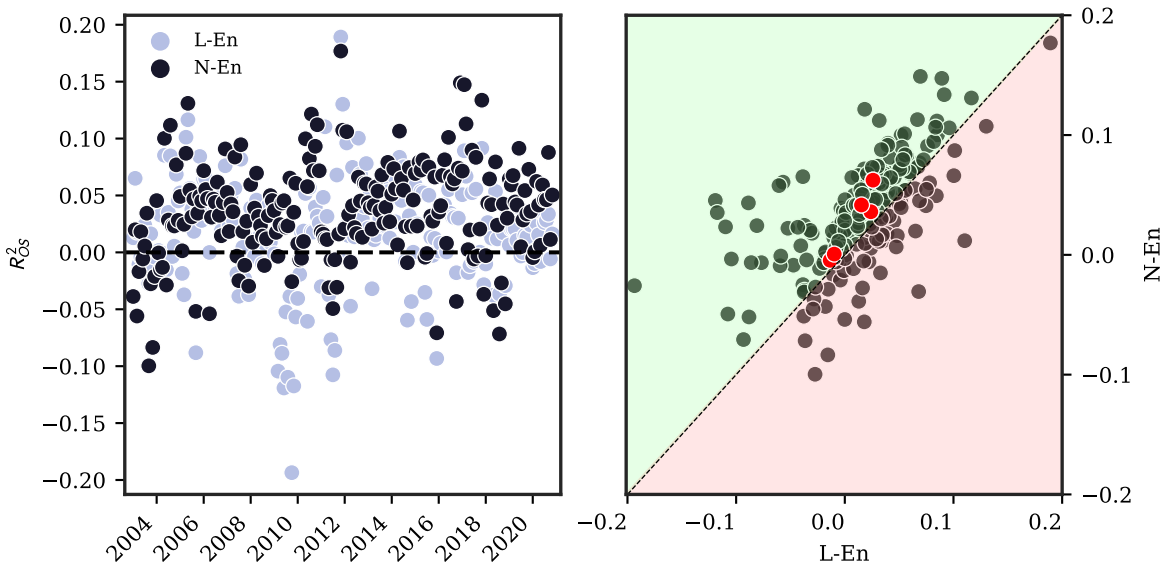

Fig. 4. Comparing Linear and Nonlinear Ensembles

The left panel of the figure shows monthly R^2_{OS} for the testing sample from 2003 through 2020 for the linear (L-En) and nonlinear (N-En) ensembles. The right panel compares the two by showing the resulting R^2_{OS} for L-En on the x-axis and for N-En and the y axis. The green-shaded area represents a relative outperformance in terms of predictability for N-En, while the red-shaded area represents the opposite. The five red circles represent the Coronavirus selloff from December 2019 through April 2020.

In this section, we highlight the usefulness of modeling nonlinear interaction effects

between option characteristics. The left panel of Figure 4 compares the monthly R^2_{OS} for N-En in dark blue and for L-En in light blue. The figure provides multiple insights: First, both ensembles tend to beat a naive benchmark of predicting zero delta-hedged returns in most months. Second, N-En is less prone to experience predictability crashes. While L-En does a poor job of predicting future option returns during and after turbulent times (2008-2012 for example), N-En has predicted returns much more consistently. Third, for N-En, the predictive power has stayed roughly constant over time, confirming not only that nonlinear models leverage the information embedded in option characteristics, but also that the resulting predictability patterns are highly persistent over time.[16] At the same time, this suggests that our methods pick up more than just plain mispricing, especially in recent years, under the assumption that modern financial markets, and in particular the options market, has become informationally more efficient.

The panel on the right shows a scatter plot of the resulting R^2_{OS} of L-En on the x-axis and N-En on the y-axis. The green-shaded area represents a relative outperformance of the nonlinear ensemble, whereas the red-shaded area represents the opposite. For 71.2% of the months in our sample, we find that N-En outperforms L-En and if so, quite comfortably. In the figure, we have also highlighted in red circles the period of December 2019 through April 2020, which are the months surrounding the Coronavirus selloff in February and March 2020. Nonlinear ensembles were better able to deal with this huge exogenous shock, beating L-En in all of the five months. The R^2_{OS} for N-En dips slightly below zero only in January and reaches pre-crisis levels of 3 to 6% in March. While the pandemic-driven selloff constituted a large exogenous shock to financial markets, the relationship between option characteristics and future returns quickly went back to normalcy after the initial reaction. This speaks in favor of using nonlinear machine learning methods to identify persistent predictability patterns in the options market, especially given that the sample we used to train the models to uncover predictability patterns in option returns during the Coronavirus selloff ended in December of 2017. In Figure IA6.1 we repeat the exercise using the cross-sectional out-of-sample $R^2_{OS;XS}$ defined in Equation (5). N-En beats the linear ensemble in 85.6% of the months and predictability does not dip below 1.5% of variation of return spreads explained during the Coronavirus selloff.

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

5.2. Machine Learning Portfolios

To gauge if the predictability of machine learning methods is also economically significant, we follow Gu et al. (2020) and form portfolios using machine learning forecasts. Specifically, each month, we sort individual equity options into 10 decile portfolios based on the machine learning models' (L-En and N-En) expected return forecasts. Then, we calculate the one-month-ahead average realized return of individual equity options in each decile. Finally, we compute the average long-short portfolio return of a zero-net investment portfolio by buying options with the highest expected return forecast (decile 10) and financing this investment by writing options with the lowest expected return forecast (decile 1).

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.223	−1.395	1.646	−0.848	−1.720	−1.841	1.950	−0.944	***
2	−0.724	−0.678	1.847	−0.367	−0.715	−0.650	1.848	−0.352	
3	−0.492	−0.379	1.851	−0.205	−0.401	−0.350	1.741	−0.201	
4	−0.322	−0.220	1.845	−0.119	−0.224	−0.197	1.655	−0.119	
5	−0.178	−0.106	1.829	−0.058	−0.099	−0.093	1.644	−0.057	
6	−0.047	−0.035	1.795	−0.019	0.007	−0.029	1.630	−0.018	
7	0.083	0.048	1.798	0.027	0.111	0.042	1.662	0.025	
8	0.220	0.114	1.794	0.064	0.229	0.139	1.767	0.078	
9	0.387	0.220	1.807	0.122	0.394	0.286	1.923	0.149	
Hi	0.711	0.522	1.936	0.270	0.835	0.786	2.286	0.344	
H-L	1.934	1.917	1.237	1.550	2.555	2.627	1.274	2.062	***
		(14.19)		(6.54)		(15.79)		(8.14)	

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. 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.

Table 3 shows the average predicted and one-month-ahead realized portfolio returns. For both ensemble classes, the average predictions are comparable to the returns that 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 2.63% with a Sharpe ratio of 2.06. The realized payoff is also close to the predictions made by N-En (2.55% per month, on average). To validate the usefulness of allowing for

nonlinearities when predicting future option returns, we perform a comparison of realized return spreads in the last column of Table 3, which highlights that N-En is better than L-En at identifying particularly overpriced options. This translates into the large H-L spread between the two models based on the difference-in-differences (diff-in-diff) test results; i.e., the diff-in-diff return spread generated by N-En vs. L-En is economically and statistically significant at 0.71% (= 2.63% – 1.92%) per month, showing that the cross-sectional variations in option returns are more accurately captured by the nonlinear models.[17]

	Lo	2	3	4	5	6	7	8	9	Hi
N^U	634.65	797.43	820.74	808.51	794.23	784.62	769.00	738.25	672.45	519.95
m	1.03	1.00	1.00	1.00	1.00	1.01	1.01	1.02	1.02	1.05
ttm	124.66	156.54	172.37	178.87	182.27	184.51	184.72	185.19	187.54	205.52
$\hat{r} > 0$	0.31	0.35	0.38	0.39	0.41	0.42	0.44	0.45	0.47	0.52
$s(r) = s(\hat{r})$	0.69	0.65	0.62	0.58	0.55	0.53	0.48	0.47	0.48	0.52
% Calls	0.43	0.42	0.44	0.46	0.47	0.51	0.55	0.60	0.65	0.75
Spread	0.20	0.16	0.13	0.12	0.11	0.11	0.11	0.12	0.13	0.15
OI	0.06	0.08	0.10	0.12	0.12	0.12	0.11	0.11	0.09	0.08
delta	0.02	0.02	0.03	0.03	0.03	0.05	0.08	0.13	0.17	0.27
gamma	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
vega	0.15	0.16	0.18	0.19	0.20	0.20	0.20	0.20	0.20	0.21
theta	−0.31	−0.19	−0.15	−0.13	−0.12	−0.11	−0.11	−0.11	−0.12	−0.12

Table 4: N-En Portfolio Decomposition

The table shows summary statistics for the ten machine learning portfolios following the nonlinear ensemble N-En. All measures are averaged over time. N^U denotes the number of individual stocks underlying the options in the portfolios, m the moneyness, ttm the time-to-maturity. $\hat{r} > 0$ denotes the share of positive returns in the portfolio and $s(r) = s(\hat{r})$ the share for which N-En correctly predicted the realized return's sign. % Calls is the share of call options in the portfolio, Spread the average option bid-ask spread, OI the relative open interest and *delta*, *gamma*, *vega*, and *theta* the respective option Greeks. *gamma* is expressed for a 1% move in the underlying stock ($gamma \times \frac{S}{100}$) and *vega* and *theta* in terms of the underlying stock price ($\frac{x}{S}$ for $x \in [vega, theta]$).

Table 4 provides summary statistics for the decile portfolios based on the nonlinear ensemble N-En. The high and low portfolio tend to include options from $N^U = 520$ and $N^U = 634$ underlyings, respectively. The portfolios use options from the least number of individual stocks, but still rely on average on options from about a third of the stocks included in the sample. While we find little change in the average moneyness of the included options, the Low portfolio includes more short-term and fewer call options (% Calls = 0.43), while the High portfolio includes more long-term call options (% Calls

[17]We have repeated this exercise using dollar open interest of the respective options as weights. The results are in Table IA7.1 and confirm the findings discussed here.

= 0.75). The relative bid ask spreads of the options are also highest for the extreme portfolios, while the outstanding open interest is lowest for options in these portfolios. The *delta* (of the unhedged option), *theta* and *vega* are increasing in the expected option return, while *gamma* is roughly comparable across portfolios. Considering a long position of the respective option portfolios, we find that the probability of a positive payoff ($\hat{r} > 0$) increases monotonically in the expected return. While only 31% of the options included in Low have a positive return in month $t + 1$, 52% of the options in High have. Focusing on whether the predicted return direction is correct, $s(r) = s(\hat{r})$, we find it to be highest for Low with 69% of the return directions correctly predicted. The ratio first decreases and then starts to increase once more for portfolios 9 and 10. For portfolio 10 (High), it stands at 52%. Figure IA7.1 to Figure IA7.3 in the Online Appendix provide a visual representation of the portfolio decomposition, split by the put and call contracts included.

Figure IA7.4 shows the relative share of call options in the machine learning portfolios for each year in our testing sample. We find that it increased after the financial crisis. Over time, the call share of the High portfolio has decreased, while it stayed roughly constant for Low. Another interesting aspect of the machine learning portfolios is to analyze in how far N-En places all options trading on a given stock in the same bucket, or whether it distinguishes between over- and under-priced contracts *within* a single underlying. For this, Figure IA7.5 shows the share of options of a given underlying that ended up in the same portfolio. We find that this measure is highest for the High and Low portfolios, but that it is mostly below 50%. This suggests that N-En does not simply identify underlying stocks for which all options are considered mispriced, but also picks up on the variation *within* the option chain of a given stock.

How persistent is our trading strategy, i.e., 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 potentially change rapidly, we can provide indicative evidence for this on the options-bucket level (see Section 4 for a definition). We focus on how the portfolio mode for all stock-bucket combinations in our sample changes from one month to another.[18] With this, we can understand more about the persistence of the machine learning predictions. Figure IA7.6 provides the results to this exercise: we first note that the diagonal is the lightest-shaded area, highlighting that transitioning from one portfolio to the same or a neighboring is most likely. Second, the lightest areas overall are at the

[18]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.

High and especially the Low portfolio. If options of a given stock-bucket combination are classified as being highly overpriced (Low portfolio), they tend to stay in this portfolio in the next months. The same applies to a lesser degree for the most underpriced stocks (High portfolio).

In a next step, we investigate in how far the profitability of our machine learning portfolios changes with the state of the economy, and whether N-En manages to beat L-En regardless of this state. The results in Table IA7.2 shows that payoffs are generally amplified during bad economic states and that N-En manages to beat L-En in all market phases. This holds for a wide range of measures which approximate the economic state.

Table IA7.3 in the Appendix provides the H-L results for different option buckets. Predictability is concentrated in short-term options, for which we also find the highest payoffs when following the investments proposed by our machine learning models. Throughout all option buckets we find higher raw (Avg) and risk-adjusted (SR) returns for N-En than for L-En. This difference is consistently, highly statistically significant.

5.2.1. Risk Attribution

One possible explanation for these results is that the machine learning models are best at predicting the most risky option positions, which should translate to higher future realized returns. To understand whether this is the case, we compute risk-adjusted returns for the H-L portfolios for N-En, either for the pooled sample of all options, or split by option buckets. We consider a wide range of candidate models, which have been proposed by earlier studies to explain the returns of a variety of financial instruments. To date, we are still lacking a concise factor model for option returns. Therefore, we resort to the most prominent models from other asset classes, including the CAPM, the Fama and French (2015) five factor model enhanced by the momentum factor of Carhart (1997) (FF6, Fama and French, 2018) and additionally by the liquidity factor of Pástor and Stambaugh (2003) (FF6+PS), a model using the Agarwal and Naik (2004) option-market factors (AN), a model with the intermediary leverage bearing capacity of Grünthaler et al. (2020), and finally the factor model on optionable stocks proposed by Bali and Murray (2020) (BM).

The results are provided in Table 5. Raw and risk-adjusted returns for all option types are virtually the same. The risk exposure picked up by the candidate models does not suffice to explain the return spreads generated by the nonlinear machine learning methods.

			CAPM		FF6		FF6+PS	
Full Sample			2.715	(16.64)	2.638	(22.38)	2.636	(21.31)
					Buckets			
$\tau \leq 90$	atm		1.961	(21.23)	1.916	(27.75)	1.915	(27.25)
	itm	C	2.359	(16.26)	2.290	(19.10)	2.289	(18.85)
	itm	P	2.196	(10.48)	2.147	(10.52)	2.148	(10.61)
	otm	C	4.025	(13.95)	3.961	(18.36)	3.957	(17.92)
	otm	P	4.416	(17.76)	4.322	(19.27)	4.322	(19.08)
$\tau > 90$	atm		1.963	(15.77)	1.895	(22.42)	1.893	(21.96)
	itm	C	2.841	(13.80)	2.754	(16.18)	2.754	(16.20)
	itm	P	1.360	(11.51)	1.343	(14.19)	1.341	(13.52)
	otm	C	3.506	(13.62)	3.445	(17.48)	3.442	(17.09)
	otm	P	2.776	(14.44)	2.728	(15.76)	2.724	(14.96)

			AN		BM		LBC	
Full Sample			2.629	(18.27)	2.513	(31.28)	2.627	(16.84)
					Buckets			
$\tau \leq 90$	atm		1.911	(23.67)	1.887	(29.40)	1.935	(23.70)
	itm	C	2.238	(17.49)	2.174	(18.33)	2.277	(19.38)
	itm	P	2.158	(11.86)	1.984	(8.77)	2.074	(10.33)
	otm	C	3.881	(16.35)	3.775	(15.78)	3.816	(15.13)
	otm	P	4.448	(18.27)	4.058	(21.85)	4.300	(17.33)
$\tau > 90$	atm		1.881	(14.79)	1.808	(20.24)	1.939	(15.41)
	itm	C	2.766	(12.91)	2.557	(15.27)	2.797	(13.14)
	itm	P	1.263	(15.12)	1.188	(12.10)	1.256	(13.84)
	otm	C	3.419	(13.09)	3.274	(16.77)	3.382	(12.31)
	otm	P	2.821	(14.45)	2.690	(20.63)	2.859	(14.82)

Table 5: Common Factor Models and Machine Learning Predictions

The table shows the risk-adjusted returns of the high-minus-low portfolio following the predictions by N-En using risk factor models proposed in the literature. Risk-adjusted returns are provided for the CAPM, the Fama and French (2015) 5-factor model plus momentum (Carhart, 1997), FF6; the Fama and French (2015) 5-factor model plus momentum and the liquidity factor of Pástor and Stambaugh (2003), FF6+PS; the model following Agarwal and Naik (2004) including the returns of at-the-money and out-of-the-money index options plus the market factor, AN; the model for optionable stocks by Bali and Murray (2020), including the spread between implied and realized volatility by Bali and Hovakimian (2009), its difference through time by An et al. (2014), the call-minus-put implied volatility spread by Cremers and Weinbaum (2010), and the market factor, BM; and a model including the market factor and the leverage bearing capacity of financial intermediaries proposed in Grünthaler et al. (2020), LBC. We also provide the information for option buckets defined in Section 4.

5.2.2. Impact of Transaction Costs

So far we have assumed that the investor can buy and sell each option at mid-point between bid and ask – that is with zero effective spreads. Prior research has shown that transaction costs in option markets can be large (Ofek et al., 2004). We now turn to the impact of trading at different transaction prices by changing the ratio of effective to quoted spreads. For that, Table 6 shows the full-sample payoffs to the high-minus-low trading strategy assuming that investors have to pay an effective spread of 15% or 25% of the quoted bid-ask spread. As a simple example, with an effective spread of 15%, we would buy an option with $bid = \$10; ask = \$15$ at \$12.875 and sell it at \$12.125. We consider H-L payoffs, such that we go long options in decile 10 and short options in decile 1. For options that expire after our holding period of one month, we consider the effective spreads for trading in and out of the position, for options that expire within the holding period, we only do so for engaging into the position and cash in the payoff at maturity.

			Eff. Spread = 15			Eff. Spread = 25		
			H-L	t	SR	H-L	t	SR
Full Sample			1.334	(10.97)	1.171	0.474	(4.41)	0.431
					Buckets			
$\tau \leq 90$	atm		1.381	(17.49)	1.834	1.007	(13.92)	1.384
	itm	C	1.198	(10.46)	1.073	0.437	(3.29)	0.381
	itm	P	1.564	(9.50)	1.284	1.101	(8.27)	0.950
	otm	C	2.836	(10.22)	1.118	2.057	(7.65)	0.824
	otm	P	3.285	(12.18)	1.214	2.725	(10.36)	1.025
$\tau > 90$	atm		0.604	(6.31)	0.675	−0.291	(−2.44)	−0.311
	itm	C	0.132	(0.94)	0.110	−1.624	(−6.50)	−1.072
	itm	P	0.218	(2.56)	0.232	−0.569	(−4.75)	−0.556
	otm	C	1.051	(5.33)	0.575	−0.544	(−2.61)	−0.295
	otm	P	1.153	(6.69)	0.655	0.142	(0.79)	0.080

Table 6: Machine Learning Predictions and Effective Spreads

The table shows the realized returns of the high-minus-low portfolio following the predictions by N-En after accounting for transaction costs through effective spreads, which we define as a fraction of the quoted spreads provided OptionMetrics. H-L denotes the average returns controlling for spreads of 15% and 25%, t denotes the corresponding t-statistic and SR the resulting Sharpe ratio. We also provide the information for option buckets as defined in Section 4.

The strategy using all available option contracts remains highly profitable for 15% and 25% effective spreads. The average monthly returns are 1.33% or 0.47% (t-statistic: 10.97 vs. 4.41) with Sharpe ratios of 1.17 and 0.43. If we increase the effective spreads further to 50%, the unconditional strategy fails to generate positive returns. Consistent

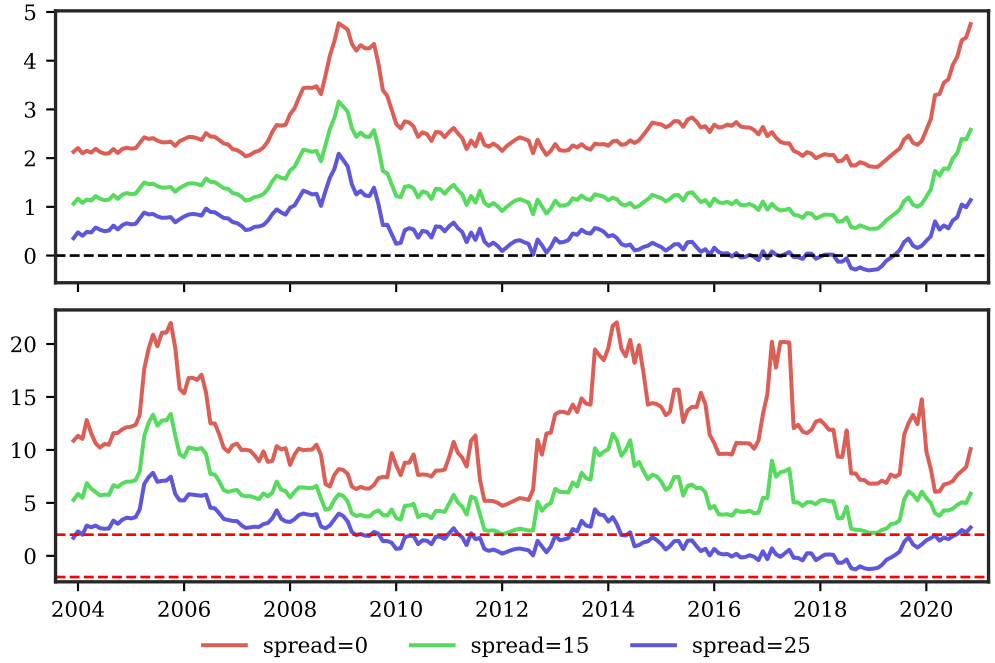

Fig. 5. Rolling One-Year Returns and Effective Spreads

The upper panel in this figure shows rolling cumulative returns over twelve months for the high-minus-low portfolio following the predictions made by N-En using all available option contracts. The lower panel shows corresponding rolling Newey and West (1987) t-statistics with a lag of twelve months. We compare the resulting profitability for zero effective spreads (trading at the mid price), as well as effective spreads of 15% and 25%.

with the evidence for the various options under the assumption of zero effective spreads in Table IA7.3, the profitability of short-term options is generally amplified. For both levels of effective spreads, short-term options generate significantly positive returns, which are the highest in absolute terms for out-of-the-money options, and highest in risk-adjusted terms for at-the-money options (SR: 1.78 and 1.34).[19] Contrarily, all returns for long-term options turn insignificant or even negative for the effective spreads of 25%. While long-term options make up a larger portion of our sample, their value seems to be distorted by amplified spreads, such that the same signals that generate significantly positive returns for short-term options fail to do so for options with maturity of more than three months.

The evidence in Table 6 deals with average returns and profitability of the resulting trading strategies for the entire sample of 215 months. To understand how the predictability we uncover through nonlinear machine learning methods is distributed over time, we

[19]For otm short-term puts, we retain significant profitability even at 50% effective spreads, for atm short-term options and otm calls payoffs are positive, but not significantly different from zero.

show rolling returns for the last 12 months in the upper panel of Figure 5. The lower panel depicts the statistical significance, as measured by rolling Newey and West (1987) t-statistics with 12 lags. The high-minus-low portfolio shows the largest payoffs during turbulent times. Annual returns are amplified during and after the financial crisis in 2008-2010, remain high surrounding the Taper Tantrum in 2013 and increase once more during the Coronavirus selloff at the beginning of 2020. The realized returns are highly significant at all times, with t-statistics regularly around or above 10. Assuming effective spreads of 15% significantly lowers the generated returns, but they remain consistently positive over time. The same applies when increasing effective spreads to 25%, with the notable exception of the years 2016 to 2019, where returns are zero or turn slightly negative. Whereas returns are significantly different from zero for effective spreads of 15% throughout the whole sample period, the same is only observed for the beginning of the sample in case spreads are 25%.

We provide the same results split by option buckets in Figure IA7.8 (t-statistics in Figure IA7.9). In contrast to using all available option contracts, investing in short-term at-the-money and out-of-the-money as well as in-the-money put options generates positive returns in all months and for both 15% and 25% effective spreads, which are statistically significant in most months. Long-term options tend to suffer more severely from imposing transaction costs, but tend to survive 15% effective spreads for the first half of our sample. Muravyev and Pearson (2020) argue that effective spreads are much lower than the ones we measure here using OptionMetrics' end-of-day quoted spreads, since investors have the ability to time trades over the entire trading day. Using high-frequency data on single stock options, the authors find that effective transaction costs are about a quarter smaller than their conventional estimates. Moreover, consider the impact of execution timing, the reduction can be increased to more than 60%. We are therefore confident that the signals N-En generates are valuable and require further discussions on the sources of these highly profitable trading strategies.

5.3. Which Characteristics Matter?

We next analyze the importance of option characteristics groups. Since we use a large number of characteristics (a total of 270) in our estimation, we look at the relative importance of various characteristics groups, following Lundberg and Lee (2017).

Optimally, we would re-estimate the model after excluding the characteristics in each group sequentially. This approach, however, is infeasible, given the large computational

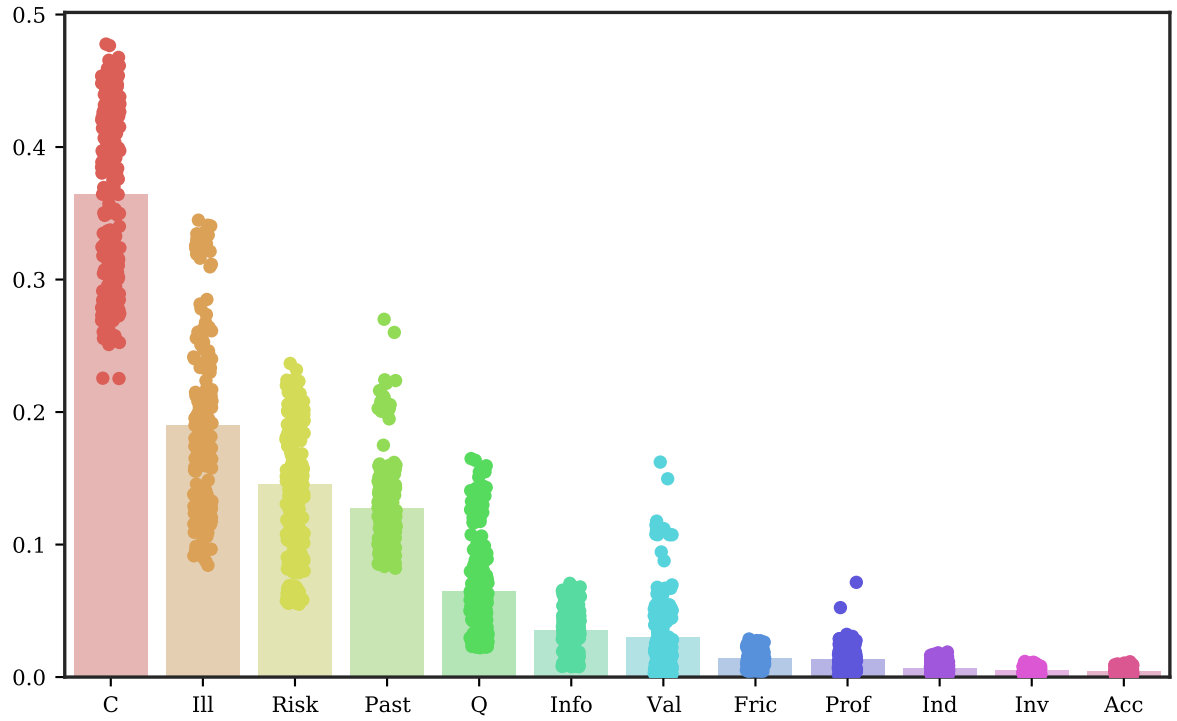

Fig. 6. Feature Group Importance for Nonlinear Forecast Ensemble

The figure shows the feature group importance for the twelve feature groups defined in Appendix IA4 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.

burden required by multiple estimations of the models considered.[20] Instead, Lundberg and Lee (2017) propose the use of SHAPs (SHapley Additive exPlanations), which approximate the effect of this feature exclusion and are based on cooperative game theory.[21]

The relative feature group importance for N-En is provided in Figure 6, in which the groups are sorted by their total importance over the entire testing sample. It is directly apparent that 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

[20]We use this approach in Section 5.4 and re-estimate each model for three subsamples of the input characteristics.

[21]The authors have created a Python package: `https://github.com/slundberg/shap`.

essential when making option return forecasts. Measures about illiquidity and risk are the next-most important predictors. Included are, for example, risk-neutral moments of the underlying, proxies for variance risks, option Greeks, and bid-ask spreads of the underlying and the option itself. Of secondary importance, but nevertheless aiding in the prediction process for N-En are characteristics in the groups Past Prices, which includes measures of stock and option momentum,[22] Quality and Informed Trading. The dots in the figure represent the group importance during each month in the testing sample. On some occasions we find that illiquidity-based features are the most important. For most months, however, it is the group of contract-based information that influences the resulting predictions the most. We want to highlight a last observation from this figure: N-En relies much more on information drawn from options to make its predictions. In fact, the impact of most groups comprised solely of stock characteristics is very small in comparison.

In line with this, Figure 7 shows the 10 most important characteristics, along with the dispersion of their relative importance across the testing sample. The most influential characteristic by far is the implied volatility of the option, followed by the bid-ask spread of the underlying stock. Industry momentum (indmom), the stock's realized variance (rv), and the variance risk premium (ivrv) are also highly important. The reliance also on stock-based characteristics corroborates the findings by Cao et al. (2019).

Figure IA8.1 shows the rank of the importance of each feature group for N-En over the 18 years in our testing sample, with rank 1 denoting the lowest and rank 12 the highest importance. We find that contract-based characteristics are the most important in each year, with risk- and illiquidity-based characteristics, as well as information about past prices routinely ending up on the second or third highest rank. Illiquidity has played the second largest role in impacting N-En's predictions after the financial crisis, whereas risk-based information did so before and during 2008.

5.4. Impact of the Information Set

The previous section indicates that option-based characteristics and especially the option's location on the underlying's implied volatility surface are of high importance. The signals derived from these characteristics using modern machine learning techniques translate not only into option return predictability, but also highly profitable trading strategies. However, what happens if we assume that the investor restricts her attention

[22] Approximated by the characteristic *iv_rank*.

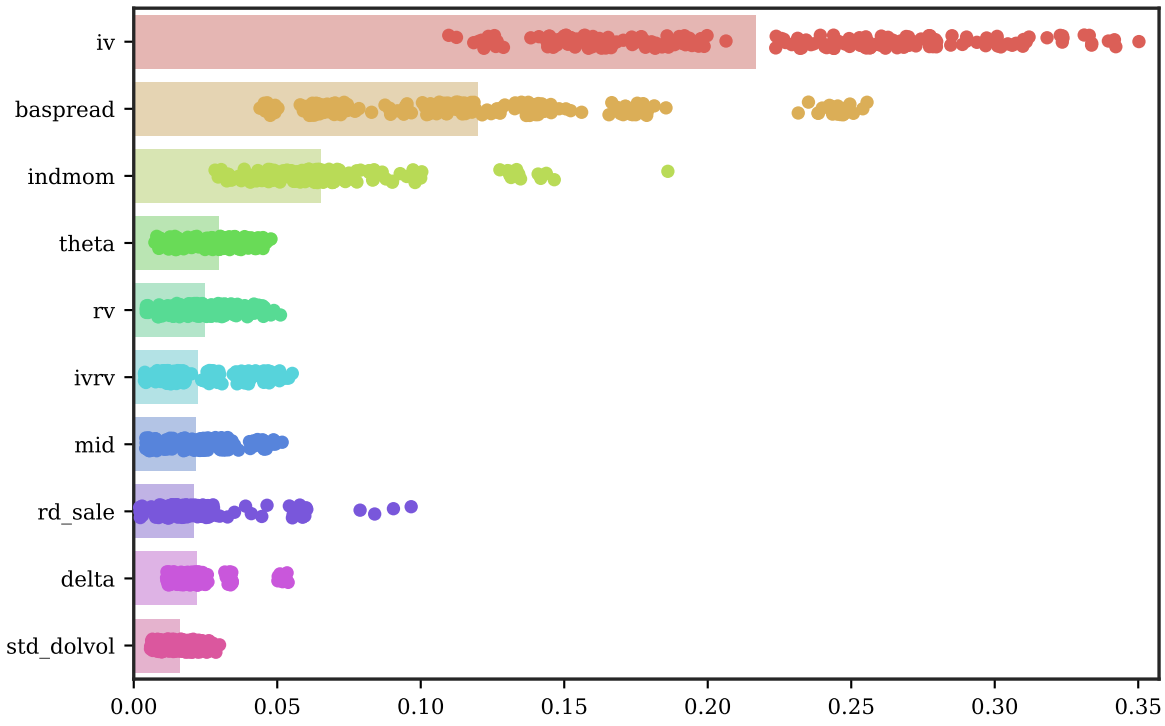

Fig. 7. Most Important Features

The figure shows the ten most influential features for the predictions of the nonlinear ensemble (N-En) following the importance using SHAP values (Lundberg and Lee, 2017). The values are scaled such that they sum to one across all 270 characteristics. The bars represent the mean feature group importance for the entire testing sample, the dots the dispersion of the group importance for the months in the testing sample.

to only a subset of characteristics? Likewise, how well will her predictions turn out if she chooses to consider only stock-based characteristics? To answer these questions, we re-estimate each model using only stock-based characteristics (S), option-based characteristics (O), or only those characteristics operating on the bucket- or contract-level (B+C) and contrast the resulting predictability with that of the full information set. Figure 8 provides the resulting full-sample R^2_{OS} values.

Using all available information produces the highest R^2_{OS} consistent with the idea that more information leads to better forecasts if the models used are sufficiently able to capture this information. Less weight is put on uninformative characteristics and important nonlinear interactions between them are taken advantage of. Restricting to only option-based information (O) comes in as the second place. The benefit of adding stock-based to the 77 option-based characteristics is substantial, given that the R^2_{OS} drops

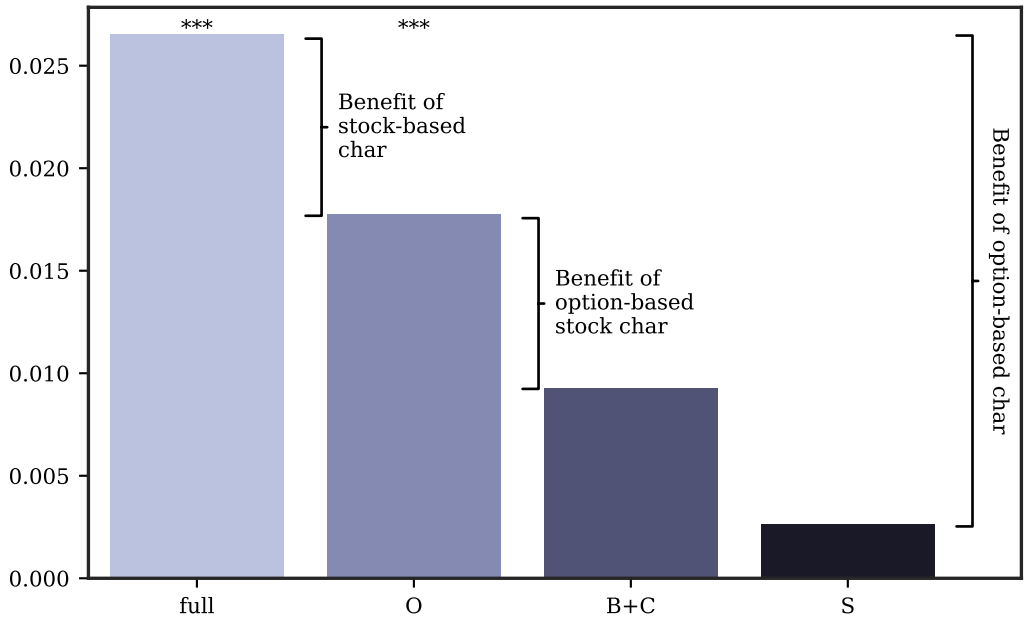

Fig. 8. Restricting the Information Set for N-En

The figure shows the out-of-sample R^2 defined in Equation (4) for N-En with restricted access to the full set of characteristics. The full model is shown in the left bar for reference, and is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA4. ***, **, * 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.

from 2.65% to 1.77% (both significant at the 1% level) for the whole sample if we exclude the whole stock-based characteristics block. Only considering the subset stock-based characteristics (S), however, is detrimental to uncovering option return predictability. The out-of-sample R^2 drops to 0.26%. The benefit of option-derived characteristics is huge when making informed forecasts of future delta-hedged option returns. As an additional check, we also consider whether option-contract and option-bucket information is sufficiently informative, which would render the addition of option-based characteristics for the underlying pointless. We strongly reject this idea, given that the inclusion of option-based characteristics for the underlying boosts out-of-sample predictability R^2_{OS} from 0.93% (B+C) to 1.77% (O). Consistent with the feature group importance shown in Figure 6, contract-based characteristics are highly important, but alone do not suffice to forecast future single-equity option returns. Analyzing the cross-sectional out-of-sample $R^2_{OS;XS}$ in Figure IA9.1 yields a comparable picture.

The increase in predictability shows that adding information from multiple sources is useful when predicting single-equity option returns. However, we are also interested

in making a statistical statement about this, answering whether the forecasts of N-En using more information are statistically more informative. For completeness, we also consider the linear ensemble (L-En) once more to understand in how far allowing for nonlinearities helps when using only a restricted set of characteristics. The resulting Diebold and Mariano (1995) forecast comparisons are provided in Panel A of Table 7.

Panel A: Diebold and Mariano (1995) Forecast Comparison

	L-En			N-En			
	O	S	full	B+C	O	S	full
L-En: B+C	**2.79**	−0.81	**3.82**	1.70	**2.60**	0.71	**3.64**
L-En: O		**-2.43**	**3.40**	1.08	**2.22**	0.05	**3.28**
L-En: S			**6.68**	**2.56**	**3.53**	1.50	**4.90**
L-En: full				−0.56	1.23	−1.92	**2.86**
N-En: B+C					**4.53**	**-1.98**	**4.83**
N-En: O						**-5.09**	**3.30**
N-En: S							**8.34**

Panel B: Forecast Correlation

	L-En			N-En			
	O	S	full	B+C	O	S	full
L-En: B+C	**0.89**	0.47	**0.77**	0.67	0.60	0.32	0.58
L-En: O		0.44	**0.84**	0.63	0.66	0.31	0.63
L-En: S			0.53	0.34	0.28	0.61	0.33
L-En: full				0.55	0.61	0.37	0.70
N-En: B+C					**0.86**	0.38	**0.76**
N-En: O						0.35	**0.84**
N-En: S							0.50

Table 7: Forecast Comparison

Panel A of the table shows Diebold and Mariano (1995) test statistics defined in Equation (8) to compare the forecasts for models with restricted access to the full set of characteristics. The full model is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA4. 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%).

For the linear and nonlinear ensemble, we find that adding more information is always worthwhile. At the same time, we find a clear hierarchy, which puts the informational content of option-based characteristics above that of stock-based characteristics. Models using both bucket and contract information (B+C) in their respective ensemble class beat models relying solely on stock-based information (S), but are outperformed by models also

leveraging the information inherent in option-based information about the underlying (O). The full model for L-En and N-En performs significantly better than all three alternative model specifications. Comparing the forecasts of the linear and nonlinear ensemble, we find that N-En restricted to option-based information (O) manages to surpass all but the full L-En model. In contrast, the full L-En model only manages to provide marginally more accurate forecasts than N-En restricted to stock-based characteristics (S).[23] In line with our intuition that (*i*) more information is always better and (*ii*) that nonlinear interactions are most valuable when making informed investment decisions in the options space, we find that the full nonlinear ensemble manages to outperform all four L-En specifications.

The impression from the Diebold and Mariano (1995) comparison carries over to evidence from forecast correlations in Panel B of Table 7. We find the largest similarity of the resulting predictions for linear or nonlinear ensembles estimated on option-based information (both B+C and O) and the full information set ($\rho_{t+1} > 0.7$). The forecast correlation of models using only option- and only stock-based characteristics is particularly low. Interestingly, the predictions made by the two full model specifications are highly correlated at $\rho_{t+1}^{\text{L-En: full; N-En: full}} = 0.7$.

The effectiveness of forecasts depends on the characteristics of the option. We therefore test how the return predictability changes for various option buckets (defined in Section 4) when restricting the information set in Table 8. Most of the times, the resulting best model uses all information. Notable exceptions are in the case of ITM short-term puts for both N-En and L-En. Using the nonlinear ensemble, the resulting predictability for the full model is significant for all short-term option buckets, and all but the in-the-money put bucket in case of long-term options. Predictions made with models restricted to just information based on the underlying stock tend to underperform substantially, even generating negative predictability for long-term options.

To conclude this section, we have established the importance of using option-based information for predicting future option returns. While more information is always beneficial when allowing for nonlinearities, if one has to restrict themselves to only a subset of available information, it is best to focus on option-based, as opposed to stock-based information. Given limited attention and limited information processing capacity of individual investors, this insight may prove useful for timing and engaging into investments in the options market.

[23] The t-statistic amounts to -1.92, corresponding to statistical significance at the 10% level.

			L-En				N-En			
TTM	Mon.		B+C	O	S	full	B+C	O	S	full
	atm		-0.003	-0.002	-0.000	0.004	0.011**	0.018***	0.006	0.022***
	itm	C	0.011	0.014	0.002	0.018*	0.007	0.029	0.016**	0.059***
$\tau \leq 90$	itm	P	0.026***	0.027**	0.010*	0.016	0.063***	0.060**	0.028***	0.061**
	otm	C	0.005	0.008*	0.002	0.012***	0.016***	0.023***	0.007***	0.027***
	otm	P	0.020***	0.021***	0.003	0.027***	0.033***	0.041***	0.005	0.048***
	atm		-0.012	-0.007	-0.013	0.005	0.003	0.013	-0.003	0.025***
	itm	C	-0.004	0.002	-0.011	0.014	0.013	0.033**	-0.002	0.047***
$\tau > 90$	itm	P	-0.033	-0.030	-0.029	-0.026	0.004	0.007	-0.013	0.015
	otm	C	-0.006	-0.003	-0.005	0.002	0.001	0.009	0.001	0.017***
	otm	P	-0.004	-0.001	-0.006	0.009*	0.002	0.012	0.001	0.023***

Table 8: Option-bucket performance for different information sets

The table shows the out-of-sample R^2 defined in Equation (4) for models with restricted access to the full set of characteristics for options in a respective bucket, as defined in Section 4. The full model is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA4. The information yielding the best information set per setup is underlined. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

6. 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, costly arbitrage, 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 investor attention and informational frictions, costly arbitrage, and option mispricing provide an explanation to the observed return predictability in the options market.

6.1. Presence of Informed Investors

We analyze if the return predictability is concentrated in options with lower presence of informed investors and higher trading by non-institutional traders. We hypothesize that option return predictability originates partly by informational frictions, such that the information implied from stock- and option-based characteristics is not directly incorporated into option prices. Informed or sophisticated traders are able to detect and process information sooner, and are therefore able to exploit mispricing. Hence, in the presence and trading of many informed traders, information is impounded into prices faster, or put differently, information diffuses faster. Consequently, slow information diffusion leads to additional option return predictability that is more pronounced for options with a lower share of informed traders.

We use two proxies for the presence of informed investors. First, we consider the case of institutional investors. Institutional investors are deemed to be more sophisticated and to exhibit superior ability in processing and reacting to information (see, among others, Chang, Hsieh, and Wang, 2015). Boehmer and Kelley (2009) find that stocks with a higher degree of institutional ownership are priced more efficiently. Following Eisdorfer et al. (2020), we take institutional ownership in the underlying stock as a proxy for the fraction of informed traders in the options market. Hence, option return predictability should be weaker for options with higher share of institutional investors. Additionally, we use analyst coverage as a second proxy for the presence of informed investors. The previous literature has provided empirical evidence that analyst coverage is an important channel through which information is incorporated into asset prices (Womack, 1996; Barber, Lehavy, McNichols, and Trueman, 2001; Gleason and Lee, 2003; Jegadeesh, Kim, Krische, and Lee, 2004; Kelly and Ljungqvist, 2012). Ellul and Panayides (2018) show that the termination of analysts covering a specific stock impairs price efficiency, causes more informed trading and leads to higher profitability of insider traders. Therefore, we hypothesize that higher analyst coverage of the underlying stocks translates into less option return predictability.

In order to investigate the resulting differences in return predictability, we form quintile splits at time t of the stocks in our sample, either by institutional ownership, or number of analysts covering the firm (Q1–Q5). Subsequently, we contrast the predictability of options written on stocks with different levels of institutional ownership (analyst coverage). Figure 9 depicts the R^2_{OS} values for the sub-samples. Higher institutional ownership directly translates to lower predictability of option contracts by N-En, confirming

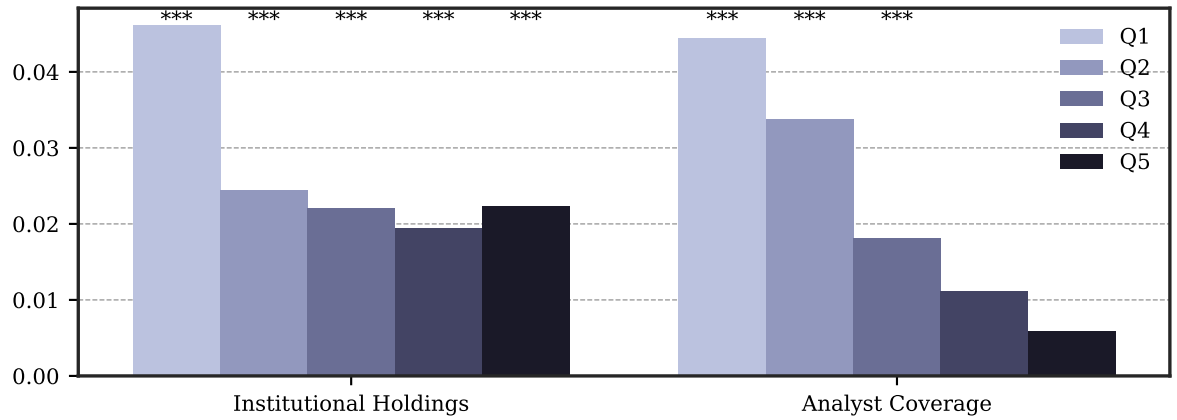

Fig. 9. Predictability and Profitability Conditional on Informed Investor Presence

The figure shows out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En for different quintiles of institutional ownership and analyst coverage of the underlying stocks, respectively. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7).

our hypothesis. We find the highest predictability for stocks in the lowest institutional ownership quintile at $R^2_{OS} = 4.6\%$. For stocks with higher institutional ownership, we find levels of predictability at around 1.9% to 2.5%. The predictability uncovered is statistically significant for all levels of institutional ownership. Turning to analyst coverage, option return predictability is strictly monotonically decreasing in the number of analysts covering a firm. In the lowest quintile, $R^2_{OS} = 4.4\%$, which drops to an insignificant 0.6% in Q5. Figure IA10.1 in the Online Appendix shows similar findings in the case of cross-sectional out-of-sample predictability.

The above findings motivate us to study the effect of high and low institutional ownership (analyst coverage) on the economic significance of return predictability in more detail. Precisely, we perform bivariate portfolio level analyses to study how option return predictability varies across different levels of institutional ownership and analyst coverage. We sort options into quintiles based on the level of institutional ownership (or analyst coverage) of the underlying stock at the end of month t. Subsequently, within each quintile, options are further sorted into quintiles by the one-month-ahead expected return forecast of N-En.

Panel A of Table 9 presents the results for the 25 (5x5) portfolios sorted by expected returns conditional on institutional ownership. It is evident from the last column that H-L return spreads are significantly positive regardless of the level of institutional ownership. Monthly returns range from 2.49% to 1.40%. The spreads between the lowest and highest return forecast quintiles are nearly monotonically decreasing such that the difference-in-

	Low Pred.	2	3	4	High Pred.	H-L
			Institutional Holdings			
Low	-1.985***	-0.505***	-0.222	-0.067	0.508**	2.492***
2	-1.138***	-0.342**	-0.177	-0.059	0.304	1.441***
3	-1.091***	-0.338**	-0.149	-0.024	0.311	1.402***
4	-1.027***	-0.308**	-0.129	0.031	0.445**	1.473***
High	-1.270***	-0.379***	-0.131	0.055	0.512**	1.782***
H-L	0.715***	0.126**	0.091**	0.122***	0.004	-0.710***
			Analyst Coverage			
Low	-2.056***	-0.483***	-0.076	0.213	0.870***	2.926***
2	-1.439***	-0.348**	-0.068	0.143	0.646***	2.085***
3	-0.967***	-0.229*	-0.063	0.097	0.458**	1.425***
4	-0.799***	-0.209	-0.077	0.039	0.406*	1.205***
High	-0.736***	-0.272**	-0.094	-0.003	0.279	1.015***
H-L	1.321***	0.211***	-0.017	-0.216***	-0.591***	-1.911***

Table 9: Bivariate Portfolios of Institutional Attention and Expected Returns

The table shows realized returns for quintile portfolios following the predictions by the nonlinear ensemble N-En within quintiles sorted by analyst coverage in the upper panel and institutional ownership share in the lower panel. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level as defined in Equation (7). We also show the realized returns and significance for the resulting high-minus-low portfolios.

difference return spread is economically (-0.71% per month) and statistically significant at the 1% level.

Panel B of Table 9 shows even stronger results for analyst coverage. The return spreads in the last column confirm pervasive option return predictability for all analyst coverage quintiles by the N-En model. The return spreads are strictly monotonically decreasing in the degree of analyst coverage, confirming our hypotheses that option return predictability is directly related to institutional attention. Furthermore, this leads to a negative and highly significant difference between the high-minus-low spreads of the highest and lowest analyst coverage quintile of -1.91% per month. Taken together, the results for analyst coverage and institutional ownership are in line with the notion that the presence of informed traders and the resulting faster diffusion of information reduces option return predictability.

6.2. Option Demand by Types of Market Participants

We now investigate in how far option demand by professional traders and public customers translates into return predictability. For this, we obtain open/close option trans-

action data for various types of market participants from four NASDAQ exchanges and the CBOE C1 exchange from January 2011 through December 2020.[24] The open/close data set unfortunately does not provide a clear distinction between institutional and retail trading. As an approximation, we separately consider "public customers", which we denote as Customers (C), and the joint trading activity of "proprietary traders" and "professional customers", which we summarize as Professionals (P). While this classification does not guarantee that trades of institutions do not end up in the group of C, the relative concentration of informed trades is expected to be much higher in group P compared to group C.[25] If anything, under our assumption that institutional trading leads to more efficient prices, our results should be even stronger if we were able to adequately differentiate between retail and institutional options trading. Eisdorfer et al. (2020) construct investor groups in the options market in a similar fashion to understand whether retail-driven trades exhibit more behavioral biases.

An investor may express pessimistic views about the underlying stock by purchasing puts or alternatively short-selling call options. The opposite holds true for betting on optimistic expectations. We therefore consider signed open interest in individual option contracts over the last month, and analyze how buying or selling pressure by C and P impacts the resulting return predictability during month $t+1$. Define the signed volume over the last month for contract i as:

$$Volume_{i,t}^{x;Buy} = Volume_{i,t}^{x;OpenBuy} + Volume_{i,t}^{x;CloseBuy} \tag{13}$$

$$Volume_{i,t}^{x;Sell} = Volume_{i,t}^{x;OpenSell} + Volume_{i,t}^{x;CloseSell}, \tag{14}$$

for option j at time t and $x \in$ [C, P]. Net demand for each investor group is consequently defined as the difference between the daily buy and sell volume, which we denote as V^x. We calculate monthly net demand of option i by summing up the respective daily signed volume. We scale the resulting signed volume measures V^x by the total *non-signed* volume of public customers, professional traders, and proprietary trading desks to make it comparable across stocks. Furthermore, we aggregate individual option volume $V_{i,t}^x$ at the level of the underlying to arrive at $V_{S,t}^x$. Following Garleanu et al. (2009), demand in one option has price effects also on neighboring options of the same underlying. While any demand potentially increases the informativeness of market prices, we argue that the

[24]The four NASDAQ exchanges are NASDAQ PHOTO, NASDAQ Options Market, NASDAQ ISE, and NASDAQ GEMX.

[25]We also note that the classification into "professional customer" differs between the CBOE and NASDAQ exchanges.

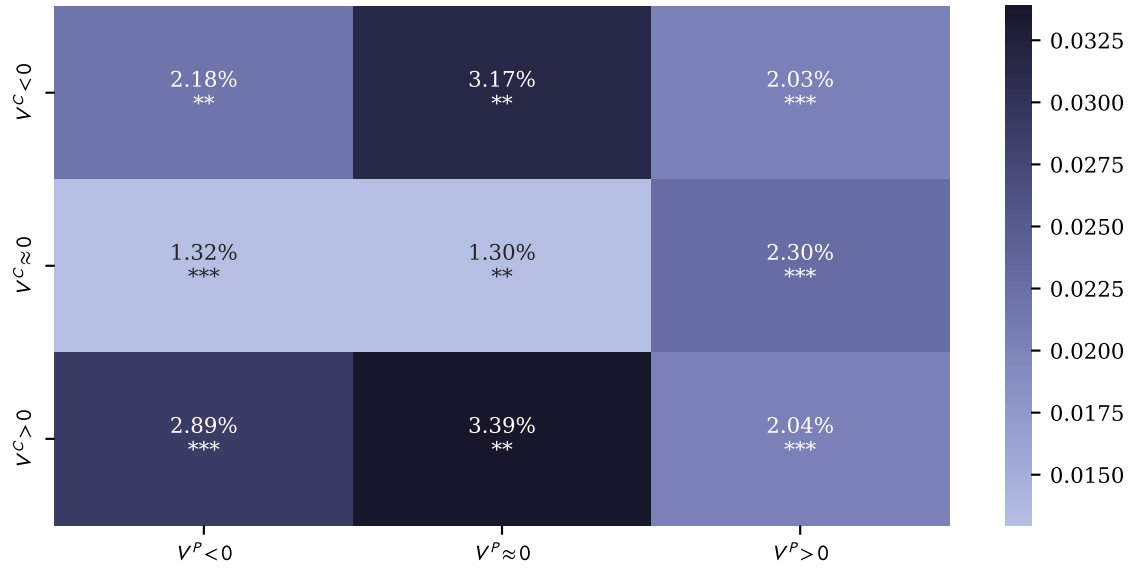

Fig. 10. Predictability Conditional on Option Demand

The figure shows the out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En conditional on the net open interest of retail and institutional investors, I. We independently form terciles based on the cross-sectional distribution of I^R and I^I. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

impact of institutional demand is amplified. This in turn should correspond to decreasing return predictability.

Figure 10 shows the resulting R^2_{OS} for public customers (C) vs. professional traders (P). We define terciles based on the cross-sectional distribution of V^x during month t and analyze the resulting predictability for $t+1$. We find multiple interesting results: First, option return predictability is highest for options with little to no demand by professionals ($V^P \approx 0$), while it is lowest if demand by both investor groups is missing ($R^2_{OS} = 1.30\%$). Second, compared to these cases, a lack of customer demand if demand by professionals is present leads to much lower predictability, consistent with our hypothesis that professionals are quicker to impound information into market prices. Our model (N-En) manages to identify these situations of mispricing and allows an investor to profit from them. The model also uncovers high predictability in cases where public customers are buying options of an underlying and professionals facilitate their buying intent by shorting, which is in line with Garleanu et al. (2009) who argue that professional traders tend to act as market markers. We add to this by showing that they tend to act in this way in cases where N-En uncovers high predictability of the respective option from which professionals can profit.

6.3. Costly Arbitrage

Our main results have shown a high level of option return predictability, especially in comparison to the cross-section of stock returns (Gu et al., 2020). Moreover, we find that option return predictability is stable over time and does not vanish towards the end of our sample period. Hence, one might wonder if limits-to-arbitrage are preventing competitive arbitrageurs from benefiting of pricing anomalies underlying option return predictability.

As liquidity is considered one facet of costly arbitrage (Atilgan, Bali, Demirtas, and Gunaydin, 2020), the analysis on the impact of transaction costs on the profitability on machine learning portfolios in Section 5.2.2 has hinted at its importance. Whereas conservative estimates of transaction costs have a substantial influence on long-term options, trading short-term at-the-money and out-of-the-money options remains profitable in all months of our sample.

Idiosyncratic risk has been identified as the main arbitrage cost in case of stocks (Shleifer and Vishny, 1997; Pontiff, 2006). Rather than relying on idiosyncratic risk of option excess returns, we focus on the bid-ask spread of the underlying stock as an impediment to trading the underlying. Arbitrageurs who have identified an anomaly in a given option will want to hedge their directional exposure from the underlying stocks. High bid-ask spreads of the underlying stock therefore present costs for arbitrageurs in delta-hedging their option positions. Consequently, pricing errors may persist even in the presence of arbitrageurs.

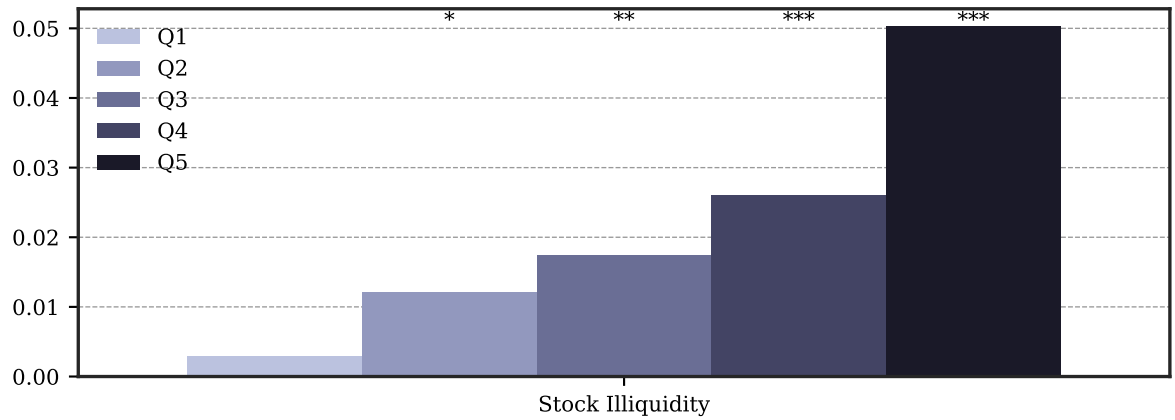

Fig. 11. Predictability and Profitability Conditional on Stock Illiquidity – R^2_{OS}

The left panel of the figure shows out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En for different quintiles of the underlying stock's illiquidity measured by the relative bid ask spread, $baspread = \frac{ask-bid}{mid}$. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

In order to understand how costly arbitrage is influencing option return predictability, we investigate the effect of stock bid-ask spreads on out-of-sample R^2_{OS}s. In Figure 11, we have sorted our option sample into quintiles based on the underlying stock's bid-ask spread. Figure 11 shows that options on stocks with higher bid-ask spreads exhibit higher return predictability. This relationship is strictly monotone and amplified for the quintile of the least liquid underlying stocks. Here, we find predictability of approximately 5%.[26]

	Low Pred.	2	3	4	High Pred.	H-L
			Stock Illiquidity			
Low	-0.585***	-0.188*	-0.082	0.019	0.189	0.774***
2	-0.770***	-0.210*	-0.063	0.058	0.358**	1.127***
3	-1.008***	-0.235	-0.015	0.158	0.520**	1.528***
4	-1.348***	-0.407***	-0.084	0.167	0.674***	2.022***
High	-2.567***	-0.916***	-0.241	0.292	1.209***	3.776***
H-L	-1.981***	-0.734***	-0.164	0.267**	1.017***	2.998***

Table 10: Bivariate Portfolios of Stock Illiquidity and Expected Returns

The table shows realized returns for quintiles portfolios following the predictions by the nonlinear ensemble N-En within quintiles of the underlying stock's illiquidity measured by the relative bid ask spread, $baspread = \frac{ask-bid}{mid}$. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level as defined in Equation (7). We also show the realized returns and significance for the resulting high-minus-low portfolios.

We also form bivariate portfolios by sorting options into the expected return quintiles within each stock bid-ask quintile. Table 10 reports the results. Return spreads are significantly positive regardless of the level of stock bid-ask spreads. Moreover, the one-month-ahead average realized return spreads between the lowest and highest expected return forecast quintiles are strictly monotonically increasing in stock bid-ask spreads. Most importantly, the diff-in-diff return spread is economically large at 3% per month and highly statistically significant, showing that the return predictability in the cross-section of equity options is significantly influenced by the level of illiquidity of underlying stocks. Overall, the results suggest that option return predictability is at least in part tied to costly arbitrage.

6.4. Option Mispricing

We expect to find higher levels of predictability for options that are mispriced as the nonlinear ensemble model should manage to identify these opportunities and correctly propose shorting over- and investing into undervalued options.

[26]Evidence from cross-sectional $R^2_{OS;XS}$ confirms these findings, see Figure IA10.3.

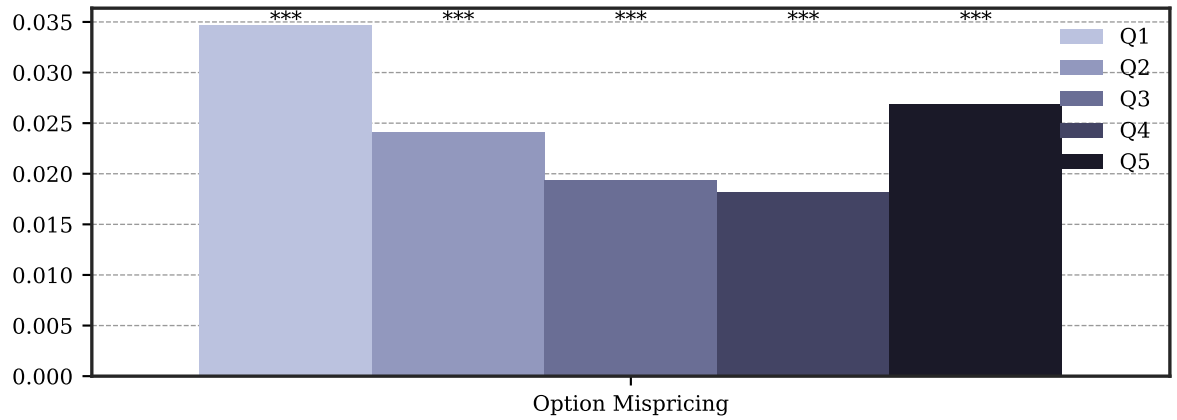

Fig. 12. Predictability and Profitability Conditional on Option Mispricing – R^2_{OS}

The left panel of the figure shows out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En for different quintiles of option mispricing. We calculate option mispricing using the stocks' realized volatility over the last quarter and calculate "fair prices" for short-term at-the-money options using it as input to the Black and Scholes (1973) model. Mispricing is then defined as $\log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the obtained "fair price". ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5% level as defined in Equation (7).

We follow Eisdorfer et al. (2020) to quantify option mispricing as the ratio between theoretical and observed option prices. The theoretical option price is given by the Black and Scholes (1973) pricing model, where we use each underlying stock's realized volatility over the past quarter, estimated using high-frequency price data from the NYSE TAQ database, as our estimate for the expected volatility. For all short-term at-the money options, we compare the log of the theoretical price with the log of the price observed in the market, i.e., Mispricing $= \log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the theoretical and O the observed mid price. A high (low) level of mispricing thus denotes overvaluation (undervaluation). Averaging over all short-term at-the-money options, we obtain one level of mispricing per underlying stock at each point in time.

The predictability for different mispricing quintiles is provided in Figure 12. Consistent with our hypothesis that predictability clusters within over- and under-valued options, we find the highest level of predictability in the most undervalued, followed by the most overvalued options. This create a U-shaped predictability pattern in the option's mispricing.

This predictability pattern also translates into additional profitability, as depicted in Table 11. The high-minus-low spread within each option mispricing portfolio is highly significant, generating monthly profits upwards of 1.52% per month. Interestingly, the difference between these spreads for under- and overvalued stocks is insignificant at -0.06%

	Low Pred.	2	3	4	High Pred.	H-L	vs. Fair
				Option Mispricing			
Under	-1.278***	-0.304**	-0.054	0.156	0.678***	1.956***	0.441***
2	-1.101***	-0.255*	-0.045	0.132	0.593***	1.694***	0.179**
Fair	-0.995***	-0.234	-0.087	0.069	0.520**	1.515***	---
4	-0.994***	-0.232*	-0.031	0.134	0.545**	1.539***	0.024
Over	-1.467***	-0.381***	-0.103	0.059	0.427*	1.894***	0.379***
Over-Under	-0.190*	-0.077	-0.050	-0.097*	-0.251**	-0.062	---

Table 11: Bivariate Portfolios of Option Mispricing and Expected Returns

The table shows realized returns for quintiles portfolios following the predictions by the nonlinear ensemble N-En within quintiles sorted by option mispricing. We calculate option mispricing using the stocks' realized volatility over the last quarter and calculate "fair prices" for short-term at-the-money options using it as input to the Black and Scholes (1973) model. Mispricing is then defined as $\log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the obtained "fair price". We also show the realized returns and significance for the resulting high-minus-low portfolios. Column "vs. Fair" tests statistical differences between the H-L portfolios for under- and overvalued options, and fairly valued options. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level as defined in Equation (7).

per month, suggesting that significant trading opportunities exist within over- *and* undervalued options. Instead, when we compare the high-minus-low portfolio returns of mispriced options with the portfolios using only fairly priced options (so $O \approx \widetilde{O}$), we find that they are significantly different from each other, with higher returns both for undervalued (0.44%) and overvalued options (0.38% per month). To summarize, the employed machine learning methods manage to pick up on mispricing opportunities in the options market and allow investors to profit from them.

7. Conclusion

An extensive literature examines cross-sectional determinants of stocks, bonds, currencies, mutual funds, and hedge funds. However, research on cross-sectional predictors of option returns is relatively scarce and not very well understood. In this paper, we close this gap in the literature and identify variables that predict the cross-sectional differences in delta-hedged option returns. Predicting option returns is of foremost relevance for retail and institutional investors as the importance of option markets for hedging and speculation purposes has strongly increased in the past years.

In this paper, we apply machine learning to predict individual U.S. option returns using a set of 77 option-based and 193 stock-based characteristics in the period from 1996 to 2020. Empirically, we derive several results that enhance our knowledge on

empirical asset pricing. First, we show that the complexity of the machine learning models matters 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 when accounting for realistic transaction costs. Third, we find that characteristics describing the option's location on the underlying's implied volatility surface are the most relevant to successfully predict option returns. Finally, we document that the sources of return predictability are related to informational frictions in the options market and the existence of costly arbitrage. In line with this notion, we find that option return predictability is higher for options on stocks characterized by (i) low institutional ownership and low analyst coverage, (ii) with little demand by professional traders, but substantial interest by public customers, such as retail investors, (iii) for illiquid underlying stocks, and (iv) mispriced options.

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.

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.

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.

Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. Review of Financial Studies 16, 527–566.

Bali, T. G., Goyal, A., Huang, D., Jiang, F., Wen, Q., 2021. Different strokes: Return predictability across stocks and bonds with machine learning and big data. Working paper.

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

Bali, T. G., Murray, S., 2020. In search of a factor model for optionable stocks. Working paper.

Barber, B., Lehavy, R., McNichols, M., Trueman, B., 2001. Can investors profit from the prophets? security analyst recommendations and stock returns. Journal of Finance 56, 531–563.

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

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.

Boehmer, E., Kelley, E. K., 2009. Institutional investors and the informational efficiency of prices. Review of Financial Studies 22, 3563–3594.

Bollen, N. P., Whaley, R. E., 2004. Does net buying pressure affect the shape of implied volatility functions? Journal of Finance 59, 711–753.

Bollerslev, T., Todorov, V., Xu, L., 2015. Tail risk premia and return predictability. Journal of Financial Economics 118, 113–134.

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

Buraschi, A., Jackwerth, J., 2001. The price of a smile: Hedging and spanning in option markets. Review of Financial Studies 14, 495–527.

Byun, S.-J., Kim, D.-H., 2016. Gambling preference and individual equity option returns. Journal of Financial Economics 122, 155–174.

Cao, J., Han, B., 2013. Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics 108, 231–249.

Cao, J., Han, B., Tong, Q., Zhan, X., 2021. Option return predictability. Review of Financial Studies, forthcoming.

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

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

Chang, C.-C., Hsieh, P.-F., Wang, Y.-H., 2015. Sophistication, sentiment, and misreaction. Journal of Financial and Quantitative Analysis 50, 903–928.

Chen, L., Pelger, M., Zhu, J., 2021. Deep learning in asset pricing. Working paper.

Christoffersen, P., Goyenko, R., Jacobs, K., Karoui, M., 2018. Illiquidity premia in the equity options market. Review of Financial Studies 31, 811–851.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

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

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

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

Eisdorfer, A., Goyal, A., Zhdanov, A., 2020. Cheap options are expensive. Swiss finance institute research paper.

Ellul, A., Panayides, M., 2018. Do financial analysts restrain insiders' informational advantage? Journal of Financial and Quantitative Analysis 53, 203–241.

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

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

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

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. Journal of Finance 75, 1327–1370.

Filippou, I., Rapach, D., Taylor, M. P., Zhou, G., 2020. Exchange rate prediction with machine learning and a smart carry portfolio. Working paper.

Fong, K. Y., Holden, C. W., Trzcinka, C. A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Review of Financial Studies 33, 2326–2377.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics pp. 1189–1232.

Garleanu, N., Pedersen, L. H., Poteshman, A. M., 2009. Demand-based option pricing. Review of Financial Studies 22, 4259–4299.

Giglio, S., Liao, Y., Xiu, D., 2021. Thousands of alpha tests. Review of Financial Studies 34, 3456–3496.

Gilad-Bachrach, R., Rashmi, K., 2015. Dart: Dropouts meet multiple additive regression trees. Cornell University: Cornell, Ithaca, NY, USA .

Gleason, C. A., Lee, C. M. C., 2003. Analyst forecast revisions and market price discovery. Accounting Review 78, 193–225.

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

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

Goyenko, R., Zhang, C., 2021. The joint cross section of option and stock returns predictability with big data and machine learning. 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., 2020. 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., 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.

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.

Jegadeesh, N., Kim, J., Krische, S. D., Lee, C. M. C., 2004. Analyzing the analysts: When do recommendations add value? Journal of Finance 59, 1083–1124.

Jensen, T. I., Kelly, B. T., Pedersen, L. H., 2021. Is there a replication crisis in finance? Tech. rep., National Bureau of Economic Research.

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

Kelly, B. T., Ljungqvist, A., 2012. Testing asymmetric-information asset pricing models. Review of Financial Studies 25, 1366–1413.

Kelly, B. T., Palhares, D., Pruitt, S., 2020a. Modeling corporate bond returns. Working paper.

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.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25, 1097–1105.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2016. Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474 .

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

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., 2020. Market efficiency in the age of big data. Working paper, National Bureau of Economic Research.

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

Muravyev, D., 2016. Order flow and expected option returns. Journal of Finance 71, 673–708.

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

Muravyev, D., Pearson, N. D., Pollet, J. M., 2021. Is there a risk premium in the stock lending market? evidence from equity options. Journal of Finance, forthcoming.

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

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.

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.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.

Steel, M. F., 2020. Model averaging and its use in economics. Journal of Economic Literature 58, 644–719.

Tian, M., Wu, L., 2021. Limits of arbitrage and option risk premium. 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., Xiao, X., 2021. Default risk and option returns. Working paper.

Womack, K. L., 1996. Do brokerage analysts' recommendations have investment value? Journal of Finance 51, 137–167.

Wu, W., Chen, J., Yang, Z., Tindall, M. L., 2021. A cross-sectional machine learning approach for hedge fund return prediction and selection. Management Science 67, 4577–4601.

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 IA1. 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):

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{i,s,t+1} - g(z_{i,s,t}))^2 \tag{IA1}$$

Appendix IA2. 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 (IA1)) 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 IA2.1 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 IA2.1: 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 IA3. Option-Based Characteristics

This section describes a broad set of the 77 option-based characteristics, motivated by earlier studies on the cross-section of option and/or stock returns. Out of the 77 we compute, 42 characteristics operate on the level of the underlying stock, 19 on the level of option buckets (that is, we differentiate between different parts of the time-to-maturity and moneyness domain of options, described in Section 4.3), and 16 on the level of individual option contracts.

IA3.1. 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. **Proportional bid-ask spread (*pba*).** Following Cao and Wei (2010), we use the proportional bid-ask spread as a measure of illiquidity

$$pba = \frac{\sum_j VOL_j \times \frac{ask_j - bid_j}{0.5 \times (ask_j + bid_j)}}{\sum_j VOL_j},$$

where VOL_j denotes the trading volume in option j, ask_j and bid_j the bid and ask spread of option j, respectively.

14. **Dollar trading volume (*dvol*).** Following Cao and Wei (2010), we include the

dollar trading volume across all options,

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

15. **Absolute illiquidity (*ailliq*).** Following Cao and Wei (2010), we introduce the absolute illiquidity as
$$ailliq = \frac{\sum_j \times \frac{|\Pi_t|}{DVOL_j}}{\sum_j VOL_j},$$
where $DVOL_j$ denotes the dollar trading volume in option j.
16. **Percentage illiquidity (*pilliq*)**. Following Cao and Wei (2010), we introduce the absolute illiquidity as
$$ailliq = \frac{\sum_j \times \frac{|\Pi_t|}{DVOL_j \times O_j}}{\sum_j VOL_j},$$
where $DVOL_j$ denotes the dollar trading volume in option j, and O_j the price of option j.
17. **Trading volume (*vol*).** Following Cao and Wei (2010), we include the trading volume across all options as defined as $\sum_j VOL_j$, with VOL_j being the volume in option j.
18. **Number of traded options (*nopt*).** The average number of options per underlying stock per month.
19. **Total open interest (*toi*).** Open interest across all options on an underlying.
20. **Volatility uncertainty (*volunc*)**. Following Cao et al. (2019), we calculate monthly volatility-of-volatility based on different measures of daily volatility estimates. As a first measure, we take implied volatilities of call options that have a delta of 0.5 and 30 days to maturity. As a second measure, the estimate an EGARCH(1,1) model with daily stock returns over a rolling window of the past twelve months. For both measures of volatility, we calculate the return of volatility as $\frac{\Delta\sigma}{\sigma} = \frac{\sigma_t - \sigma_{t-1}}{\sigma_{t-1}}$, where σ_t is the volatility on day t. Subsequently, we calculate for each measure a monthly volatility-of-volatility estimate as the standard deviatio of the daily percentage in volatility. Next, we rank stocks based on the two measures. Finally, we compute ***volunc*** as the average of the ranking percentile of the two individual volatility-of-volatility measures. Note that Cao et al. (2019) includes a third volatility-of-volatility measure based on realized variance based on intraday data.

21. **Atm iv volatility (*ivvol*).** Following Baltussen, van Bekkum, and van der Grient (2018), volatility of atm implied volatility scaled by average implied volatility, that is

$$VOV_t^{1M} = \frac{\sqrt{\frac{1}{20}\sum_{j=t-19}^{t}\left(\sigma_j^{IV} - \bar{\sigma}_t^{IV}\right)^2}}{\bar{\sigma}_t^{IV}},$$

where $\bar{\sigma}_t^{IV} = (1/20)\sum_{j=t-19}^{t}\sigma_j^{IV}$, and σ_j^{IV} is implied volatility.

22. **Variance spread (*ivrv*)**. Following Bali and Hovakimian (2009), the realized-implied volatility spread, defined as

$$ivrv = RVol - IVol,$$

where $RVol$ is the realized volatility over month t and $IVol$ is the average volatility implied by atm call and atm put options observed at the end of month t.

23. **Variance spread (*ivrv_ratio*)**. The realized-implied volatility ratio, defined as

$$ivrv = \frac{IVol}{RVol},$$

where $RVol$ is the realized volatility over month t and $IVol$ is the average volatility implied by atm call and atm put options observed at the end of month t.

24. **Near-the-money call minus put implied volatility (*civpiv*)**. Following Bali and Hovakimian (2009), the implied volatility spread of call and put options, defined as

$$civpiv = CVol - PVol,$$

where $CVol$ and $PVol$ denote call and put near-the-money implied volatility, respectively.

25. **Atm call minus put implied volatility based on implied volatility surface data (*atm_civpiv*)**. The implied volatility spread of call and put options, defined as

$$atm\_civpiv = CVol - PVol,$$

where $CVol$ and $PVol$ denote call and put atm implied volatility based on implied volatility surface data, respectively.

26. **Change in atm call IV (*dciv*)**. Following An et al. (2014), the change in the

implied volatility at-the-money call options, defined as

$$dciv = CVol_t - CVol_{t-1},$$

where $CVol_t$ denotes month-t call implied volatility based on implied volatility surface data.

27. **Change in atm put IV (*dpiv*)**. Following An et al. (2014), the change in the implied volatility at-the-money put options, defined as

$$dpiv = PVol_t - PVol_{t-1},$$

where $PVol_t$ denotes month-t put implied volatility based on implied volatility surface data.

28. **Change in atm put minus call IV (*atm_dcivpiv*)**. Following An et al. (2014), the change in the implied volatility spread of call and put options, defined as

$$atm\_dcivpiv = (CVol_t - CVol_{t-1}) - (PVol_t - PVol_{t-1}),$$

where $CVol_t$ and $PVol_t$ denote month-t call and put atm implied volatility based on implied volatility surface data, respectively.

29. **IV skew (*skewiv*)**. Following Xing, Zhang, and Zhao (2010), an implied volatility smirk measure as the difference between the implied volatilities of otm puts and atm calls, denoted by VOL^{OTMP} and VOL^{ATMC}, respectively, that is

$$skewiv = VOL^{OTMP} - VOL^{ATMC}.$$

We compute monthly ***skewiv*** by averaging over daily ***skewiv***.

30. **Weighted put-call spread (*vs_level*)**. Following Cremers and Weinbaum (2010), the call-put spread is

$$VS_t = IV_t^{calls} - IV_t^{puts} = \sum_{j=1}^{N_t} w_{j,t} \left(IV_{j,t}^{call} - IV_{j,t}^{put} \right),$$

where j denotes pairs of put and call options with the same strike and time to maturity, $w_{j,t}$ are weights, N_t denotes the number of valid pairs of options on day t, and $IV_{j,t}$ denotes Black and Scholes (1973) implied volatility. Average open interest in the call and puts are used as weights.

31. **Change in weighted put-call spread (*vs_change*)**. Following Cremers and Weinbaum (2010), we compute changes in ***vs_level***.
32. **Put-call parity violations (*pcpv*)**. We follow Ofek et al. (2004) and record violations of put-call-parity via the midpoints of option quotes and the closing price of the stock. Precisely, ***pcpv*** is given as

$$pcpv = 100 \log\left(\frac{S}{S^*}\right),$$

where S denotes the stock price, and $S^* = PV(K) + C - P$, where $PV(K)$ is the present value of the strike price K, and C and P denote the prices of a call and put option, respectively. Ofek et al. (2004) focus on ATM and intermediate maturities (i.e. between 91 and 182 days). As the authors filter for dividends and we want to exclude as little stocks as possible, we focus on ATM and short-term maturities, which are also studied by Ofek et al. (2004), but in a sub-analysis. We compute an average over the previous month.
33. **Implied shorting fees in options (*shrtfee*).** Muravyev et al. (2021) propose an option-based shorting fee measure as

$$h_Q^{imp} = \frac{1}{\delta}\left(1 - \left(1 - \frac{S_t - C_t + P_t - PV(D) - PV(K)}{S_t}\right)^{1/k}\right),$$

where C_t and P_t are the midpoints of quoted call and put quote prices, $PV(D)$ is the present value of dividends with ex-dividend dates before the expiration date, k denotes the time to expiration, S_t is the current stock price and K is the strike price, and δ is the one-day discount factor. We take the median of the implied borrowing fees from put-call pairs.
34. **Implied volatility duration (*ivd*).** Measure for the expected timeliness of the resolution of uncertainty, following Schlag, Thimme, and Weber (2020). It is defined as

$$ivd = \sum_{j=1}^{J} \frac{\Delta IV_j^2}{\sum_j \Delta IV_j^2} \times \tau_j,$$

where $\Delta IV_j^2 = IV_{j,\tau_j}^2 - IV_{j,\tau_{j-1}}^2$ is the difference between the non-annualized squared IVs for all options at τ_j and those at τ_{j-1}, and $(\tau_1, \ldots, \tau_8) = (30, 60, 91, 122, 152, 182, 273, 365)$ (days).

IA3.2. 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. **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,

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

2. **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.
3. **Illiquidity measure based on zero returns (*pzeros*)** As in Lesmond, Ogden, and Trzcinka (1999), we take the proportion of zero return days as a measure of liquidity. We compute their measure on a monthly basis as

$$\text{pzeroes} = \frac{\text{\# of zero return days}}{T},$$

   where T denotes the number of days in a month.
4. **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.
5. **Amihud measure of illiquidity (*amihud*)** Following Amihud (2002), the mea-

sure 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.

6. **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 $roll$.

7. **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.

8. **Std.dev of the Amihud measure (*stdamihud*)** The standard deviation of the daily Amihud (2002) measure within a month.
9. **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 $gammaps = -\gamma$.

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

month.

13. **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

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

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

IA3.3. 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. **Implied volatility (*iv*).** Following Buchner and Kelly (2020), the Black and Scholes (1973) implied volatility of the option contract.
7. **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.
8. **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.
9. **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.
10. **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 volatility. We scale *vega* by the price of the underlying stock to make it comparable in the cross-section.
11. **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.
12. **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.

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

16. **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 IA4. Classification of Characteristics

In this section we provide a detailed summary of the characteristics used in our analyses. We consider a total of 270 characteristics, of which 77 are derived from option-based information and the remainder from stock-based information. This information source is provided in the table below. We further provide the instrument level the respective characteristic relates to. Here, we consider three different levels. "Contract" information relate to a single option contract. Examples for this are the open interest or the option's delta. Since flow-based measures cannot be estimated for individual option contracts due to migrating moneyness and fleeting time-to-maturity, we construct various option buckets, outlined in Section 4. The characteristics on this level are denoted by instrument level "Bucket". The final group is that of characteristics operating on the level of the "Underlying" stock.

As a last set of information we also group the characteristics into 12 groups. Group "Accruals" contains five characteristics, "Contract" six, "Frictions" contains four characteristics, "Illiquidity" 28, "Industry" 90, "Informed Trading" 17, "Investment" 11, "Past Prices" 13, "Profitability" 16, "Quality" 29, "Risk" 40, and "Value" 10. The grouping for stock-based characteristics follows the intuition formed by Green et al. (2017) and Jensen et al. (2021). We group the remaining option-based characteristics accordingly.

Table IA4.1: Classification of Characteristics

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

Continued on Next Page

Table IA4.1 from previous page

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

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ill	Amihud Illiquidity	Green et al. (2017)	Underlying	Underlying	Illiquidity
illiq	Illiquidity	Bao et al. (2011)	Options	Bucket	Illiquidity
ind_10	Industry code		Underlying	Underlying	Industry
ind_11	Industry code		Underlying	Underlying	Industry
ind_12	Industry code		Underlying	Underlying	Industry
ind_13	Industry code		Underlying	Underlying	Industry
ind_14	Industry code		Underlying	Underlying	Industry
ind_15	Industry code		Underlying	Underlying	Industry
ind_16	Industry code		Underlying	Underlying	Industry
ind_17	Industry code		Underlying	Underlying	Industry
ind_18	Industry code		Underlying	Underlying	Industry
ind_19	Industry code		Underlying	Underlying	Industry
ind_20	Industry code		Underlying	Underlying	Industry
ind_21	Industry code		Underlying	Underlying	Industry
ind_22	Industry code		Underlying	Underlying	Industry
ind_23	Industry code		Underlying	Underlying	Industry
ind_24	Industry code		Underlying	Underlying	Industry
ind_25	Industry code		Underlying	Underlying	Industry
ind_26	Industry code		Underlying	Underlying	Industry
ind_27	Industry code		Underlying	Underlying	Industry
ind_28	Industry code		Underlying	Underlying	Industry
ind_29	Industry code		Underlying	Underlying	Industry
ind_30	Industry code		Underlying	Underlying	Industry
ind_31	Industry code		Underlying	Underlying	Industry
ind_32	Industry code		Underlying	Underlying	Industry
ind_33	Industry code		Underlying	Underlying	Industry
ind_34	Industry code		Underlying	Underlying	Industry
ind_35	Industry code		Underlying	Underlying	Industry
ind_36	Industry code		Underlying	Underlying	Industry
ind_37	Industry code		Underlying	Underlying	Industry
ind_38	Industry code		Underlying	Underlying	Industry
ind_39	Industry code		Underlying	Underlying	Industry
ind_40	Industry code		Underlying	Underlying	Industry

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ind_41	Industry code		Underlying	Underlying	Industry
ind_42	Industry code		Underlying	Underlying	Industry
ind_43	Industry code		Underlying	Underlying	Industry
ind_44	Industry code		Underlying	Underlying	Industry
ind_45	Industry code		Underlying	Underlying	Industry
ind_46	Industry code		Underlying	Underlying	Industry
ind_47	Industry code		Underlying	Underlying	Industry
ind_48	Industry code		Underlying	Underlying	Industry
ind_49	Industry code		Underlying	Underlying	Industry
ind_50	Industry code		Underlying	Underlying	Industry
ind_51	Industry code		Underlying	Underlying	Industry
ind_52	Industry code		Underlying	Underlying	Industry
ind_53	Industry code		Underlying	Underlying	Industry
ind_54	Industry code		Underlying	Underlying	Industry
ind_55	Industry code		Underlying	Underlying	Industry
ind_56	Industry code		Underlying	Underlying	Industry
ind_57	Industry code		Underlying	Underlying	Industry
ind_58	Industry code		Underlying	Underlying	Industry
ind_59	Industry code		Underlying	Underlying	Industry
ind_60	Industry code		Underlying	Underlying	Industry
ind_61	Industry code		Underlying	Underlying	Industry
ind_62	Industry code		Underlying	Underlying	Industry
ind_63	Industry code		Underlying	Underlying	Industry
ind_64	Industry code		Underlying	Underlying	Industry
ind_65	Industry code		Underlying	Underlying	Industry
ind_66	Industry code		Underlying	Underlying	Industry
ind_67	Industry code		Underlying	Underlying	Industry
ind_68	Industry code		Underlying	Underlying	Industry
ind_69	Industry code		Underlying	Underlying	Industry
ind_70	Industry code		Underlying	Underlying	Industry
ind_71	Industry code		Underlying	Underlying	Industry
ind_72	Industry code		Underlying	Underlying	Industry
ind_73	Industry code		Underlying	Underlying	Industry

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ind_74	Industry code		Underlying	Underlying	Industry
ind_75	Industry code		Underlying	Underlying	Industry
ind_76	Industry code		Underlying	Underlying	Industry
ind_77	Industry code		Underlying	Underlying	Industry
ind_78	Industry code		Underlying	Underlying	Industry
ind_79	Industry code		Underlying	Underlying	Industry
ind_80	Industry code		Underlying	Underlying	Industry
ind_81	Industry code		Underlying	Underlying	Industry
ind_82	Industry code		Underlying	Underlying	Industry
ind_83	Industry code		Underlying	Underlying	Industry
ind_84	Industry code		Underlying	Underlying	Industry
ind_85	Industry code		Underlying	Underlying	Industry
ind_86	Industry code		Underlying	Underlying	Industry
ind_87	Industry code		Underlying	Underlying	Industry
ind_88	Industry code		Underlying	Underlying	Industry
ind_89	Industry code		Underlying	Underlying	Industry
ind_90	Industry code		Underlying	Underlying	Industry
ind_91	Industry code		Underlying	Underlying	Industry
ind_92	Industry code		Underlying	Underlying	Industry
ind_93	Industry code		Underlying	Underlying	Industry
ind_94	Industry code		Underlying	Underlying	Industry
ind_95	Industry code		Underlying	Underlying	Industry
ind_96	Industry code		Underlying	Underlying	Industry
ind_97	Industry code		Underlying	Underlying	Industry
ind_98	Industry code		Underlying	Underlying	Industry
ind_99	Industry code		Underlying	Underlying	Industry
indmom	Industry momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
invest	Capital expenditures and inventory	Green et al. (2017)	Underlying	Underlying	Investment
iv	Implied volatility	Buchner and Kelly (2020)	Options	Contract	Contract
iv_rank	Implied volatility rank vs. last year		Options	Bucket	Past Prices
ivarud30	Option implied variance asymmetry	Huang and Li (2019)	Options	Underlying	Risk
ivd	Implied volatility duration	Schlag et al. (2020)	Options	Underlying	Risk
ivrv	Implied volatility minus realized volatility	Bali and Hovakimian (2009)	Options	Underlying	Risk

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
ivrv_ratio	Implied volatility minus realized volatility ratio		Options	Underlying	Risk
ivslope	Implied volatility slope	Vasquez (2017)	Options	Underlying	Risk
ivvol	Volatility of atm volatility	Baltussen et al. (2018)	Options	Underlying	Risk
ldso	Log changes in the stock to option volume	Roll et al. (2010)	Options	Underlying	Informed Trading
lev	Leverage	Green et al. (2017)	Underlying	Underlying	Quality
lgr	Growth in long-term debt	Green et al. (2017)	Underlying	Underlying	Quality
lso	Log of stock vs. option volume	Roll et al. (2010)	Options	Underlying	Informed Trading
maxret	Maximum daily return	Green et al. (2017)	Underlying	Underlying	Risk
mid	Option mid price		Options	Contract	Contract
modos	Modified stock vs. option volume	Johnson and So (2012)	Options	Underlying	Informed Trading
mom12m	12-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
mom1m	1-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
mom36m	36-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
mom6m	6-month momentum	Green et al. (2017)	Underlying	Underlying	Past Prices
moneyness	Moneyness		Options	Contract	Contract
ms	Financial statement score	Green et al. (2017)	Underlying	Underlying	Quality
mve	Size	Green et al. (2017)	Underlying	Underlying	Quality
mve_ia	Industry-adjusted size	Green et al. (2017)	Underlying	Underlying	Quality
nincr	Number of earnings increases	Green et al. (2017)	Underlying	Underlying	Quality
nopt	Number of options trading		Options	Underlying	Illiquidity
ocgo	Disposition Effect	Bergsma et al. (2020)	Options	Bucket	Past Prices
oi	Open interest		Options	Contract	Illiquidity
oistock	Open interest vs. stock volume		Options	Bucket	Informed Trading
operprof	Operating profitability	Green et al. (2017)	Underlying	Underlying	Quality
optspread	Option bid-ask spread		Options	Contract	Illiquidity
orgcap	Organization capital	Green et al. (2017)	Underlying	Underlying	Quality
P	Put-flag		Options	Contract	Contract
pba	Proportional bid-ask spread	Cao and Wei (2010)	Options	Underlying	Illiquidity
pchcapx_ia	Industry-adjusted % change in capital expenditures	Green et al. (2017)	Underlying	Underlying	Investment
pchcurrat	% change in current ratio	Green et al. (2017)	Underlying	Underlying	Quality
pchdepr	% change in depreciation	Green et al. (2017)	Underlying	Underlying	Quality
pchgm_pchsale	% change in gross margin - % change in sales	Green et al. (2017)	Underlying	Underlying	Quality
pchquick	% change in quick ratio	Green et al. (2017)	Underlying	Underlying	Quality

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
pchsale_pchinvt	% change in sales - % change in inventory	Green et al. (2017)	Underlying	Underlying	Profitability
pchsale_pchrect	% change in sales - % change in A/R	Green et al. (2017)	Underlying	Underlying	Profitability
pchsale_pchxsga	% change in sales - % change in SG&A	Green et al. (2017)	Underlying	Underlying	Profitability
pchsaleinv	% change in sales-to-inventory	Green et al. (2017)	Underlying	Underlying	Profitability
pcpv	Put-call parity deviations	Ofek et al. (2004)	Options	Underlying	Frictions
pcratio	Put-call ratio	Blau et al. (2014)	Options	Underlying	Informed Trading
pctacc	Percent accruals	Green et al. (2017)	Underlying	Underlying	Accruals
pfht	Modified illiquidity measure based on zero returns	Fong et al. (2017)	Options	Bucket	Illiquidity
pifht	An extended FHT measured based on zero returns		Options	Bucket	Illiquidity
pilliq	Percentage illiquidity	Cao and Wei (2010)	Options	Underlying	Illiquidity
piroll	Extended Roll's measure	Goyenko et al. (2009)	Options	Bucket	Illiquidity
pricedelay	Price delay	Green et al. (2017)	Underlying	Underlying	Illiquidity
ps	Financial statements score (Piotroski)	Green et al. (2017)	Underlying	Underlying	Quality
pzeros	Illiquidity measure based on zero returns	Lesmond et al. (1999)	Options	Bucket	Illiquidity
quick	Quick ratio	Green et al. (2017)	Underlying	Underlying	Quality
rd	R&D increase	Green et al. (2017)	Underlying	Underlying	Investment
rd_mve	R&D to market capitalization	Green et al. (2017)	Underlying	Underlying	Quality
rd_sale	R&D to sales	Green et al. (2017)	Underlying	Underlying	Quality
realestate	Real estate holdings	Green et al. (2017)	Underlying	Underlying	Quality
retvol	Return volatility	Green et al. (2017)	Underlying	Underlying	Risk
rnk182	182-day risk-neutral kurtosis		Options	Underlying	Risk
rnk273	273-day risk-neutral kurtosis		Options	Underlying	Risk
rnk30	30-day risk-neutral kurtosis		Options	Underlying	Risk
rnk365	365-day risk-neutral kurtosis		Options	Underlying	Risk
rnk91	91-day risk-neutral kurtosis		Options	Underlying	Risk
rns182	182-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns273	273-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns30	30-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns365	365-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
rns91	91-day risk-neutral skewness	Borochin et al. (2020)	Options	Underlying	Risk
roaq	Return on assets	Green et al. (2017)	Underlying	Underlying	Profitability
roavol	Earnings volatility	Green et al. (2017)	Underlying	Underlying	Quality
roeq	Return on equity	Green et al. (2017)	Underlying	Underlying	Profitability

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
roic	Return on invested capital	Green et al. (2017)	Underlying	Underlying	Profitability
roll	Roll's measure of illiquidity	Roll (1984)	Options	Bucket	Illiquidity
rsup	Revenue surprise	Green et al. (2017)	Underlying	Underlying	Profitability
rv	Realized variance	Cao et al. (2019)	Underlying	Underlying	Risk
salecash	Sales to cash	Green et al. (2017)	Underlying	Underlying	Value
saleinv	Sales to inventory	Green et al. (2017)	Underlying	Underlying	Value
salerec	Sales to receivables	Green et al. (2017)	Underlying	Underlying	Value
season1	Seasonal return - 1 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
season2	Seasonal return - 2 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
season3	Seasonal return - 3 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
season4	Seasonal return - 4 year historical	Heston and Sadka (2008); Keloharju et al. (2016)	Underlying	Underlying	Past Prices
secured	Secured debt	Green et al. (2017)	Underlying	Underlying	Quality
securedind	Secured debt indicator	Green et al. (2017)	Underlying	Underlying	Quality
sgr	Sales growth	Green et al. (2017)	Underlying	Underlying	Investment
shrtfee	Implied shorting fees	Muravyev et al. (2021)	Options	Underlying	Frictions
sin	Sin stocks	Green et al. (2017)	Underlying	Underlying	Quality
skewiv	IV skew	Xing et al. (2010)	Options	Underlying	Informed Trading
so	Stock vs. option volume	Roll et al. (2010)	Options	Underlying	Informed Trading
sp	Sales to price	Green et al. (2017)	Underlying	Underlying	Quality
std_dolvol	Volatility of liquidity (dollar trading volume)	Green et al. (2017)	Underlying	Underlying	Illiquidity
std_turn	Volatility of liquidity (share turnover)	Green et al. (2017)	Underlying	Underlying	Illiquidity
stdacc	Accrual volatility	Green et al. (2017)	Underlying	Underlying	Accruals
stdamihud	Standard deviation of Amihud's illiquidity measure		Options	Bucket	Illiquidity
stdcf	Cash flow volatility	Green et al. (2017)	Underlying	Underlying	Risk
tang	Dept capacity/firm tangibility	Green et al. (2017)	Underlying	Underlying	Quality
tb	Tax income to book income	Green et al. (2017)	Underlying	Underlying	Past Prices
theta	Theta	Buchner and Kelly (2020)	Options	Contract	Risk
tlm30	Tail loss measure	Vilkov and Xiao (2012)	Options	Underlying	Risk
toi	Total option open interest		Options	Underlying	Illiquidity
ttm	Time-to-maturity		Options	Contract	Contract
turn	Share turnover	Green et al. (2017)	Underlying	Underlying	Risk
turnover	Option turnover		Options	Bucket	Illiquidity
underlying_return	Return of the underlying		Underlying	Underlying	Frictions

Continued on Next Page

Table IA4.1 from previous page

Feature	Description	Source	Information Source	Instrument Level	Group
vega	Vega	Buchner and Kelly (2020)	Options	Contract	Risk
vol	Trading volume in options		Options	Underlying	Illiquidity
volga	Volga	Buchner and Kelly (2020)	Options	Contract	Risk
volunc	Volatility uncertainty	Cao et al. (2019)	Options	Underlying	Risk
vs_change	Change in weighted put-call spread	Cremers and Weinbaum (2010)	Options	Underlying	Informed Trading
vs_level	Weighted put-call spread	Cremers and Weinbaum (2010)	Options	Underlying	Informed Trading
zerotrade	Zero trading days	Green et al. (2017)	Underlying	Underlying	Illiquidity

Not Continued on Next Page

The table provides a detailed summary of the characteristics. For each characteristic, the table shows the name (Feature), a description of the characteristic (Description) and the source, if applicable (Source). Furthermore, it displays if the characteristic is derived form option-based or stock-based information (Information Source), the instrument level the characteristic relates to (Instrument Level), and to which characteristic group (Group) it belongs to.

Appendix IA5. Additional Summary Statistics

IA5.1. Summary Statistics for the Underlying Stocks

	Mean	Std	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: Time-Series Distribution							
Number of stocks in the sample each month	1747.31	142.84	1560.1	1664.0	1755.5	1850.75	1922.2
Stock coverage of stock universe (EW)	38.05	10.09	24.88	28.07	40.04	46.64	50.98
Stock coverage of stock universe (VW)	83.81	5.0	77.26	81.25	84.63	86.98	89.63
Stock traded at NYSE or AMEX	51.86	2.36	48.74	50.95	52.01	53.32	54.35
Stock already included in previous month	86.11	5.87	82.36	84.31	86.56	88.74	90.25
Panel B: Time-Series Average of Cross-Sectional Distributions							
Firm size in million	7945	27163	240	550	1516	4793	15532
Firm size CSRP percentile	72	18	45	59	75	87	94
Firm volatility CSRP percentile	45	25	11	23	44	66	81

Panel C: Time-Series Average of Industry Distribution

FF-12 Industry	Optionable Stocks	CRSP sample	FF-12 Industry	Optionable Stocks	CRSP sample
Consumer nondurables	4.48%	4.92%	Telecom	3.47%	2.56%
Consumer durables	0.52%	0.47%	Utilities	2.74%	2.21%
Manufacturing	9.34%	8.15%	Wholesale	11.11%	8.49%
Energy	4.92%	2.68%	Healthcare	11.4%	9.24%
Chemicals	2.45%	1.59%	Finance	10.08%	24.08%
Business Equipment	20.22%	14.44%	Other	19.28%	21.16%

Table IA5.1: 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.

IA5.2. Delta-Hedged Option Return per Bucket

Table IA5.2: Delta-Hedged Option Return per Bucket

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel A: Long Term (N=6,903,957)							
Delta-Hedged Return	0.15	5.03	-4.15	-1.84	-0.27	1.47	4.73
Days to Maturity	269.01	185.56	112.0	141.0	200.0	324.0	570.0
Moneyness	1.04	0.38	0.71	0.86	1.0	1.15	1.36
Implied volatility	45.85	23.08	23.55	29.88	40.13	55.64	75.54
Absolute Delta	0.44	0.24	0.13	0.25	0.43	0.63	0.79
Panel B: Long Term Atm (N=2,495,362)							
Delta-Hedged Return	0.2	3.63	-2.85	-1.41	-0.22	1.24	3.6
Days to Maturity	241.32	167.94	112.0	141.0	176.0	234.0	506.0
Moneyness	1.0	0.06	0.93	0.96	1.0	1.05	1.08
Implied volatility	38.96	18.75	20.94	26.02	34.42	46.64	62.97
Absolute Delta	0.5	0.12	0.33	0.4	0.5	0.6	0.66
Panel C: Long Term Itm Call (N=865,963)							
Delta-Hedged Return	0.14	3.91	-2.99	-1.25	-0.14	1.09	3.46
Days to Maturity	288.93	198.5	113.0	141.0	203.0	386.0	628.0
Moneyness	0.77	0.12	0.59	0.71	0.8	0.86	0.89
Implied volatility	52.16	22.33	29.0	36.08	47.1	63.07	82.1
Absolute Delta	0.81	0.07	0.72	0.75	0.81	0.87	0.92

Continued on Next Page

Table IA5.2 from previous page

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel D: Long Term Itm Put (N=639,054)							
Delta-Hedged Return	0.11	2.45	-1.98	-0.85	-0.09	0.8	2.38
Days to Maturity	280.43	192.15	113.0	141.0	203.0	355.0	598.0
Moneyness	1.42	0.73	1.13	1.16	1.25	1.44	1.79
Implied volatility	54.63	31.31	25.24	33.16	46.53	67.18	93.66
Absolute Delta	0.69	0.15	0.5	0.59	0.69	0.8	0.89
Panel E: Long Term Otm Call (N=1,528,873)							
Delta-Hedged Return	0.25	7.3	-6.89	-3.29	-0.44	2.96	8.14
Days to Maturity	285.64	193.94	113.0	143.0	204.0	381.0	603.0
Moneyness	1.36	0.35	1.13	1.17	1.26	1.42	1.67
Implied volatility	47.24	23.26	23.87	30.61	41.46	58.08	78.48
Absolute Delta	0.3	0.14	0.12	0.19	0.29	0.4	0.48
Panel F: Long Term Otm Put (N=1,374,705)							
Delta-Hedged Return	-0.03	5.58	-5.05	-2.71	-0.7	1.65	5.36
Days to Maturity	282.94	188.97	113.0	142.0	204.0	358.0	598.0
Moneyness	0.74	0.13	0.55	0.67	0.78	0.85	0.88
Implied volatility	48.75	22.56	27.2	33.19	43.05	58.12	77.45
Absolute Delta	0.16	0.08	0.05	0.1	0.16	0.23	0.27
Panel G: Short Term (N=5,225,887)							
Delta-Hedged Return	-0.21	5.36	-4.65	-2.07	-0.43	1.04	4.1
Days to Maturity	44.39	23.02	17.0	21.0	49.0	52.0	80.0
Moneyness	1.01	0.18	0.84	0.92	1.0	1.07	1.17
Implied volatility	50.08	28.44	23.33	30.7	42.74	61.37	86.07
Absolute Delta	0.48	0.26	0.14	0.26	0.47	0.69	0.84
Panel H: Short Term Atm (N=3,242,014)							
Delta-Hedged Return	-0.11	4.08	-3.67	-1.79	-0.4	0.99	3.5
Days to Maturity	40.22	22.35	17.0	19.0	46.0	51.0	79.0
Moneyness	1.0	0.05	0.93	0.96	1.0	1.04	1.07
Implied volatility	41.61	21.33	21.09	26.94	36.37	50.33	68.81
Absolute Delta	0.49	0.19	0.24	0.35	0.49	0.63	0.74

Continued on Next Page

Table IA5.2 from previous page

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Panel I: Short Term Itm Call (N=471,757)							
Delta-Hedged Return	0.19	5.14	-2.72	-1.15	-0.23	0.61	2.71
Days to Maturity	49.39	22.92	18.0	22.0	50.0	77.0	81.0
Moneyness	0.82	0.08	0.72	0.79	0.85	0.88	0.9
Implied volatility	65.11	30.15	34.89	43.82	58.16	79.04	103.74
Absolute Delta	0.85	0.07	0.75	0.8	0.85	0.9	0.94
Panel J: Short Term Itm Put (N=382,619)							
Delta-Hedged Return	-0.41	3.87	-2.59	-1.14	-0.31	0.46	1.96
Days to Maturity	47.64	23.03	18.0	22.0	50.0	77.0	80.0
Moneyness	1.29	0.34	1.12	1.14	1.19	1.31	1.53
Implied volatility	72.13	41.5	33.13	43.58	61.24	88.33	124.21
Absolute Delta	0.8	0.11	0.64	0.72	0.81	0.89	0.94
Panel K: Short Term Otm Call (N=537,456)							
Delta-Hedged Return	-0.32	9.09	-9.78	-5.0	-0.96	3.36	9.64
Days to Maturity	54.05	21.64	21.0	47.0	50.0	78.0	81.0
Moneyness	1.23	0.17	1.12	1.14	1.18	1.26	1.4
Implied volatility	61.01	30.5	30.14	38.92	53.55	76.05	101.68
Absolute Delta	0.21	0.11	0.08	0.13	0.2	0.29	0.36
Panel L: Short Term Otm Put (N=592,041)							
Delta-Hedged Return	-0.85	7.47	-7.99	-4.29	-1.43	1.42	6.3
Days to Maturity	52.39	22.06	19.0	46.0	50.0	78.0	81.0
Moneyness	0.82	0.09	0.7	0.78	0.84	0.88	0.89
Implied volatility	60.34	29.45	31.8	39.81	52.79	73.11	98.45
Absolute Delta	0.14	0.07	0.05	0.08	0.13	0.19	0.24

Not Continued on Next Page

The table reports the descriptive statistics of delta-hedged option returns for the period 1996 to 2020. Delta-hedged option returns are measured over a period of one calendar month, or until option maturity. Delta-hedging is performed daily. Days to maturity is the number of calendar days until option expiration. Moneyness is the ratio between the underlying's stock price and the option's strike price. Option implied volatility is provided by OptionMetrics. Absolute delta is the absolute value of the Black-Scholes delta. We differentiate between different parts of the time-to-maturity and moneyness domain of a single option, which we refer to as "buckets", as defined in Section 4. Specifically, we separately consider predictability for short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (itm: $K/S > 1.1$ for puts, $K/S < 0.9$ for calls), out-of-the-money (otm: $K/S < 0.9$ for puts, $K/S < 1.1$ for calls) calls and puts, and at-the-money options (atm: $0.9 \leq K/S \leq 1.1$), as well as time-to-maturity and moneyness combinations. Each panel shows statistics for pooled options belonging to one bucket.

IA5.3. Number of Options per Bucket

	Mean	Sd	10-Pctl	Q1	Median	Q3	90-Pctl
Long Term	14.98	26.69	1.05	2.38	5.86	16.48	39.51
Long Term Atm	6.47	8.98	1.0	1.69	3.31	7.78	15.89
Long Term Itm Call	3.58	5.29	1.0	1.0	1.9	3.99	7.91
Long Term Itm Put	3.33	4.49	1.0	1.0	1.83	3.68	7.51
Long Term Otm Call	4.87	7.09	1.0	1.08	2.45	5.75	11.41
Long Term Otm Put	5.34	8.57	1.0	1.03	2.4	6.16	12.76
Short Term	10.93	15.89	1.0	2.4	5.83	13.35	26.48
Short Term Atm	7.38	8.84	1.0	2.28	4.44	9.36	17.38
Short Term Itm Call	2.4	2.77	1.0	1.0	1.42	2.7	4.74
Short Term Itm Put	2.43	2.79	1.0	1.0	1.37	2.68	4.9
Short Term Otm Call	2.69	3.34	1.0	1.0	1.54	3.03	5.47
Short Term Otm Put	3.08	4.17	1.0	1.0	1.74	3.48	6.42

Table IA5.3: Number of Options for Option Buckets

The table reports summary statistics on the number of options within certain regions of the time-to-maturity and moneyness domain, denoted by "buckets", as defined in Section 4. The time-to-maturity and moneyness domain is divided into short- and long-term options ($\leq$ vs. $>$ 90 days to maturity), in-the-money (itm: $K/S > 1.1$ for puts, $K/S < 0.9$ for calls) out-of-the-money (otm: $K/S < 0.9$ for puts, $K/S < 1.1$ for calls) calls and puts, and at-the-money options (atm: $0.9 \leq K/S \leq 1.1$), as well as time-to-maturity and moneyness combinations. We first compute descriptive statistics for each underlying stock and subsequently take the average across all stocks in the sample. Values here correspond to the number of options *per underlying stock*. The sample period is from January 1996 to December 2020.

Appendix IA6. Model Comparison

Additional Analyses

IA6.1. 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	1.47	1.38	−0.93	**-4.86**	**3.25**	**8.94**	**5.63**	**7.45**	**10.02**	**10.23**
Lasso		−0.27	**-3.11**	**-5.21**	**3.88**	**4.53**	**3.91**	**3.55**	**5.09**	**6.10**
ENet			**-3.27**	**-4.95**	**3.77**	**4.68**	**4.14**	**3.59**	**5.19**	**6.19**
PCR				**-2.56**	**6.06**	**6.22**	**6.49**	**5.14**	**6.08**	**8.04**
PLS					**6.28**	**10.74**	**8.06**	**9.49**	**11.21**	**11.22**
L-En						**3.64**	**2.86**	**2.92**	**3.47**	**5.97**
GBR							**-3.47**	0.52	**-1.98**	**4.56**
RF								**2.18**	1.13	**8.22**
Dart									−1.36	**2.05**
FFN										**4.56**

The table shows Diebold and Mariano (1995) test statistics following Equation (8), using cross-sectional errors as inputs, for the nine models and two ensembles considered in the paper. A positive number indicates that the model in the column outperforms the row model. If it is highlighted in light blue (blue), this outperformance is statistically significant at the 1% level (5% level).

IA6.2. Cross-sectional Comparison between N-En and L-En

N-En beats L-En in 86.5% of the months when tasked with predicting future cross-sectional return spreads for delta-hedged single-equity options:

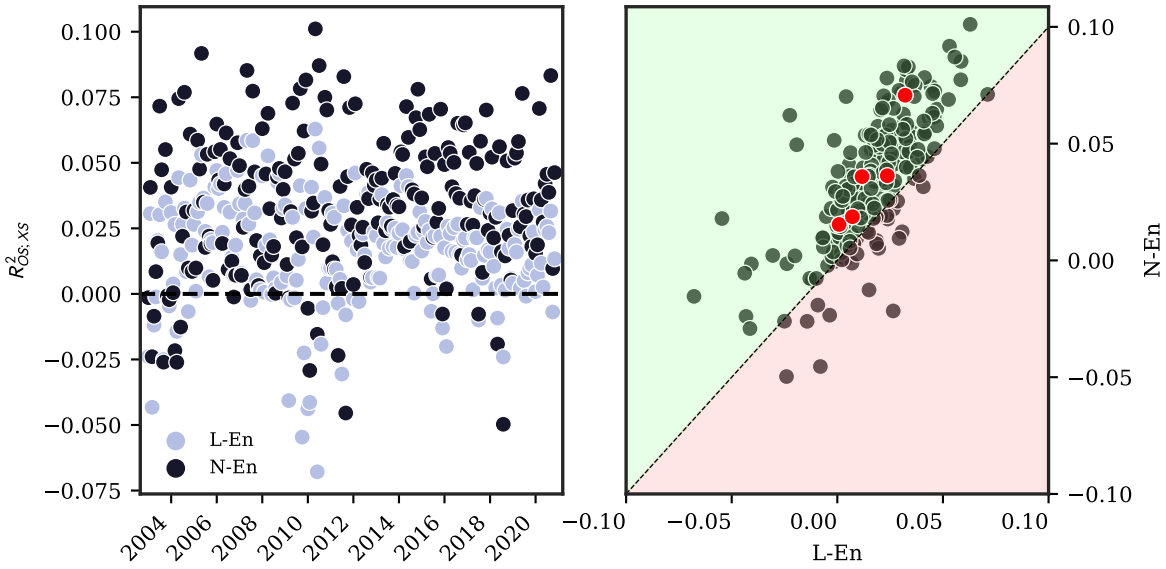

Fig. IA6.1. Comparing Linear and Nonlinear Ensembles – $R^2_{OS;XS}$

The left panel of the figure shows monthly cross-sectional $R^2_{OS;XS}$ for the testing sample from 2003 through 2020 for the linear (L-En) and nonlinear (N-En) ensembles. The right panel compares the two by showing the resulting $R^2_{OS;XS}$ for L-En on the x-axis and for N-En and the y axis. The green-shaded area represents a relative outperformance in terms of predictability for N-En, while the red-shaded area represents the opposite. The five red circles represent the Coronavirus selloff from December 2019 through April 2020.

Appendix IA7. Machine Learning Portfolios

Additional Analyses

IA7.1. Performance for ML Portfolios using Value-weighted Returns

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Lo	−1.308	−0.967	1.335	−0.725	−1.612	−1.556	1.845	−0.843	***
2	−0.722	−0.482	1.450	−0.332	−0.709	−0.637	1.564	−0.408	
3	−0.491	−0.319	1.434	−0.222	−0.397	−0.393	1.393	−0.282	
4	−0.321	−0.220	1.432	−0.153	−0.223	−0.220	1.272	−0.173	
5	−0.178	−0.154	1.520	−0.101	−0.098	−0.125	1.260	−0.099	
6	−0.047	−0.096	1.488	−0.065	0.006	−0.077	1.321	−0.059	
7	0.082	−0.026	1.485	−0.017	0.109	−0.038	1.444	−0.027	
8	0.219	0.027	1.475	0.018	0.228	0.042	1.494	0.028	
9	0.385	0.084	1.450	0.058	0.393	0.135	1.607	0.084	
Hi	0.695	0.314	1.580	0.199	0.854	0.460	1.852	0.248	
H-L	2.004	1.281	1.240	1.033	2.466	2.016	1.466	1.375	***
		(11.52)		(8.64)		(14.28)		(9.57)	

Table IA7.1: Trading on Value-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 by the respective option's dollar open interest. 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.

IA7.2. Summary Statistics for ML Portfolios – Put and Call Composition

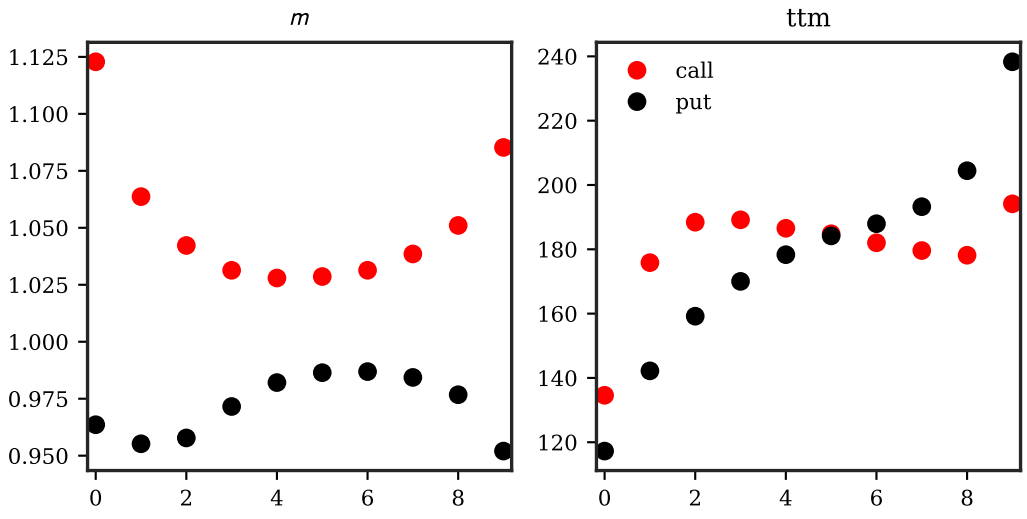

Fig. IA7.1. Machine Learning Portfolios – Moneyness and Days-to-maturity

This figure shows the average moneyness and time-to-maturity for the decile portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We split the information by put and call options included.

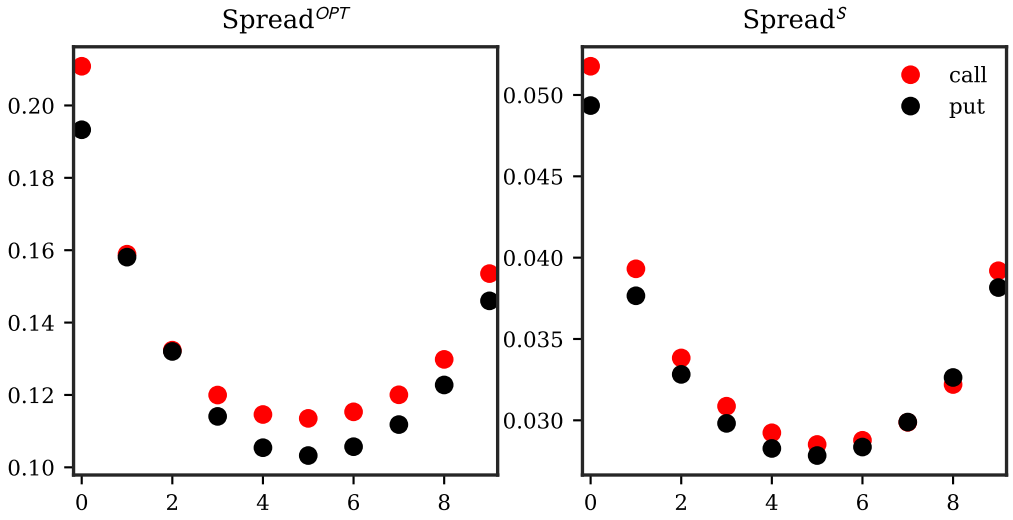

Fig. IA7.2. Machine Learning Portfolios – Spreads of the Options and the Underlyings

This figure shows the bid-ask spread of the options included in the left and of the underlying stocks in the right panel for the decile portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We split the information by put and call options included.

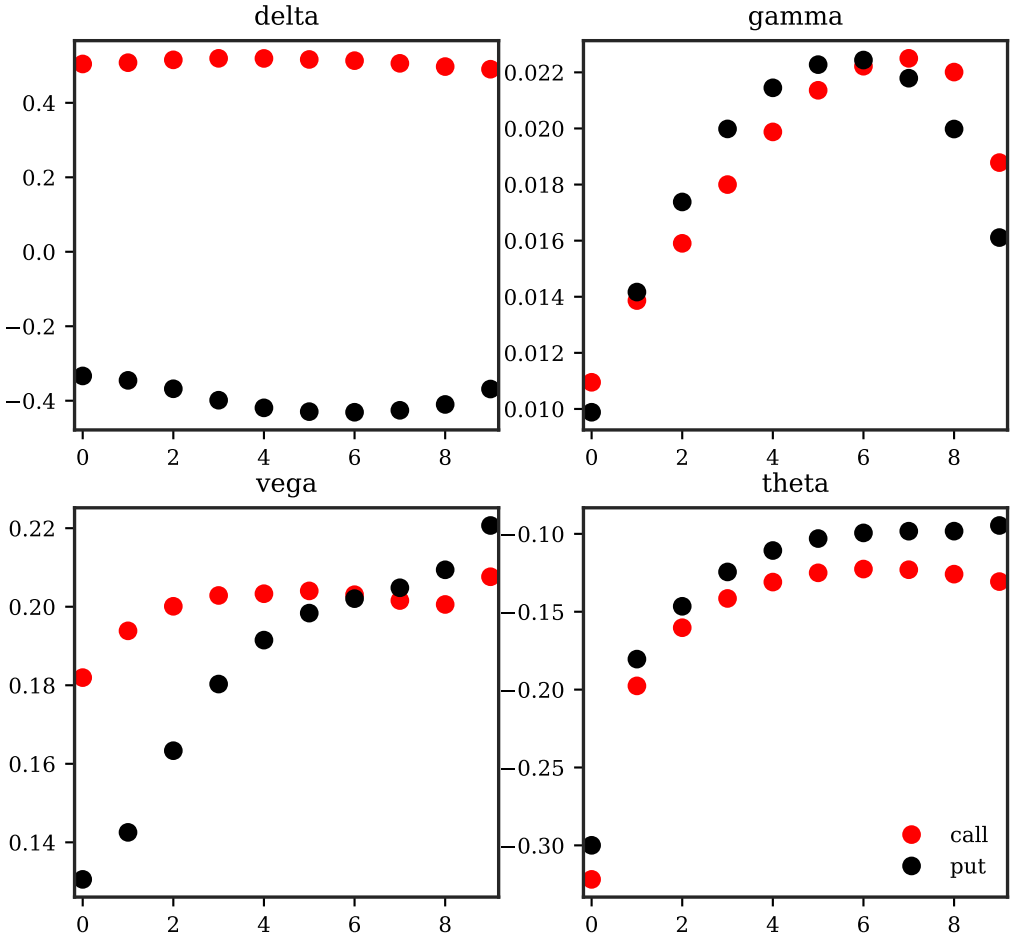

Fig. IA7.3. Machine Learning Portfolios – Greeks

This figure shows option Greeks for the decile portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We split the information by put and call options included. We show the (unhedged) delta of the option, the *gamma*, *vega*, and *theta*. *gamma* is expressed for a 1% move in the underlying stock ($gamma \times \frac{S}{100}$) and *vega* and *theta* in terms of the underlying stock price ($\frac{x}{S}$ for $x \in [vega, theta]$).

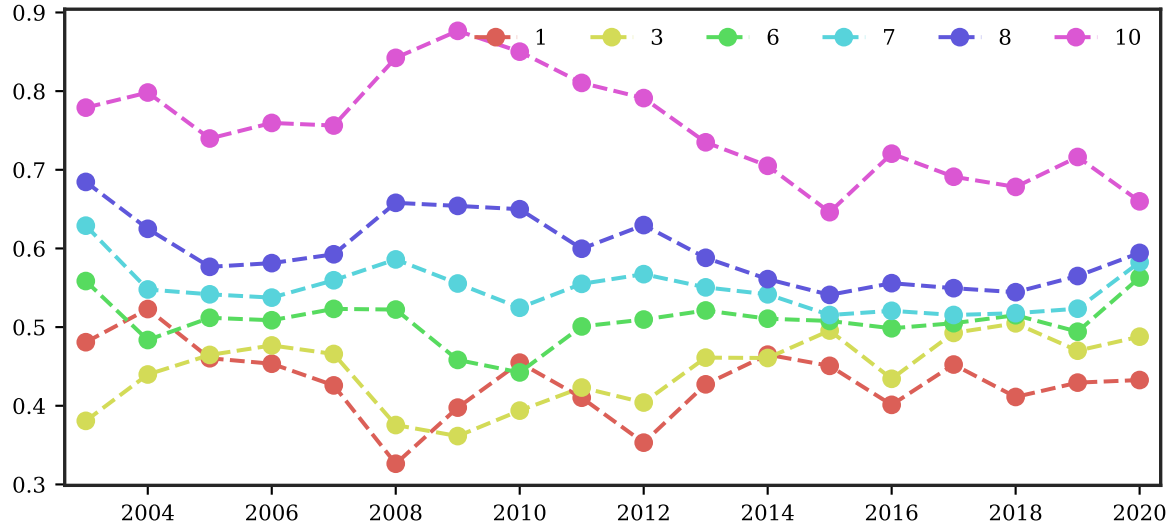

Fig. IA7.4. Machine Learning Portfolios – Call Share over Time

The figure shows the share of call options included in the portfolios sorted on expected returns following the predictions of the nonlinear ensemble N-En. We provide average numbers per year in our testing sample from 2003 through 2020.

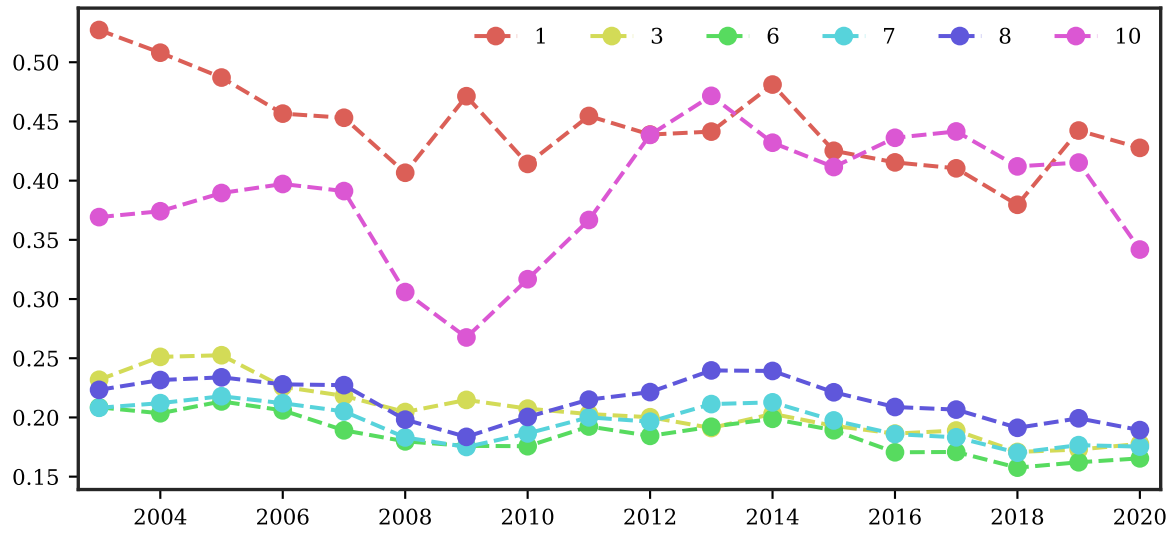

Fig. IA7.5. Machine Learning Portfolio – Underlying-Option Concentration

The figure shows the share of underlying stocks for which all options written on that stocks are classified into a single portfolio by the nonlinear ensemble N-En. We provide average numbers per year in our testing sample from 2003 through 2020.

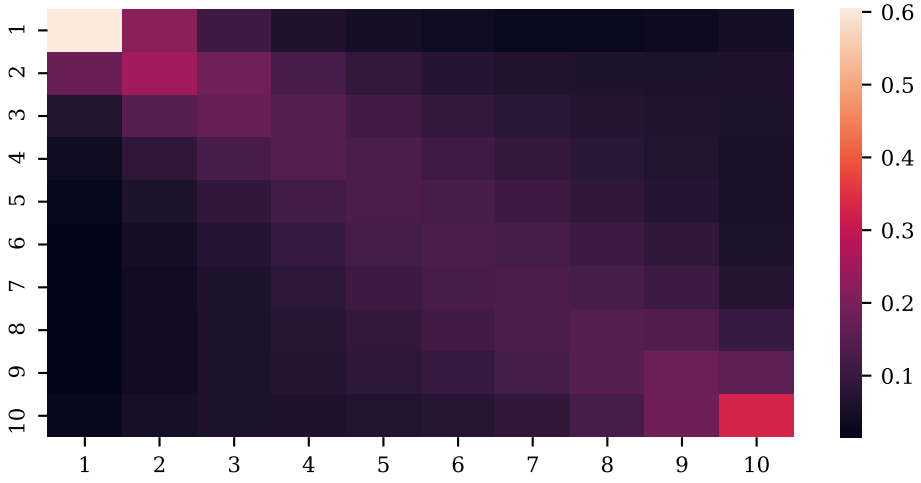

Fig. IA7.6. ML Portfolio Transition Matrix by Option Bucket

The figure shows the relative likelihood of options for a particular underlying transitioning from one portfolio to another in the next month. Since we cannot estimate this transition for single options due to their fleeting moneyness and time-to-maturity, we use changes in the portfolio mode for a given Permno-bucket combination as an approximation. Buckets are defined as in Section 4.3.

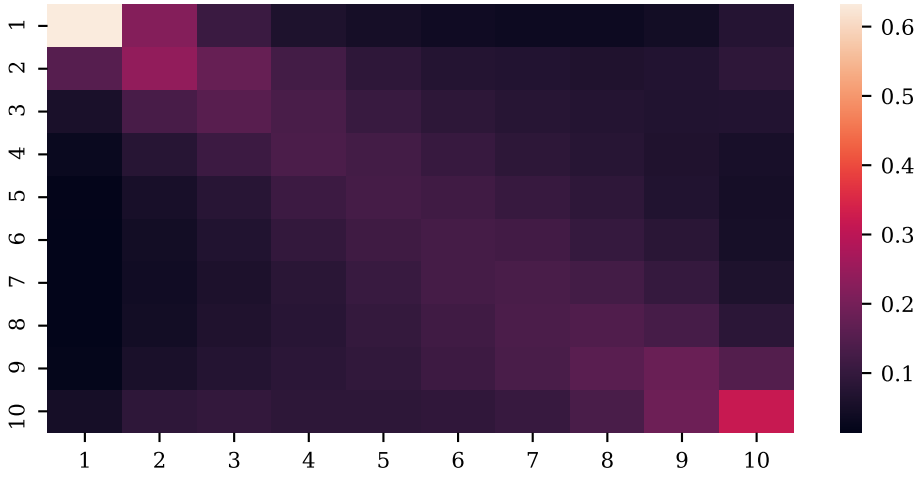

Fig. IA7.7. ML Portfolio Transition Matrix by Underlying

The figure shows the relative likelihood of options for a particular underlying transitioning from one portfolio to another in the next month. Since we cannot estimate this transition for single options due to their fleeting moneyness and time-to-maturity, we use changes in the portfolio mode for a given Permno as an approximation.

IA7.3. Performance for ML Portfolios in Different Market Phases

	L-En				N-En				
	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
Low VIX	1.689	1.725	0.584	2.953	2.428	2.321	0.726	3.197	***
		(21.03)		(7.28)		(26.34)		(7.24)	
High VIX	2.181	2.112	1.634	1.293	2.683	2.936	1.598	1.838	***
		(8.99)		(7.36)		(10.91)		(9.59)	
Low EPU	2.007	1.876	0.750	2.503	2.672	2.508	0.827	3.034	***
		(21.15)		(5.84)		(23.64)		(9.00)	
High EPU	1.860	1.959	1.587	1.234	2.437	2.747	1.599	1.718	***
		(8.49)		(7.37)		(10.01)		(8.78)	
Neg. CFNAI	1.904	1.931	1.430	1.350	2.455	2.696	1.469	1.835	***
		(9.29)		(5.61)		(11.63)		(7.64)	
Pos. CFNAI	1.967	1.902	0.982	1.938	2.667	2.550	1.012	2.520	***
		(15.33)		(4.48)		(14.93)		(5.44)	
Low FED Stress	1.832	1.764	0.854	2.065	2.534	2.418	0.942	2.567	***
		(19.30)		(5.17)		(19.35)		(6.03)	
High FED Stress	2.160	2.256	1.780	1.267	2.602	3.090	1.724	1.792	***
		(7.79)		(6.35)		(10.01)		(7.95)	
Low SENT	2.060	1.780	1.170	1.521	2.603	2.464	1.180	2.089	***
		(13.77)		(4.49)		(16.10)		(5.42)	
High SENT	1.860	2.159	0.946	2.281	2.578	2.683	1.045	2.568	**
		(15.78)		(7.41)		(10.92)		(9.22)	

Table IA7.2: Trading on Machine Learning Predictions – Market Phases

The table shows returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods for different sample splits capturing economic states. We split the sample by the median VIX from 2003 through 2020, the median EPU index by Baker, Bloom, and Davis (2016), by the sign of the Chicago Fed National Activity Index, the median of the St. Louis Fed Stress Index, and the median of the sentiment index proposed by Baker and Wurgler (2006). Pred denotes the average predicted return within the respective portfolio, Avg the average realized return, SD the standard deviation of realized returns and finally SR the realized Sharpe ratio. All values are given per month. The last column (N vs. L) gives the significance of comparing the mean realized returns for N-En and L-En. ***, **, * correspond to N-En beating L-En significantly at the 1%, 5%, 10% level, respectively.

IA7.4. Profitability of Machine Learning Portfolios Per Bucket

			L-En				N-En				
TTM	Mon.	Type	Pred	Avg	SD	SR	Pred	Avg	SD	SR	N vs. L
	atm		1.761	1.556	0.831	1.874	2.207	1.942	0.805	2.413	***
	itm	C	1.960	1.614	1.159	1.393	3.099	2.352	1.226	1.919	***
$\tau \leq 90$	itm	P	2.284	1.468	0.940	1.562	2.580	2.255	1.354	1.666	***
	otm	C	2.055	3.430	2.412	1.422	3.655	4.013	2.611	1.537	**
	otm	P	1.708	2.967	2.453	1.210	3.282	4.121	2.783	1.481	***
	atm		1.705	1.367	0.983	1.390	1.927	1.950	0.995	1.959	***
	itm	C	1.744	1.856	1.218	1.524	2.385	2.804	1.443	1.943	***
$\tau > 90$	itm	P	2.085	0.918	0.890	1.030	2.036	1.391	0.992	1.402	***
	otm	C	1.897	2.554	1.947	1.312	2.798	3.468	1.996	1.737	***
	otm	P	1.649	1.951	1.574	1.239	2.405	2.660	1.779	1.495	***

Table IA7.3: Trading on Machine Learning Predictions – Buckets

The table shows the returns to option portfolios sorted by the predictions made by the linear (L-En) and nonlinear ensemble (N-En) methods for the option buckets defined in Section 4 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.

*IA7.5. Impact of Spreads on Profitability (**N-En**) – Option Buckets*

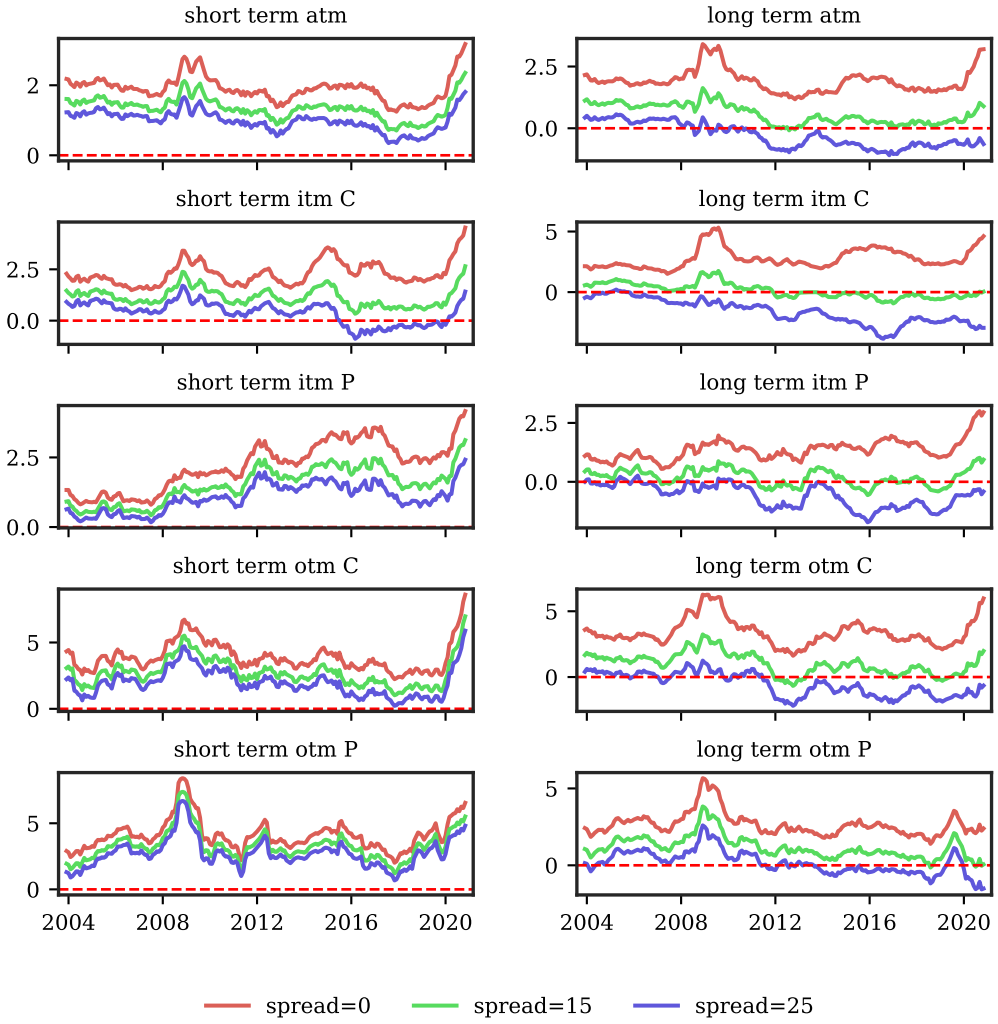

Fig. IA7.8. Rolling One-year Returns and Effective Spreads

The figure shows rolling cumulative returns over 252 trading days for the high-minus-low portfolio following the predictions made by N-En for the different option buckets defined in Section 4. We compare the resulting profitability for zero effective spreads (trading at the mid price), as well as effective spreads of 15% and 25%.

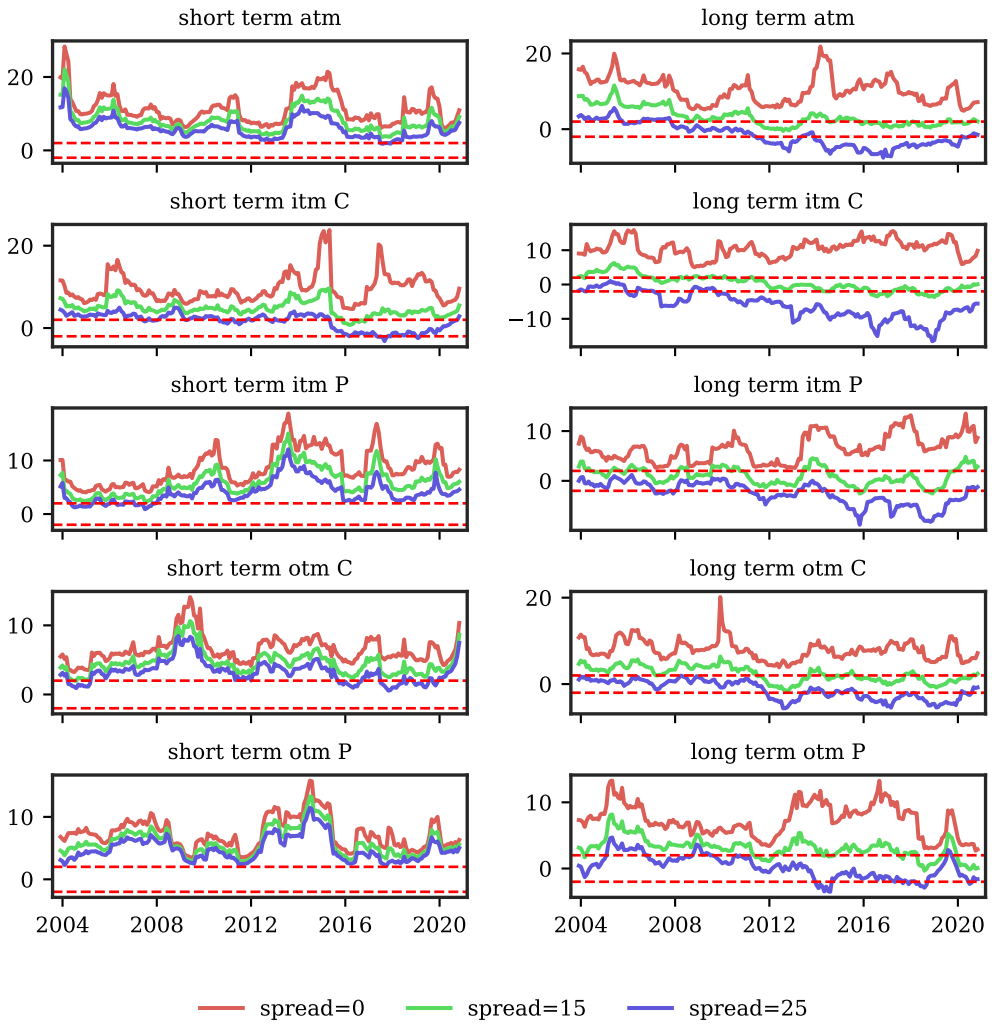

Fig. IA7.9. Rolling One-year t-statistics and Effective Spreads

The figure shows rolling Newey and West (1987) t-statistics with a lag of twelve months over 252 trading days for the high-minus-low portfolios following the predictions made by N-En for the different option buckets defined in Section 4. We compare the resulting profitability for zero effective spreads (trading at the mid price), as well as effective spreads of 15% and 25%.

Appendix IA8. Which Characteristics Matter?

Additional Analyses

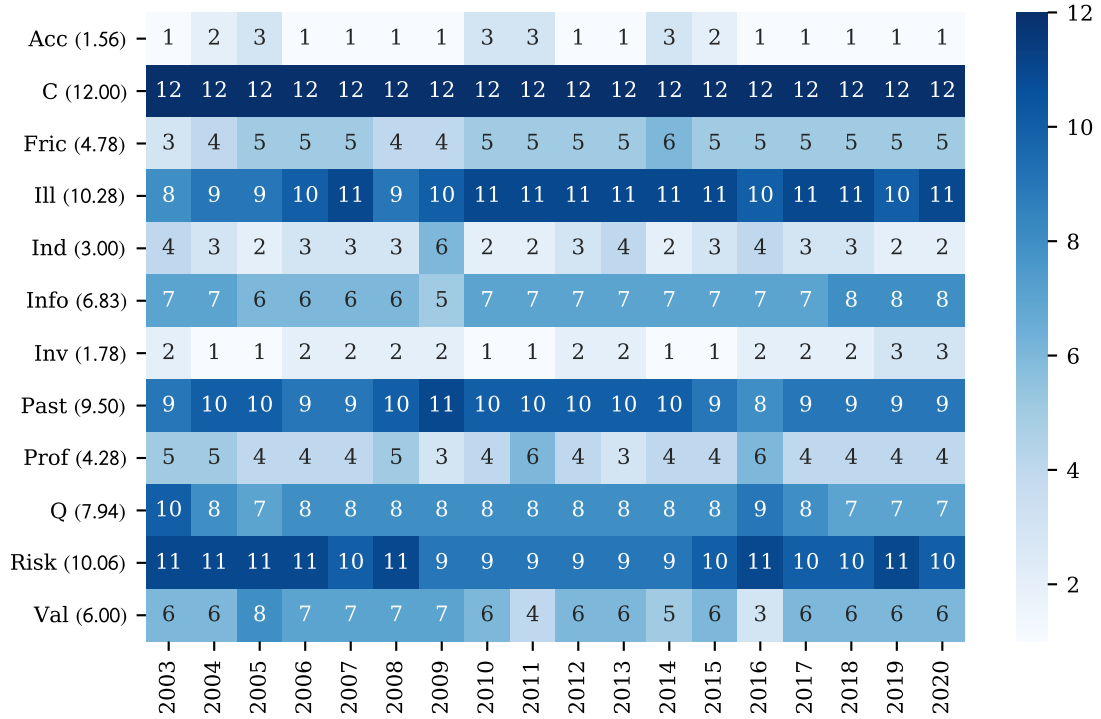

Fig. IA8.1. Feature Group Importance for N-En Over Time

The figure shows the time evolution of ranking the twelve feature groups defined in Appendix IA4 by their importance for the nonlinear ensemble (N-En). 1 denotes the lowest-ranking feature group, 12 the highest-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.

Appendix IA9. Impact of the Information Set

Additional Analyses

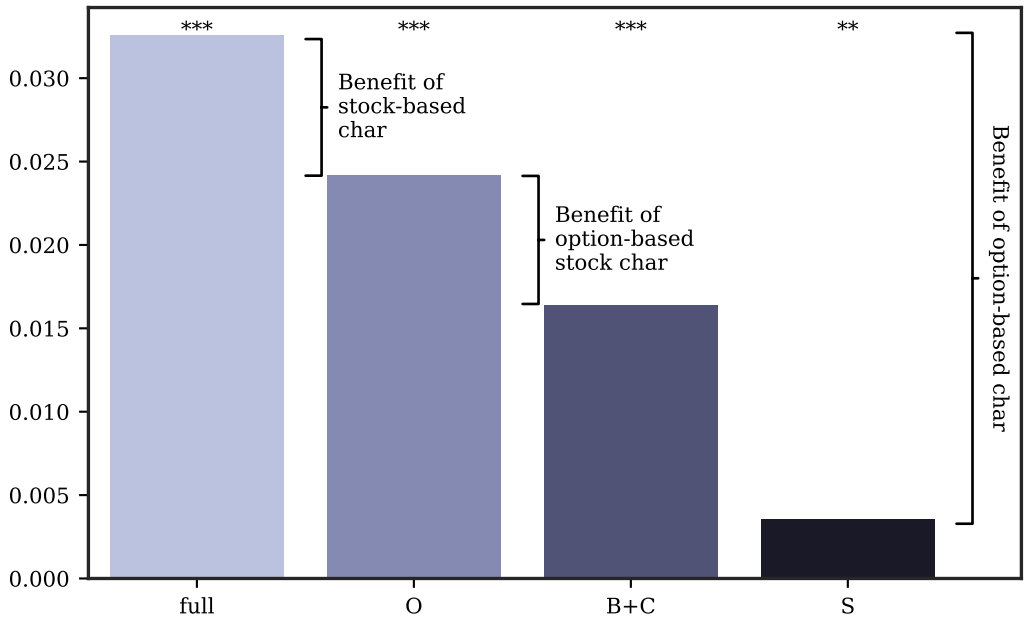

Fig. IA9.1. Restricting the Information Set for N-En – $R^2_{OS;XS}$

The figure shows the cross-sectional out-of-sample $R^2_{OS;XS}$ defined in Equation (5) for N-En with restricted access to the full set of characteristics. The full model is shown in the left bar for reference, and is compared with models using all option-based information (O), models using only bucket- and contract-based information (B+C) and models using only stock-based information (S). The distinction of the information source is provided in Appendix IA4. ***, **, * 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.

Appendix IA10. Sources of Option Return Predictability

Additional Analyses

IA10.1. Presence of Informed Investors

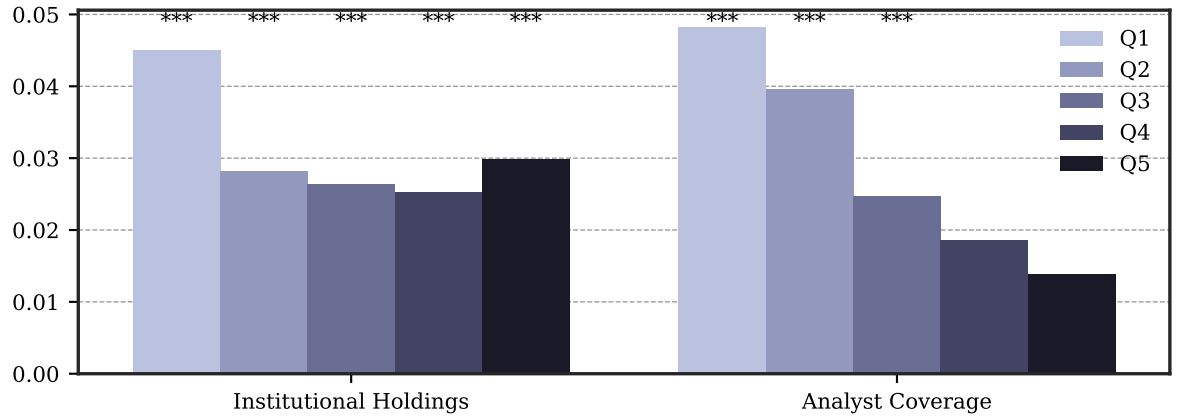

Fig. IA10.1. Predictability and Profitability Conditional on Informed Investor Presence – $R^2_{OS;XS}$

The left panel of the figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En for different quintiles of institutional ownership and analyst coverage of the underlying stocks. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7). The right panel shows the resulting Sharpe ratios of buying options in the highest predicted return decile and selling options in the lowest for each subsample.

IA10.2. Option Demand

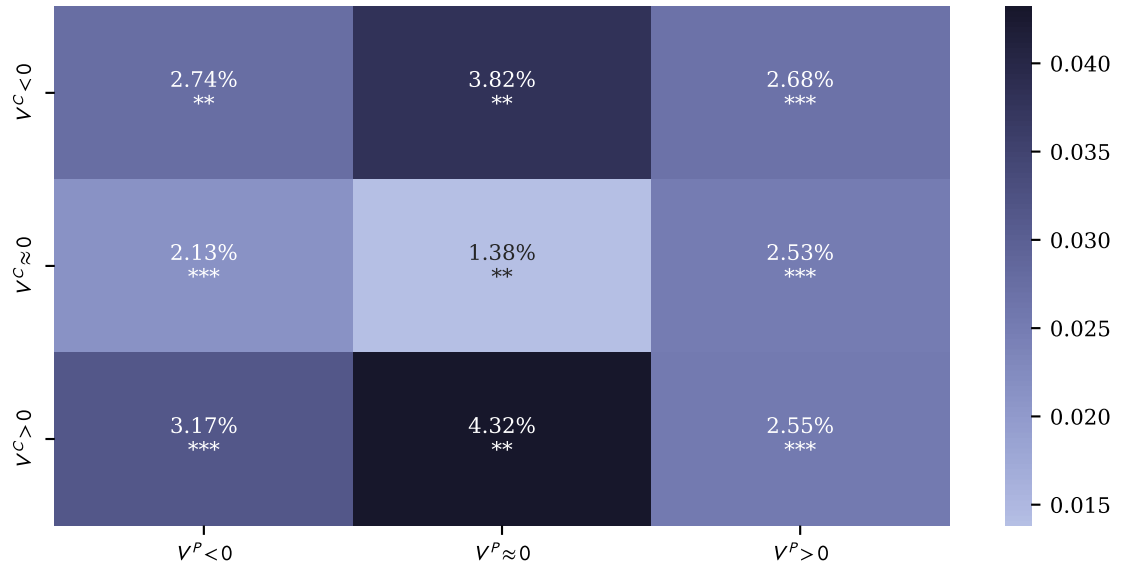

Fig. IA10.2. Predictability Conditional on Option Demand – $R^2_{OS;XS}$

The figure shows the cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En conditional on the net open interest of retail and institutional investors, I. We independently form terciles based on the cross-sectional distribution of I^R and I^I. ***, **, * denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

IA10.3. Stock Illiquidity

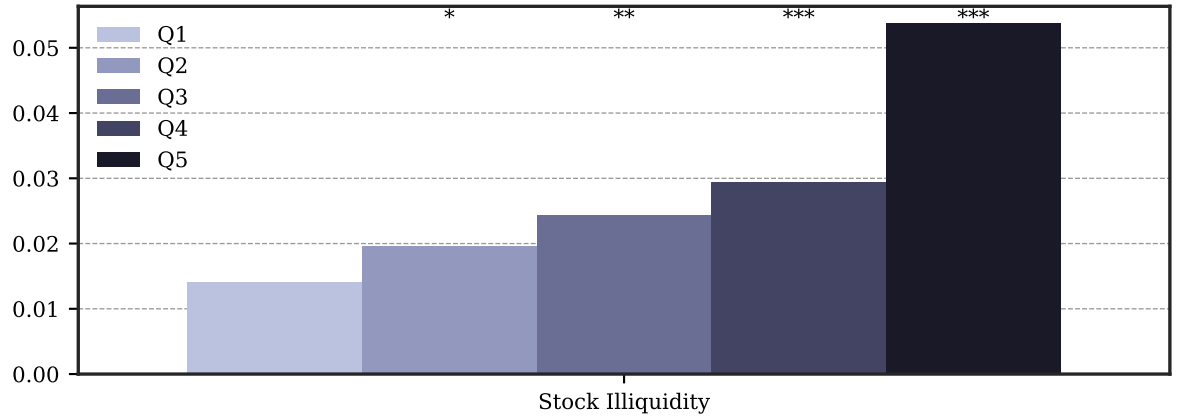

Fig. IA10.3. Predictability and Profitability Conditional on Stock Illiquidity – $R^2_{OS;XS}$

The left panel of the figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En for different quintiles of the underlying stock's illiquidity measured by the relative bid ask spread, $baspread = \frac{ask-bid}{mid}$. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7). The right panel shows the resulting Sharpe ratios of buying options in the highest predicted return decile and selling options in the lowest for each subsample.

IA10.4. Option Illiquidity

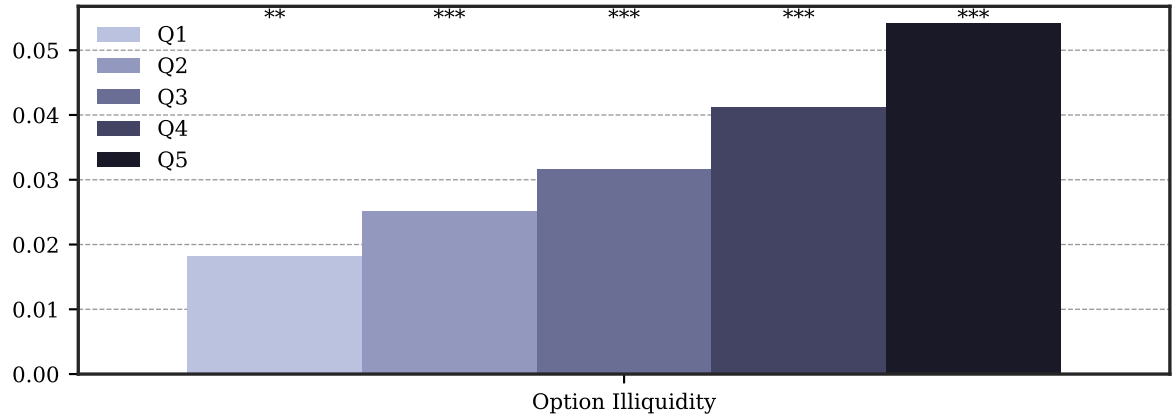

Fig. IA10.4. Predictability and Profitability Conditional on Option Illiquidity – R^2_{OS}

The left panel of the figure shows out-of-sample R^2_{OS} as defined in Equation (4) using the nonlinear ensemble N-En for different quintiles of the options' illiquidity measured by the relative bid ask spread, $optspread = \frac{ask-bid}{mid}$. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

	Low Pred.	2	3	4	High Pred.	H-L
			Option Illiquidity			
Low	-0.977***	-0.228	-0.056	0.085	0.445**	1.422***
2	-1.130***	-0.258**	-0.064	0.096	0.554***	1.684***
3	-1.333***	-0.313**	-0.074	0.086	0.587***	1.920***
4	-1.620***	-0.386***	-0.099	0.136	0.710***	2.330***
High	-2.061***	-0.535***	-0.109	0.173	0.817***	2.878***
H-L	-1.084***	-0.307***	-0.053	0.088	0.372***	1.457***

Table IA10.1: Bivariate Portfolios of Option Illiquidity and Expected Returns

The table shows realized returns for quintiles portfolios following the predictions by the nonlinear ensemble N-En within quintiles sorted by options' illiquidity measured by the relative bid ask spread, $optspread = \frac{ask-bid}{mid}$. ***, **, * denotes statistical significance at the 1%, 5% and 10%-level as defined in Equation (7). We also show the realized returns and significance for the resulting high-minus-low portfolios.

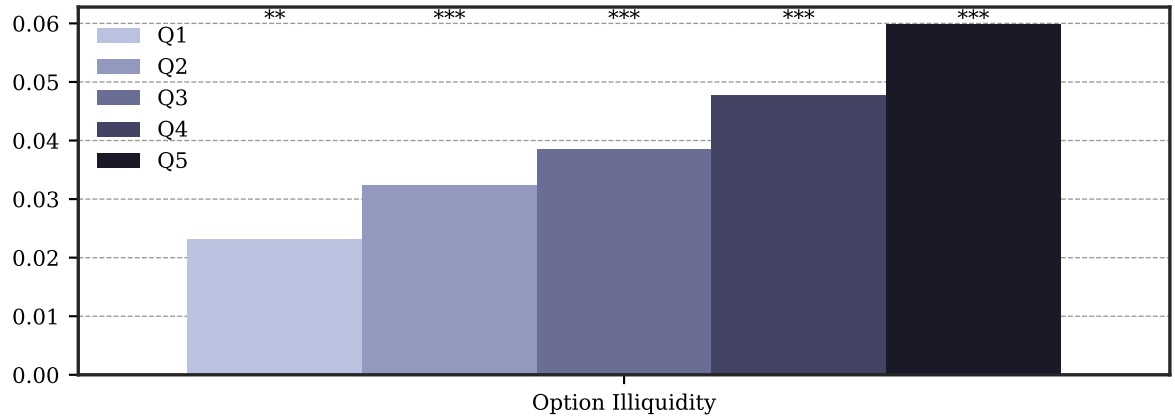

Fig. IA10.5. Predictability and Profitability Conditional on Option Illiquidity – $R^2_{OS;XS}$

The left panel of the figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En for different quintiles of the options' illiquidity measured by the relative bid ask spread, $optspread = \frac{ask-bid}{mid}$. ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7). The right panel shows the resulting Sharpe ratios of buying options in the highest predicted return decile and selling options in the lowest for each subsample.

IA10.5. Option Mispricing

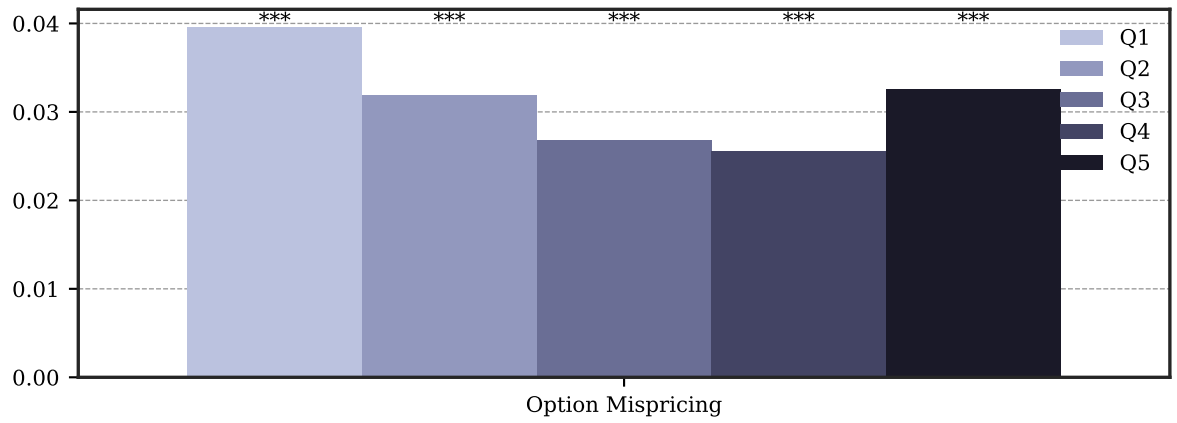

Fig. IA10.6. Predictability and Profitability Conditional on Option Mispricing – $R^2_{OS;XS}$

The left panel of the figure shows cross-sectional out-of-sample $R^2_{OS;XS}$ as defined in Equation (5) using the nonlinear ensemble N-En for different quintiles of option mispricing. We calculate option mispricing using the stocks' realized volatility over the last quarter and calculate "fair prices" for short-term at-the-money options using it as input to the Black and Scholes (1973) model. Mispricing is then defined as $\log(O/\widetilde{O})$, where $\widetilde{O}$ denotes the obtained "fair price". ***, **, * above the bars denotes statistical significance at the 0.1%, 1% and 5%-level as defined in Equation (7).

References

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.

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., 2020. Cheap options are expensive. Swiss finance institute research paper.

Fong, K. Y., Holden, C. W., Trzcinka, C. A., 2017. What are the best liquidity proxies for global research? Review of Finance 21, 1355–1401.

Gilad-Bachrach, R., Rashmi, K., 2015. Dart: Dropouts meet multiple additive regression trees. Cornell University: Cornell, Ithaca, NY, USA .

Goyenko, R. Y., Holden, C. W., Trzcinka, C. A., 2009. Do liquidity measures measure liquidity? Journal of Financial Economics 92, 153–181.

Green, J., Hand, J. R. M., Zhang, X. F., 2017. The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. Review of Financial Studies 30, 4389–4436.

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. Working paper.

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., 2021. Is there a replication crisis in finance? Tech. rep., National Bureau of Economic Research.

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

Jones, C. S., Khorram, M., Mo, H., 2020. Momentum, reversal, and seasonality in option returns. Working paper.

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.

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.

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.

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.