Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis Econometrics and Management Science

Quantitative Finance

A Nonparametric Approach to Characteristic Selection for Delta-Hedged Option Returns

Author

Daan Burwinkel (457091)

University Supervisor

Dr. Andrea Naghi

Second Assessor

Dr. Anastasija Tetereva

May 1, 2023

Abstract

The paper presents a nonparametric approach to analyze the additional information provided by characteristics for the expected delta-hedged option returns across the cross-section. Our approach utilizes the adaptive group LASSO for selecting and estimating the effect of chosen characteristics on the expected returns in a nonparametric manner. The method is flexible and can handle numerous characteristics while being robust to outliers. Our results indicate that several return predictors previously identified do not contribute additional information to expected returns, and non-linearities are significant.

Contents

1 Introduction

The primary objective of investors is to augment their financial assets. Since the inception of the stock market, investors have endeavored to identify the equities that offer the greatest potential for returns. The mounting volume of stocks available in the market has accentuated the necessity of comprehending the key determinants that influence stock returns. The Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965) and Mossin (1966) represented a significant milestone in understanding the interrelationships of stock returns. According to the CAPM, an asset's beta, which indicates its responsiveness to market fluctuations, serves as a comprehensive determinant of the entire cross-section. Fama and French (1992) were the pioneers in challenging the prevailing assumptions of the CAPM and subsequently expanded its framework with additional size and value factors (Fama & French, 1993). Subsequently, Fama and French (2015) further enhanced the model by introducing profitability and investment factors, thereby establishing the five-factor model as the benchmark for elucidating the variances in stock returns. In addition to the Fama and French factors, the most widely recognized factors are the reversal factor of Jegadeesh (1990) and the momentum factor of Jegadeesh and Titman (1993). Nowadays, many scholars have endeavored to identify the key features and factors that account for the differences in stock returns. Cochrane (2011) has remarked that the cross-section of expected returns is "is once again descending into chaos" while Harvey et al. (2016) have identified "hundreds of papers and factors" that possess predictive potential for the cross-section of expected stock returns, which Cochrane (2011) calls the "Factor Zoo".

In contrast to the stock market, the cross-sectional dynamics of the equity option market have received much less attention. Specifically, the characteristics and factors that drive the expected delta-hedged option returns, which are not affected by fluctuations in the underlying stock return, have not been thoroughly explored. This is surprising, given the substantial surge in the volume of options traded in the equity options markets. In fact, between 1995 and 2012, the largest stock markets in the United States grew by approximately 300%, whereas the total notional value of financial derivatives held by the 25 largest U.S. bank holding companies surged by around 1800% (Abdel-Khalik & Chen, 2015). Furthermore, during the same period, the U.S. gross domestic product (GDP) only doubled, which highlights the substantial expansion of the option derivative market. According to Horenstein et al. (2020), the limited number of studies in this area may be attributed to the perception that options are primarily leveraged positions in underlying stocks. Nonetheless, it is crucial to comprehend the key factors that account for the variances in delta-hedged option returns cross-sectionally. This is because options are appealing investment securities for low-risk investors owing to their inherent risk limitations.

The majority of studies in both the stock and equity option markets generally employ two primary techniques to assess whether a characteristic has a statistically and economically significant predictive effect on cross-sectional returns: portfolio sorting and the Fama and MacBeth (1973) cross-sectional regression. Although both methods have important applications, they fall short in what Cochrane (2011) calls "the multidimensional challenge". Freyberger et al. (2020) address multiple shortcomings for the two methods. At first, the curse of dimensionality can limit the applicability of portfolio sorting when the number of characteristics is substantial. Consequently, we can only employ this method to evaluate a restricted range of features. For example, if we sort stocks into five portfolios based on five potential characteristics, it would result in $5^5 = 3125$ portfolios. Secondly, portfolio sorting provides limited insight in distinguishing between characteristics. For instance, if we double-sort portfolios based on two distinct features and only observe a significant variation in returns for one category (such as for small stocks), can we conclude that the other feature is not significant for expected returns? This is what Fama and French (2008) call "awkward". Thirdly, the use of portfolio sorts assumes that expected returns remain constant across different segments of the characteristic's distribution. However, in practical scenarios, the majority of the variation in both the characteristic's values and returns are typically found in the extremes of the characteristic's distribution. This is what Fama and French (2008) call "clumsy". An alternative to portfolio sorts is the cross-sectional regression method proposed by Fama and MacBeth (1973), which enables the examination of the joint predictive power of multiple characteristics on expected returns. However, there is no guarantee that the relationship between the characteristics and expected returns will be linear, and non-linearities may play a crucial role. Additionally, the use of linear regression models is vulnerable to the influence of outliers that could lead to biased point estimates.

This study proposes a nonparametric approach to identify which firm characteristics provide additional information to the cross-sectional variation in delta-hedged option returns. This approach considers the characteristics conditionally rather than individually. The nonparametric estimation method proposed in this study utilizes the adaptive group LASSO (agLASSO) procedure developed by J. Huang et al. (2010). This two-step LASSO procedure is applied to the additive model structure of the conditional mean function, as it offers advantages such as faster convergence rates. The summands of the additive model are estimated through quadratic splines, whereby the additive model summands are approximated as quadratic functions within a defined interval. Consequently, the characteristics undergo a rank transformation. The combination of quadratic splines estimation and the rank transformation allows mapping of the nonparametric estimator directly to portfolio sorts, enabling comparison of each interval to a portfolio. The sample consists of characteristics previously examined for their impact on delta-hedged option returns' cross-section.

This framework has three different applications. Firstly, it examines which characteristics provide additional information for the cross-section of delta-hedged option returns. Secondly, it compares the out-of-sample performance of the nonparametric model to a linear model. The use of flexible form estimation techniques often raises concerns of in-sample overfitting. Model selection is performed once for a fixed time period, after which the model is estimated and one-month ahead delta-hedged option returns are predicted. The estimation and prediction period is then moved forward by one month. Thirdly, a rolling window estimation technique is employed using the agLASSO to investigate whether the predictive power of characteristics varies over time. This entails model selection and estimation over a rolling time period, followed by the prediction of one-month ahead delta-hedged option returns. The period is moved forward by one month, and the model selection and estimation are repeated to obtain a new prediction.

The paper contributes novel insights on the firm characteristics that offer additional information on the cross-section of expected delta-hedged option returns, given other characteristics. This aids in resolving the issue of the "multidimensional challenge." By identifying the characteristics that provide incremental information, one can associate them with factors, estimate the factors, construct factor models, and design trading approaches based on these factors.

2 Literature Review

This paper belongs to the literature that investigates the predictability of option returns using characteristics and factors. Section 2.1 reviews the relevant literature. In addition, the paper employs a regularization technique known as Adaptive Group LASSO on the additive model structure of the expected delta-hedged return equation. Section 2.2 presents an overview of the important literature related to regularization techniques.

2.1 The Cross-Section of Equity Options

2.1.1 Volatility Variables

The valuation of options is heavily influenced by volatility, and as a result, many studies have focused on equity option returns. The conventional approach in these studies is to use portfolio sorting and the Fama and MacBeth (1973) procedure. In a study by Bakshi and Kapadia (2003) , they examined the statistical properties of delta-hedged option portfolios and found that they are exposed to a negative market volatility premium, in addition to the equity premium. Driessen et al. (2009) investigated the relationship between market-wide correlation shocks and the cross-section of option returns. Their findings indicate that correlation risk exposure can account for the cross-section of both index and individual options.

Duan and Wei (2009) examined the relationship between the systematic risk of the underlying stock, measured by the ratio of systematic volatility to total volatility, and the level and slope of the implied volatility curves. They found that higher systematic risk of the underlying stock leads to higher levels and steeper slopes of the implied volatility curve, which results in differential prices and returns across individual equity options. Later, Cao and Han (2013) investigated the association between systematic risk and delta-hedged option returns. They discovered a positive correlation between return and systematic risk of the underlying, while a negative correlation was found between delta-hedged option returns and the idiosyncratic volatility of the underlying. Options with high idiosyncratic volatility in the underlying stock are difficult to hedge and have higher arbitrage costs. Therefore, market makers require a higher compensation for these options. In addition, Vasquez (2017) examined the slope of the volatility term structure and found that it is positively related to straddle portfolio returns.

Goyal and Saretto (2009) investigated the relationship between returns and the discrepancy between historical volatility and at-the-money implied volatility. They found that options with high realized volatility relative to implied volatility generate higher returns due to the inaccurate estimation of volatility dynamics in options. In a recent study, Cao et al. (2022) found a negative association between volatility uncertainty and expected delta-hedged option returns. They argue that these findings are primarily influenced by model risk and investors' propensity for risk-taking, which results in market makers demanding a higher premium. This is supported by the fact that an increase in volatility uncertainty, rather than a decrease, contributes to the effect of implied volatility uncertainty.

2.1.2 Investors' Behavior Variables

Investigating the influence of investor behavior on option returns, specifically regarding their gambling preferences and demand pressures, has been another area of interest. Bollen and Whaley (2004)studied the relationship between the implied volatility function and net buying pressure, which affects option prices in the Black and Scholes (1973) formula. They concluded that changes in this function are directly connected to net buying pressure from public order flow. Garleanu et al. (2008) complemented this study by exploring the effects of demand pressure on option prices, as continuous hedging of an option in the real world is practically unfeasible. Risk-averse intermediaries, who take the other side of end-user option demand, impact option prices and returns. End-users are generally net long in index options but net short in stock options. The researchers discovered a positive cross-sectional relationship between option expensiveness and end-user demand, while in the time-series, the demand by end-users explains the value and skew pattern of index options.

Boyer and Vorkink (2014) established a negative link between the total skewness of an option at the ex ante stage and its subsequent average return, while studying the impact of investors' propensity for gambling. They posited that option demand is largely driven by a preference for skewness, leading intermediaries to demand high compensation for bearing the unhedgeable risk arising from investors' requests for options that resemble lotteries. In a complementary study, Byun and Kim (2016) investigated the relationship between a stock's potential for a lottery-like payout (measured by its extreme positive value and skewness) and option returns, in order to explore the possibility of investors' gambling preferences in equity option pricing. They observed a negative correlation, indicating that investors overvalue options with lottery-like potential in the underlying stock, resulting in lower returns.

Ramachandran and Tayal (2021) explore the connection between equity option returns and short-sale constraints based on demand-driven option pricing and the mispricing of underlying stocks. The study finds a link between these constraints and put option returns for overpriced stocks in the cross-section. Specifically, put options on the most constrained stocks are more costly and offer lower delta-hedged returns. This observation is consistent with the findings of Garleanu et al. (2008).

2.1.3 Liquidity Variables and Others

The relationship between option returns and liquidity has garnered significant attention in recent literature. In particular, Christoffersen, Goyenko, et al. (2018) investigated this relationship for equity options and found that illiquid options generate higher returns than liquid ones, indicating the presence of a positive illiquidity premium. This finding can be explained by the fact that market maker intermediaries typically hold large net long positions in the equity stock option market and require compensation for the risks and costs associated with these positions. The results of this study are consistent with those of Garleanu et al. (2008), which highlight the impact of imbalanced demand and higher costs on expected returns.

Moreover, Choy and Wei (2016) explored the possibility of a liquidity risk premium and found evidence of a negative but weaker premium. Notably, a positive illiquidity premium is equivalent to a negative liquidity premium. The authors argue that end-users, who typically hold net short positions in the equity stock option market and often do not hedge their positions, might be more concerned about liquidity risk than market makers. The weaker liquidity premium they found could be due to their use of leveraged-adjusted hedged option returns and variables that focus on the risk premium, rather than the level effect premium investigated by Christoffersen, Goyenko, et al. (2018).

At last, Cao et al. (2021) and Vasquez and Xiao (2020) have conducted notable studies in this field. The former investigated the impact of twelve stock characteristics on the cross-section of delta-hedged returns and found that eight of them have a significant effect. Specifically, they discovered that stock price, profit margin, and firm profitability have a negative correlation with selling delta-hedged calls, while cash holdings, cash flow variance, new shares issuance, total external financing, distress risk, and dispersion of analyst forecasts have a positive correlation with selling these securities. Conversely, the opposite holds true when delta-hedged calls are bought. Additionally, they found that idiosyncratic stock volatility and illiquidity factors can explain these characteristics, whereas stock risk factors have no explanatory power for the cross-section of delta-hedged returns. The latter identified that default risk, in terms of credit rating or default probability, is a powerful predictor of future option returns. In the cross-section, high default risk leads to low delta-hedged option returns.

2.1.4 Sophisticated Methods

Various studies employ advanced methods to identify the significance of factors and characteristics in equity option returns. One such study is Christoffersen, Fournier, et al. (2018), , which used Principal Component Analysis (PCA) to develop an equity option pricing model. The study found a strong factor structure in equity options, where the first principal components of equity volatility levels, skews, and term structures are highly correlated with the corresponding S&P 500 index option factors. Another study by Horenstein et al. (2020) also utilized PCA and examined eleven characteristics and two market factors to develop a four-factor model that explains the cross-section and time-series of delta-hedged option returns. The four factors identified were firm size, idiosyncratic volatility, the difference between implied and historical volatilities, and the market volatility risk factor. The study further concluded that the cross-section of delta-hedged option returns is not associated with stock return factors.

Subsequently, Büchner and Kelly (2022) enhanced the Principal Component Analysis (PCA) methodology by using Instrumented Principal Component Analysis (IPCA) on a monthly panel of S&P 500 option returns. IPCA is capable of considering time-varying loadings and incorporating external data. Their analysis led to the identification of a three-factor model comprising of a level, slope, and skew factor.

Brooks et al. (2018). conducted research similar to ours, using linear LASSO, adaptive LASSO, and group LASSO regularization techniques to identify the characteristics that predict the cross-section of individual equity options. However, the main difference is that our paper employs the adaptive group LASSO technique on the additive model structure of the conditional mean function of expected returns. This approach has been shown to overcome some major estimation issues and outperform other LASSO techniques in model selection and out-of-sample estimation, as noted by Freyberger et al. (2020).

All of the studies mentioned previously employ variables that can be incorporated into a dataset of characteristics. However, many of these studies examine individual characteristics in isolation or in low dimensions, without assessing whether they offer incremental information conditional on other factors for the cross-section of delta-hedged returns. In addition, the studies that employ more advanced methods use a relatively limited set of characteristics, which increases the likelihood of important variables being excluded.

2.2 Regularization Techniques

The least absolute shrinkage and selection operator (LASSO) was first introduced by Tibshirani (1996). It adds a penalty term to regression analysis that aims to minimize the sum of squared residuals. The penalty term is a constrained form of ordinary least squares regression, where the sum of absolute values of regression coefficients must be lower than a given constant. This constraint leads to some parameters being forced to zero.

Hereafter, Yuan and Lin (2006) proposed an extension to the ordinary LASSO method, which allows groups of covariates to be jointly included or excluded from the model. They found that while the group LASSO had excellent performance, it was computationally intensive for large-scale problems.

Moreover, the (group) LASSO technique is prone to estimation inefficiency and selection inconsistency. In response, Zou (2006) advocated an extension to the LASSO method, referred to as adaptive LASSO. This approach establishes the conditions under which the LASSO technique is consistent, providing an alternative in scenarios where the LASSO approach fails to achieve consistency. The adaptive LASSO approach employs adaptive weights to penalize each coefficient differently, where each weight is a function of previously estimated coefficients, such as those from ordinary LASSO or ordinary least squares. Hence, the adaptive LASSO approach is a two-stage process, and the resulting estimates fulfill the oracle properties.

In order to address the same issues encountered in the group LASSO procedure, Wang and Leng (2008) proposed combining the group LASSO with the adaptive LASSO to create the adaptive group LASSO. This approach effectively resolves the problems of estimation inefficiency and selection inconsistency. Similar to the adaptive LASSO, the adaptive group LASSO involves a two-stage process. Initially, the group LASSO is used to perform variable selection and estimation in a grouped manner.

The resulting estimates are then utilized to construct variable-specific weights for use in the second stage, the adaptive LASSO. J. Huang et al. (2010) demonstrate how to implement the adaptive group LASSO in a nonparametric additive model structure employing quadratic splines, which is a central component of this study.

The LASSO procedure has been applied in various finance studies. For instance, Rapach et al. (2013) employed the LASSO method to investigate the relationship between international and US stock returns. J.-z. Huang and Shi (2011) used the adaptive group LASSO to identify macro factors with predictive power on bond risk premia. Chinco et al. (2019) employed LASSO to examine whether lagged returns of the cross-section could be used as candidate predictors in high-frequency stock returns. Goto and Xu (2015) utilized the LASSO technique to obtain a sparse estimator of the inverse covariance matrix for portfolio optimization in a mean variance context. Chinco et al. (2021) use the tuning parameter of LASSO to estimate the anomaly base rate. DeMiguel et al. (2020), Feng et al. (2020), and Freyberger et al. (2020) used different LASSO methods to focus on characteristic-based factors. Brooks et al. (2018) also use different linear LASSO techniques on the cross-section of equity option returns to identify characteristics that have conditional predictive power.

3 Data

In this section, we will examine the data used to model delta-hedged returns. Firstly, Section 3.1 provides details on the data sample and the data filtering process that is applied. Secondly, in Section 3.2 we will discuss the use of delta-hedged option returns. Finally, in Section 3.3, we will comment on the selected characteristics and the procedure used to re-scale them.

3.1 Data and Sampling Methodology

The option data on individual equities is sourced from the OptionMetrics IvyDB US database, which includes companies listed on the major American exchanges such as NYSE, NASDAQ, and NYSE American. While Green et al. (2017) found that the return predictability for hedged portfolios utilizing characteristics substantially decreased in 2003, Han et al. (2018) demonstrated that the aforementioned conclusion could be attributed to overfitting. As this research aims to investigate the return predictability of characteristics over time, both pre- and post-2003 era data is included in the sample. Therefore, the sample period covers January 1996 to December 2021. The implied volatility and the Greeks are computed by OptionMetrics using the binomial tree from Cox et al. (1979). In addition, stock returns and prices are obtained from the Center for Research on Security Prices (CRSP), and informative characteristics are sourced from the Standard and Poor's Compustat and I/B/E/S databases. The option prices are approximated by the midpoint between the closing bid and ask quotes.

The option data undergoes a data filtering process akin to that of Brooks et al. (2018). The initial step entails removing options with abnormal bid-ask spread. An abnormal bid-ask spread is defined as negative, exceeding \$5, or below the minimum tick size. Additionally, extreme option prices that are less than 50% of the exercise value or \$100 above it are removed, following the approach of Boyer and Vorkink (2014), i.e. $S_t - K$ for calls and $K - S_T$ for puts. Excluding options with unusual bid-ask spreads and extreme prices decreases the likelihood of incorporating erroneous option prices in the dataset, as explained by Duarte et al. (2019). Secondly, options that are associated with underlying stocks paying dividends over the remaining option life are eliminated. This step eliminates the early exercise premium in American-style options and guarantees that the sample comprises exclusively European-style options. Thirdly, all options violating the no-arbitrage restrictions specified in Equations 1 and 2 are excluded.

$$\text{Call: } \max(0, S_t - K, S_t - \sum_{s=1}^{\tau} e^{-rs} D_{t+s} - Ke^{-r\tau}) \leq C_t(\tau, K) \leq S_t \tag{1}$$

$$\text{Put: } \max(0, K - St, \sum_{s=1}^{\tau} e^{-rs} D_{t+s} + Ke^{-r\tau} - S_t) \leq P_t(\tau, K) \leq K \tag{2}$$

The fourth criterion for the sample is that it includes only options that are close to the money, with a moneyness between 0.95 and 1.05.

After the data filtering process, each underlying stock may have multiple available option prices. To standardize the sample, the option closest to the money is chosen, and on the first trading day of each month, the option expiring on the third Friday of the following month is selected (Brooks et al., 2018). This ensures homogeneity in moneyness and maturity. The impact of these variables on option returns has been studied by Coval and Shumway (2001) but is not within the scope of this research. Although it is worth noting that they discovered a notable diversity in the collection of equity option returns, even when considering options with identical levels of moneyness and maturity. To avoid microstructure effects in option delivery, the option portfolio is held until the first trading day of the expiration month. (Ni et al., 2005). The option data retained is for underlying stocks that can be found in the CRSP, Compustat, and I/B/E/S databases.

Table 1 presents the descriptive statistics of the option sample, which comprises of $222,391$ option-month observations for calls and $178,472$ observations for puts. The average moneyness of call and put options are both close to one, with values of 1.003 and 0.996 respectively. The mean delta values for calls and puts are 0.518 and -0.493, respectively, indicating homogeneity in moneyness across the sample. Additionally, the average time to maturity is 46.932 for calls and 46.729 for puts. Finally, the average bid-ask spread relative to option prices is 0.153 for calls and 0.126 for puts, which shows that the filtering process has successfully reduced the microstructure biases of delta-hedged option returns.

3.2 Delta-Hedged Option Returns

Since options are derivatives on stocks, there is an obvious relation between the returns of the two financial products. This study investigates which characteristics offer additional and conditional insights into the cross-sectional variation of option returns, over and above changes in the underlying stock price. To achieve this goal, we employ delta-hedged option returns. These returns are designed to be theoretically immune to shifts in the underlying stock price, and are widely utilized in practical applications (Bakshi & Kapadia, 2003; Büchner & Kelly, 2022; Cao & Han, 2013; Goyal & Saretto, 2009; Horenstein et al., 2020). A delta-hedged option portfolio is expected to generate a return equivalent to the risk-free rate, provided that it can be perfectly replicated using the underlying asset and a risk-free bond (Bertsimas et al., 2001). However, empirical studies have found negative returns for delta-hedged option portfolios (Bakshi & Kapadia, 2003; Cao & Han, 2013; Carr & Wu, 2009; Goyal & Saretto, 2009). The delta-hedged call (put) position involves purchasing a call (put) option and selling (buying) a specific number of delta shares of the underlying stock, or vice versa. To calculate the delta-hedged returns over a given period τ, we adopt the definition of Bakshi and Kapadia (2003):

$$\Pi_{it\to t+\tau} = O_{it+\tau} - O_{it} - \int_t^{t+\tau} \Delta_{iu} dS_{iu} - \int_t^{t+\tau} r_u^f (O_{iu} - \Delta_{iu} S_{iu}) du \tag{3}$$

where $O_{it+\tau}$ and O_{it} are the prices of European option i on the first trading day of the expiration month and the first trading day of the month prior to expiration respectively, $\Delta_{iu} = \frac{\partial O_{iu}}{\partial S_{it}}$ is the delta of the option at time u, S_{iu} is the price of the underlying stock at time u, and r_u^f is the annualized risk-free rate at time u, which equals the yield of the 10-year U.S. bond from Federal Reserve Economic Data (FRED). Equation 3 can be parsed into three distinct components. The first part represents the variation in the option price over the defined period, the second component adjusts for continuously delta-hedging the position, and the last component represents the expenses incurred while financing the delta-hedged position with the risk-free rate. This guarantees that the delta-hedged option gain is expressed above and beyond this rate.

Since Bakshi and Kapadia (2003) employed this equation in a simulation analysis, it must be discretized for empirical implementation (Horenstein et al., 2020). The discrete delta-hedged option gain from t until maturity $t + \tau$ is given as follows:

$$\Pi_{it\to t+\tau} = O_{it+\tau} - O_{it} - \sum_{n=0}^{N-1} \Delta_{it_n} [S_{it_{n+1}} - S_{it_n}] - \sum_{n=0}^{N-1} \frac{a_n r_{t_n}^f}{365} (O_{it_n} - \Delta_{it_n} S_{it_n}) \tag{4}$$

where a_n is the number of calendar days between t_n and t_{n+1}. This means that the option is discretely delta-hedged N times over the period $[t, t + \tau]$, such that the hedged portfolio is rebalanced on every t_n, $n = 0, 1, ..., N - 1$, where $t_0 = t$ and $t_N = t + \tau$.

The variation in prices between options makes the delta-hedged gains incomparable. Hence, Equation 5 normalizes the delta-hedged gains by dividing them with the initial invested amount at time t to create delta-hedged returns that are comparable.

$$R_{it} = \frac{\Pi_{it\to t+\tau}}{|(\Delta_{it} S_{it} - O_{it})|} \tag{5}$$

This study employs a homogeneous sample of options with identical moneyness and maturity, as well as a fixed holding period for delta-hedged portfolios. Accordingly, returns are allocated to the period that encompasses the majority of the holding period (t), thereby addressing the challenge of overlapping timeframes between characteristics and returns. This methodology is necessary for the characteristics to accurately predict delta-hedged option returns.

Table 1 displays the mean returns of delta-hedged options. Delta-hedged call options have an average return of -0.452%, while delta-hedged put options have an average return of -0.668%. The standard deviation for call and put options are 6.834% and 5.364%, respectively. These negative returns are consistent with previous research in this field.

Table 1: Option Sample Statistical Summary

Variable	Mean	Std Dev.	p10	p25	Median	p75	p90
Panel A: Call option							
Return	-0.452	6.834	-5.322	-2.688	-0.681	1.131	4.501
Moneyness	1.003	0.029	0.962	0.993	1.002	1.018	1.036
Delta	0.518	0.083	0.387	0.434	0.519	0.575	0.618
Days to maturity	46.932	2.663	44	45	46	49	51
Bid-Ask spread	0.275	0.300	0.100	0.100	0.200	0.300	0.450
Bid-Ask spread/option price	0.153	0.145	0.047	0.072	0.104	0.169	0.266
Panel B: Put Options							
Return	-0.668	5.364	-5.021	-2.639	-0.981	0.634	3.404
Moneyness	0.996	0.016	0.943	0.988	0.996	1.009	1.028
Delta	-0.493	0.083	-0.499	-0.512	-0.471	-0.421	-0.365
Days to maturity	46.729	2.558	44	45	46	47	51
Bid-Ask spread	0.269	0.294	0.100	0.100	0.200	0.300	0.450
Bid-Ask spread/option price	0.126	0.118	0.042	0.054	0.098	0.148	0.244

3.3 Characteristics

In order to identify the characteristics that offer additional and conditional insights into the cross-section of delta-hedged option returns, a significant range of informative characteristics is utilized. Specifically, the characteristics outlined in Brooks et al. (2018) and Green et al. (2017), which cover option, stock, and firm-level characteristics, are employed. Additionally, if any of the characteristics discussed in Section 2.1 are not included in the data set, they are added. Finally, well-known option features, including implied volatility, trading volume, and the Greeks (delta, gamma, vega, and theta) from Black and Scholes (1973) that are computed using the binomial tree of Cox et al. (1979), are included. These features have previously been linked to characteristic-based anomalies in past research. The dataset comprises 130 characteristics, including 100 stock-level and 30 option-level variables. For an exhaustive list of these variables and their respective calculations, please refer to the study by Brooks et al. (2018).

The characteristics used in this study are classified into two categories: (1) Option Characteristics; and (2) Stock Characteristics.For balance sheet data, we adopt the timing convention introduced in Fama and French (1993) and Freyberger et al. (2020) to address the look-ahead bias. Particularly, for the estimation period of June year t until May year $t+1$, and the prediction period of July year t until June year $t+1$, balance sheet data of the fiscal year ending in calendar year $t-1$ is used. In contrast, the frequency of stock and option characteristics is monthly, which is consistent with the frequency of returns.

To mitigate the impact of outliers and enhance the comprehensibility of the nonparametric estimator, a standardization process is carried out on the characteristics. This involves employing a rank transformation technique introduced by Freyberger et al. (2020), as presented in Equation 6.

$$F_{s,t}(C_{s,it-1}) = \frac{\text{rank}(C_{s,it-1})}{N_t + 1} \tag{6}$$

The rank transformation of $C_{s,it-1}$ maps the the cross-sectional distribution of a given characteristic s to the unit interval, $C_{s,it-1} \in [0, 1]$. Here, $\text{rank}(\min_{i=1,...,N_t} C_{s,it-1}) = 1$ and $\text{rank}(\max_{i=1,...,N_t} C_{s,it-1}) = N_t$. The method is performed for all time periods $t \in \{1, ..., T\}$ and all characteristics S. The transformation establishes a one-to-one correspondence with portfolio sorting by focusing on the rank of the characteristic within the cross-section rather than its absolute value. The transformation also mitigates the impact of outliers by narrowing the range of the transformed characteristics, which generally enhances out-of-sample performance.

4 Methodology

The methodology section is organized as follows: Section 4.1 presents an overview of methods commonly used to determine the impact of a characteristic on the cross-sectional returns, highlighting their respective strengths and weaknesses, and intuition for the proposed framework in Sections 4.1.1, and Section 4.1.2. Subsequently, Section 4.2 explores nonparametric estimation, specifically focusing on the conditional mean function of the returns and the incorporation of characteristics in Sections 4.2.1 and 4.2.2. We then present the proposed approach for estimating the conditional mean function and its corresponding interpretation in Sections 4.2.3 and 4.2.4. Finally, in Section 4.2.5, we present the proposed tuning parameters for the method.

4.1 Current Methods

4.1.1 Expected Returns

Most researchers in empirical asset pricing literature concentrate on identifying characteristics that can predict expected returns. In other words, they try to find a characteristic C in period $t-1$ that predict the excess returns of firm i in the next period, R_{it}. In fact, they try to specify the conditional mean function, m_t, written as:

$$m_t(c_1, ..., c_S) = \mathbb{E}[R_{it}|C_{1,it-1} = c_1, ..., C_{S,it-1} = c_S] \tag{7}$$

where $C_{1,it-1}, ..., C_{S,it-1}$ are the S firm characteristics.

One commonly used approach to approximate (7) for a single characteristic is portfolio sorting. However, while portfolio sorts are simple and intuitive, they also have several drawbacks. First, they suffer from the curse of dimensionality (Fama & French, 2015). The number of portfolios grows exponentially as the number of characteristics increases. For instance, if derivatives are sorted jointly using six characteristics and each characteristic is partitioned into five portfolios, the resulting number of portfolios would be $5^6 = 15,625$. This large number of portfolios can quickly become comparable to the number of firms included in the initial sample. As a result, portfolios may become insufficiently diversified, leading to reduced power in asset-pricing tests (Fama & French, 2015). Hence, this method can only analyse a small number of characteristics. Second, the approach provides limited formal guidance regarding the interpretation of return spreads in higher-order conditionally sorted portfolios. To illustrate, suppose that derivatives are divided into five portfolios based on a particular characteristic, and within each of these portfolios, five additional portfolios are constructed based on another characteristic. If a return spread is observed for the latter characteristic in only one out of five portfolios of the first characteristic, concluding that the latter characteristic has no influence on expected returns could be erroneous. This is what Fama and French (2008) call "awkward". Third, portfolio sorting assumes constant expected returns across various segments of a characteristic's distribution, which implies that the estimator of the conditional mean function is flawed. This is what Fama and French (2008) call "clumsy". In fact, portfolio sorts can be viewed as a limited type of nonparametric regression, as demonstrated in Section 4.1.2. This similarity provides a conceptual basis for the proposed methodology.

Another widely adopted approach is spanning tests, which employ the significant return patterns of portfolio sorts to create a long-short portfolio. This is achieved by purchasing the portfolio with the highest expected return and selling short the portfolio with the lowest expected return. Subsequently, the hedged return is regressed on a set of established risk factors. (Fama & French, 1993, 2015).If a time-series intercept is statistically significant, it suggests that the risk factors do not fully account for the variability in the hedged returns. Furthermore, the hedged return could be regarded as an additional risk factor that helps explain the cross-sectional variation. Gibbons et al. (1989) demonstrated that a significant time-series intercept corresponds to enhanced Sharpe ratios for a mean-variance investor. However, it should be noted that the order in which characteristics are examined matters, and thus spanning tests cannot resolve the issue of identifying which characteristics offer incremental information for expected returns in the cross-section.

The last alternative to the aforementioned methods is a linear regression. We assume linearity of the conditional mean function (7) and run a panel regression of excess returns on S characteristics as in (8).

$$R_{it} = \alpha + \sum_{s=1}^{S} \beta_s C_{s,it-1} + \epsilon_{it} \tag{8}$$

This method allows to jointly investigate the predictive power of characteristics on expected returns. An example is the regression of Fama and MacBeth (1973), in which point estimates are numerically equivalent to the estimates of (8) if characteristics are constant over time. Nevertheless, this approach has a few limitations. Firstly, there is no justification for assuming linearity of the conditional mean function in advance. Fama and French (2008) have raised concerns about the possibility of nonlinearities and have consequently made ad hoc adjustments. For example, they use the logarithm of the book-to-market value as a predictive variable. Secondly, these regressions are vulnerable to outliers that may significantly impact the point estimates. To mitigate this concern, researchers often employ data winsorization and subsequently estimate regressions separately.

4.1.2 Equivalence Portfolio Sorting and Regressions

Cochrane (2011) addresses that "Portfolio sorts are really the same thing as nonparametric cross-sectional regressions, using nonoverlapping histogram weights." Following Freyberger et al. (2020), we show the equivalence between these portfolio sorts and regressions and use the outcome to motivate the nonparametric framework.

Suppose we only consider univariate portfolios sorts for intuition, we have excess return R_{it} and characteristic C_{it-1} for derivatives $i = 1, ..., N_t$ and periods $t = 1, ..., T$. The derivatives are subsequently divided into L portfolios based on the value of the lagged characteristic C_{it-1}. Particularly, a derivative i is in portfolio l if it holds that $C_{it-1} \in I_{tl}$, where I_{tl} is an interval of the distribution of the firm characteristic at time t for portfolio l. If each time period t consists of N_t derivatives, then the number of derivatives at time t in portfolio l equals N_{tl}, $N_{tl} = \sum_{i=1}^{N_t} \mathbb{1}(C_{it-1} \in I_{tl})$. It follows that the excess return of portfolio l at time t, equals (9).

$$P_{tl} = \frac{1}{N_{tl}} \sum_{i=1}^{N_t} R_{it} \mathbb{1}(C_{it-1} \in I_{tl}) \tag{9}$$

On the other hand, a pooled time-series cross-sectional regression of excess returns on dummy variables can be used. Here, $\mathbb{1}(C_{it-1} \in I_{tl})$ is the dummy variable which equals one if the characteristic of firm i is in portfolio l at time t. This results in the following regression (10).

$$R_{it} = \sum_{l=1}^{L} \beta_l \mathbb{1}(C_{it-1} \in I_{tl}) + \epsilon_{it} \tag{10}$$

Now, let $\hat{\beta}$ be the Ordinary Least Squares (OLS) estimator, $\hat{\beta} = (X'X)^{-1}X'R$, with R the NT x 1 vector of excess returns and X the NT x L matrix of dummy variables $\mathbb{1}(C_{it-1} \in I_{tl})$. When substituting (10) in (9) and solve for the OLS estimate, (11) results.

$$\hat{\beta}_l = \frac{1}{T} \sum_{t=1}^{T} \frac{N_{tl}}{\frac{1}{T} \sum_{t=1}^{T} N_{tl}} P_{tl} \tag{11}$$

This equation can be simplified if we assume the number of derivatives in each portfolio l to be constant over time t, $N_{tl} = \bar{N}_l$, as is shown in (12).

$$\hat{\beta}_l = \frac{1}{T} \sum_{t=1}^{T} P_{tl} \tag{12}$$

Furthermore, the following holds for the difference of two OLS estimates:

$$\hat{\beta}_l - \hat{\beta}_{l'} = \frac{1}{T} \sum_{t=1}^{T} (P_{tl} - P_{tl'}) \tag{13}$$

Therefore, the coefficients of the pooled time-series cross-sectional regressions represent the average portfolio return over time. Moreover, the disparity between the coefficients of two portfolios corresponds to the mean excess return between the portfolios over time.

If we abandon the presumption of constant derivatives within each portfolio over time, which implies that the number of derivatives in the portfolios varies at each time period, the regression coefficients' outcomes usually alter. Two methods can restore the equivalence. Firstly, we can consider the varying number of derivatives in the portfolios over time and define the excess return as in (14).

$$\hat{\beta}_l - \hat{\beta}_{l'} = \frac{1}{\sum_{t=1}^{T} N_{tl}} \sum_{t=1}^{T} N_{tl} P_{tl} - \frac{1}{\sum_{t=1}^{T} N_{tl'}} \sum_{t=1}^{T} N_{tl'} P_{tl'} \tag{14}$$

Second, it is possible to use a Weighted Least Squares Estimator (WLS), $\tilde{\beta} = (X'WX)^{-1}X'WR$. The NT x NT weight matrix W accounts for the changing amount of derivatives in the portfolio by using the inverse of the number of derivatives on the diagonal. Again, the difference of the estimated coefficients equals the average excess return between the portfolios over time (15).

$$\tilde{\beta}_l - \tilde{\beta}_{l'} = \frac{1}{T} \sum_{t=1}^{T} (P_{tl} - P_{tl'}) \tag{15}$$

4.2 Nonparametric Estimation

The relation between regressions and portfolio sorts offers insight into the suggested nonparametric estimation framework and the interpretation of portfolio sorts as a specific instance of nonparametric estimation.

Suppose the conditional mean function, $m_t(c)$, is known. By means of statistical rules, the following holds for the expected value of the excess return of portfolio l (16).

$$\mathbb{E}[R_{it}|C_{it-1} \in I_{lt}] = \int_{I_{lt}} m_t(c) f_{C_{it-1}|C_{it-1} \in I_{lt}}(c) dc \tag{16}$$

The conditional mean function is multiplied with the density function of the characteristic in period $t-1$, given that the characteristic at $t-1$ is in the interval of portfolio l at time t, $f_{C_{it-1}|C_{it-1} \in I_{lt}}(c)$. Thereafter, integrate the multiplication over the appropriate interval of the characteristic distribution to construct the expected value of the excess return of portfolio l. Now, all necessary information for portfolio returns is present in the conditional mean function as in (16). However, knowing the conditional mean, $m_t(c)$, means that additional information about nonlinear relationships between characteristics and expected returns, and a more general functional form is needed.

Again consider the pooled time-series cross-sectional regression in (10). The L dummy variables of the form $\mathbb{1}(C_{it-1} \in I_{tl})$ are called constant splines in nonparametric estimation. The use of constant splines in estimating the conditional mean function involves approximating the function with a step function. However, such step functions are unsuitable as nonparametric estimators due to their nonsmoothness and undesirable theoretical properties.

To expand on the relationship between portfolio sorts and nonparametric estimation, Freyberger et al. (2020) conducted simulations and found three important results. Firstly, portfolio means or constant splines are adequate approximations of the conditional mean function, particularly for intermediate portfolios. However, they fail to capture the nonlinearities present in the extreme ends of the characteristic distribution. Secondly, higher order nonparametric estimators are effective in approximating the relationship between returns and characteristics in both intermediate portfolios and extremes. Finally, they discovered that as the number of portfolios increases, portfolio means offer better approximations in the extremes, and the mean returns and higher order nonparametric estimators' predictions become more comparable.

4.2.1 Additive Conditional Mean Function

Portfolio sorts and regressions share the characteristic of allowing the conditional examination of a range of characteristics. However, when the number of characteristics is increased in portfolio sorts, the estimators can become infeasible. If one wants to analyse the effect of four characteristics and jointly partition each characteristic into five portfolios, the number of portfolios to analyse will be $5^4 = 625$. The resulting portfolios are not only impractical to analyse, but also lack diversification.

This issue is also encountered in nonparametric regressions. Employing a completely nonparametric estimation method that involves numerous characteristics for the estimation of the conditional mean function, m_t, leads to a slow rate of convergence and imprecise estimates that are difficult to apply in practice (Stone, 1982). To be more precise, if one assumes technical regularity conditions, the optimal rate of convergence in mean square equals $N_t^{-4/(4+S)}$, with S the number of characteristics and N_t the number of observations at time t. This nonparametric rate of convergence is always smaller than its parametric equivalent, N_t^{-1}. If we assume that the conditional mean function is twice continuously differentiable, the convergence rate slows down as the number of characteristics S increases. Consequently, when there are many characteristics, the estimator's finite sample properties deteriorate.

For instance, consider a scenario where a researcher examines the impact of a single characteristic versus testing eleven characteristics simultaneously. The rate of convergence in the former scenario is denoted by $N_t^{-4/5}$, whereas the latter rate of convergence equals $(N_t^*)^{-4/15}$. Equation (17) shows that the rate of convergence in both situations are equal if the sample size in the latter situation equals the sample size of the former situation to the power of three.

$$(N_t^*)^{-4/15} = N_t^{-4/5} \Rightarrow N_t^* = N_t^3 \tag{17}$$

Furthermore, the estimates in both situations now have comparable finite sample properties. Suppose the situation with only one characteristic has a sample size of one thousand, $N_t = 1,000$. Then, the second situation must have a sample size of one billion returns to obtain the same rate of convergence and therefore similar finite sample properties. Inversely, the first situation only needs ten observations if the sample size with eleven characteristics equals one thousand, $N_t^* = 1,000$.

Identifying the characteristics that conditionally yield additional information for the cross-sectional analysis of delta-hedged option returns is challenging using two approaches. Firstly, focusing on a single characteristic in isolation is inadequate as it fails to achieve the intended goal. Secondly, analyzing multiple characteristics simultaneously necessitates a considerable sample size to ensure the convergence rate is sufficient for producing estimators with satisfactory finite sample properties. Freyberger et al. (2020) propose an additive model structure to resolve this challenge, which is a suitable remedy for the nonparametric regression framework. Equation (18) displays the division of the conditional mean function with all characteristics into a sum that comprises individual characteristic conditional mean functions.

$$m_t(c_1, ..., c_S) = \sum_{s=1}^{S} m_{ts}(c_s) \tag{18}$$

Here, $m_{ts}(\cdot)$ are unknown and to be estimated functions. The key theoretical benefit of the additive model structure is its steady convergence rate. Specifically, the rate of convergence is $N_t^{-4/5}$, which is independent of the number of characteristics S (Horowitz et al., 2006; Stone, 1985, 1986).

All additive model structures exhibit an important restriction:

$$\frac{\partial^2 m_t(c_1, ..., c_S)}{\partial c_s \partial c_{s'}} = 0 \tag{19}$$

for all $s \neq s'$. Equation (19) demonstrates that one characteristic's predictive power does not rely on another characteristic, as they are conditionally independent under the additive model. Consequently, the additive model does not consider cross-dependencies, which can be addressed by manually including interactions as regressors. These regressors involve multiplying one characteristic by another, such as examining whether a firm's size truly impacts the analysis by interacting each characteristic with size. Alternatively, one may estimate the model separately for small and large firms. Brandt et al. (2009) stress that one can always interpret characteristics as the cross-product of a more basic set of characteristics.

Adopting an additive model structure yields favorable econometric benefits, albeit with a restrictive assumption. Nevertheless, the assumption is less restrictive than the assumptions of additivity and linearity utilized in Fama-MacBeth regressions. Moreover, multivariate regressions frequently rely on comparable assumptions. Another advantage of the additive model structure is that it permits the joint estimation of a model with numerous characteristics, the identification of significant characteristics, and the simultaneous estimation of each individual characteristic's summand for the complete conditional mean function, m_t. At last, the structure also permits extrapolation to areas with limited data points on the joint characteristic support from areas with more data points by averaging over the marginal characteristic distribution. Without any further additive model assumptions, imprecise estimates would result in sparsely distributed joint characteristic situations.

4.2.2 Normalised Conditional Mean Function

A particular characteristic normalisation is utilised to enable the mapping of the nonparametric estimator to portfolio sorts while minimising outlier effects. The normalisation approach enhances the interpretability of the outcomes as one of its benefits. The conditional mean function, m_t, for S characteristics is defined as before:

$$m_t(C_{1,it-1}, ..., C_{S,it-1}) = \mathbb{E}[R_{it}|C_{1,it-1}, ..., C_{S,it-1}] \tag{20}$$

For each characteristic s, let $F_{s,t}(\cdot)$ be a known strictly monotone function with its corresponding inverse $F_{s,t}^{-1}(\cdot)$. Define the normalisation of the characteristic $\tilde{C}_{s,it-1} = F_{s,t}(C_{s,it-1})$ and

$$\tilde{m}_t(C_{1,t-1}, ..., C_{S,it-1}) = m_t(F_{1,t}^{-1}(C_{1,it-1}), ..., F_{S,t}^{-1}(C_{S,it-1})) \tag{21}$$

Then,

$$\tilde{m}_t(\tilde{C}_{1,it-1}, ..., \tilde{C}_{S,it-1}) = m_t(C_{1,it-1}, ..., C_{S,it-1}) \tag{22}$$

Equation (22) indicates that utilising normalised characteristics in the normalised conditional mean function is equivalent to utilising regular characteristics in the regular conditional mean function. Therefore, knowledge of the regular conditional mean function, m_t, is equal to knowing the normalised conditional mean function, $\tilde{m}_t$, which is the targeted estimation function, and vice versa. As no supplementary constraints are applied, this transformation does not lead to a loss of generality.

As the transformed conditional mean function, $\tilde{m}_t$, offers favorable properties and improved interpretability because of its linkage to portfolio sorts, it is preferable to estimate this function rather than the regular conditional mean function, m_t. When financial products are grouped into portfolios based on a characteristic, their absolute value is usually not considered in isolation. Instead, the rank or relative value of the characteristic in the cross-section is taken into account. To achieve this, the characteristic is normalised through a rank transformation that leads to relative values instead of absolute values, akin to portfolio sorting. A more detailed description of the rank transformation function can be found in Section 3.3. All aforementioned and future econometric theory also applies to any other monotonic transformation or the non-transformed conditional mean function.

Although knowing either of the two conditional mean functions means that one is able to reconstruct the other, the estimates in finite samples are typically different. This is shown by Equation (23), in which the arguments on the right-hand size do not coincide with the transformed conditional mean function, and therefore the equation does not hold.

$$\hat{m}_t(c_1, ...c_s) \neq \hat{\tilde{m}}_t(F_{1,t}^{-1}(c_1), ..., F_{S,t}^{-1}(C_S)) \tag{23}$$

Freyberger et al. (2020) performed simulations and empirical applications and concluded that the transformed conditional mean function, $\tilde{m}_t$, provides better out-of-sample predictions than the conditional mean function, m_t. The transformed estimator appears to be more robust to outliers due to the transformation, something that addresses to the concern of Cochrane (2011) about the sensitivity of regressions to outliers.

4.2.3 Adaptive Group LASSO

In order to identify and estimate the characteristics that offer additional information for the expected delta-hedged option returns across different entities, the adaptive group LASSO technique proposed by J. Huang et al. (2010) is employed. This approach consists of a two-step procedure specifically designed for additive model structures. With the introduction of the transformed conditional mean function, the focus now shifts to modelling the excess delta-hedged option returns as a function of these characteristics:

$$R_{it} = \sum_{s=1}^{S} \tilde{m}_{ts}(\tilde{C}_{s,it-1}) + \epsilon_{it} \tag{24}$$

where $\tilde{m}_{ts}(\cdot)$ are the unknown and to be estimated functions, and $\tilde{C}_{s,it-1}$ the rank-transformed characteristic.

The primary concept behind the LASSO procedure is to estimate functions in a nonparametric manner. In other words, the procedure evaluates the impact of characteristics that provide additional information on the cross-section of delta-hedged option returns while setting functions to zero if they do not influence these returns. As such, the procedure performs model selection by distinguishing between constant null functions and variable functions.

n portfolio sorts, constant splines are used to approximate the summands of the transformed conditional mean function within each portfolio. However, instead of using constant splines, we propose to estimate quadratic functions over the normalised characteristic distribution, which is similar to the approach taken by Freyberger et al. (2020). To illustrate, let $0 = t_0 < t_1 < ... < t_{L-1} < t_L = 1$ a sequence of ordered numbers between zero and one similar to portfolio breakpoints, and let $\tilde{I}_l$ for $l = 1, ..., L$ be the partition of the interval, $\tilde{I}_l = [t_{l-1}, t_l)$ and $\tilde{I}_L = [t_{L-1}, t_L]$. The numbers $t_0, ..., t_{L-1}$ are knots and we choose $t_l = l/L$ for all $l = 0, ..., L-1$. Due to the rank-transformation of the characteristics, the knots can be interpreted as the quantiles of the characteristics distribution, where $\tilde{I}_l$ is the l-th portfolio.

Therefore, to estimate the transformed conditional mean function, $\tilde{m}_t$, we use quadratic splines. That is, the transformed conditional mean function is approximated as a quadratic function on the intervals $\tilde{I}_l$. The functions are chosen in such a way that the endpoints are connected and the transformed conditional mean function is differentiable on the interval $[0, 1]$. Thus, we can write the summands of the transformed conditional mean function as the following series expansion:

$$\tilde{m}_{ts}(\tilde{c}) \approx \sum_{k=1}^{L+2} \beta_{tsk} p_k(\tilde{c}) \tag{25}$$

where $p_k(c)$ are known functions.

The user-specific parameter L corresponds to the number of portfolios. The precision of the approximation is enhanced as the number of portfolios increases, but at the same time, the number of parameters increases, leading to an increase in variance. Thus, the number of portfolios, L, is regarded as a smoothing parameter, as there is a trade-off between precision and variance. Compared to the constant splines used in portfolio sorts, this estimator is more smooth and flexible (Freyberger et al., 2020).

The adaptive group LASSO procedure employs the series expansion of the transformed conditional mean function. This two-step approach starts with the group LASSO estimation, where the estimates are obtained as follows:

$$\tilde{\boldsymbol{\beta}}_t = \underset{b_{sk}:s=1,\ldots,S;k=1,\ldots,L+2}{\operatorname{argmin}} \sum_{i=1}^{N}\left(R_{it} - \sum_{s=1}^{S}\sum_{k=1}^{L+2} b_{sk}p_k(\tilde{C}_{s,it-1})\right)^2 + \lambda_1 \sum_{s=1}^{S}\left(\sum_{k=1}^{L+2} b_{sk}^2\right)^{\frac{1}{2}} \tag{26}$$

where we obtain the $(L+2)$ x S matrix of estimates $\tilde{\boldsymbol{\beta}}_t$ and use the penalty parameter λ_1.

Equation (26) consists of two parts. The first step computes the sum of squared residuals, while the second step applies the group LASSO penalty function. This penalty function penalizes the entire group of k coefficients associated with a characteristic s, rather than individual coefficients, thereby identifying the entire expansion that provides sufficient incremental information for the cross-section of delta-hedged option returns. The penalty function also makes it feasible to use this procedure even when the number of characteristics exceeds the sample size. Furthermore, the penalty term λ_1 is chosen by minimising the Bayesian Information Criterion (BIC) (Yuan & Lin, 2006).

Unless the design matrix exhibits restrictive conditions, the first step of the adaptive LASSO procedure selects too many characteristics (Meinshausen & Bühlmann, 2006; Zou, 2006). The adaptive LASSO procedure not only selects informative characteristics, but also some uninformative ones. To address this issue, the second step of the procedure incorporates characteristic-specific weights in the penalty function. The weights are constructed by using the estimated coefficients from the first step.

$$w_{ts} = \begin{cases} \left(\sum_{k=1}^{L+2} \tilde{\beta}_{sk}^2\right)^{-\frac{1}{2}} & \text{if } \sum_{k=1}^{L+2} \tilde{\beta}_{sk}^2 \neq 0 \\ \infty & \text{if } \sum_{k=1}^{L+2} \tilde{\beta}_{sk}^2 = 0 \end{cases} \tag{27}$$

The second step does not select any characteristic that was not selected in the first step, by construction.

The second step constructs the estimates as follows:

$$\breve{\boldsymbol{\beta}}_t = \underset{b_{sk}:s=1,\ldots,S;k=1,\ldots,L+2}{\operatorname{argmin}} \sum_{i=1}^{N}\left(R_{it} - \sum_{s=1}^{S}\sum_{k=1}^{L+2} b_{sk}p_k(\tilde{C}_{s,it-1})\right)^2 + \lambda_2 \sum_{s=1}^{S}\left(w_{ts}\sum_{k=1}^{L+2} b_{sk}^2\right)^{\frac{1}{2}} \tag{28}$$

where the characteristics specific weights are implemented in the second part. Again, the BIC is used to construct the penalty parameter λ_2. The estimation obtained in the second step of the adaptive LASSO procedure is considered to be model-selection consistent when the sample size is large enough. This means that the procedure correctly selects non-constant functions with a high probability close to one (J. Huang et al., 2010).

The resulting estimated function equals:

$$\hat{\tilde{m}}_{ts}(\tilde{c}) = \sum_{k=1}^{L+2} \hat{\beta}_{tsk}p_k(\tilde{c}) \tag{29}$$

If the cross-section is large enough, the model selection and estimation can be done on a period-by-period basis. The main advantage of this approach is that it allows one to observe the significance of characteristics and the shape of the conditional mean function over time. However, pooling the data and estimating the conditional mean function as time-invariant would yield more accurate estimates and therefore more reliable predictions.

4.2.4 Conditional Mean Function Interpretation

In nonparametric additive models, the locations of the estimates are unidentified. Let α_s be S constants such that $\sum_{s=1}^{S} \alpha_s = 0$. Then, the following holds:

$$\tilde{m}_t(\tilde{c}_1, ..., \tilde{c}_S) = \sum_{s=1}^{S} \tilde{m}_{ts}(\tilde{c}_s) = \sum_{s=1}^{S} (\tilde{m}_{ts}(\tilde{c}_s) + \alpha_s) \tag{30}$$

The estimation locations are now determined. The constant terms in the estimation results are not relevant for the model selection, estimation of expected returns, and portfolio construction, as they do not affect these procedures. These constants are only important if one wants to visualize the estimation results of the conditional mean function of a specific characteristic.

4.2.5 Tuning Parameters

The proposed method has several tuning parameters that are specific to the research and can be manually adjusted, but can significantly alter important results. These parameters include the model selection method, penalty parameter, number of knots, and order of splines/polynomials. Modifying any of these parameters while keeping other factors constant may lead to different outcomes in terms of characteristic selection and out-of-sample predictions. Freyberger et al. (2020) conducted a simulation study to examine how sensitive the results of the proposed method are to changes in the tuning parameters. The study assumed 13 characteristics as the ”true” predictors and generated returns using a fifth-order polynomial based on these predictors. In addition, returns were also generated using a linear (first-order polynomial) process to compare the performance of different model selection methods under different data-generating processes. For a comprehensive description of the simulation setup, I please refer to the work by Freyberger et al. (2020). The main results of their study are crucial for the proposed framework, as they can help identify the characteristics that have a conditional impact on the cross-section of delta-hedged option returns.

Freyberger et al. (2020) consider the following model selection methods: the t-statistic significance value of 2, the increased t-statistic significance value of 3 for multiple testing (Harvey et al., 2016), the False Discovery Rate (FDR) p-value adjustment (Green et al., 2017), the linear single-step LASSO, the linear adaptive LASSO, the nonlinear group LASSO, and the nonlinear adaptive group LASSO. The linear single-step LASSO corresponds to the linear version of the nonlinear group LASSO, while the linear adaptive LASSO is the linear equivalent of the nonlinear adaptive group LASSO in two steps. Furthermore, they consider several information criteria: Akaike information criteria (AIC), Bayesian information criteria (BIC), adjusted BIC (Yuan & Lin, 2006), and tenfold cross-validation.

The information criteria differ in their treatment of the penalty term that balances the cost of adding parameters with the goodness of fit of the model. AIC applies a penalty equal to twice the number of parameters, while BIC uses the natural logarithm of the product of the number of observations and the number of parameters. Yuan and Lin (2006) propose an adaptation of the BIC that accounts for grouped variables. The tenfold cross-validation method partitions the dataset into ten subsamples, where nine are used for model estimation and the remaining one is used for predicting out-of-sample returns. This process is repeated nine times, ensuring that each subsample is used once for model validation. Finally, the penalty parameter with the lowest mean-squared prediction error is selected. Finally, the parameters for the number of knots and the order of the polynomial vary between 10 to 25 and 0 to 4, respectively.

At first, Freyberger et al. (2020) identified the tuning parameters that choose the pertinent characteristics with high probability while excluding the irrelevant ones. The reference model employed the nonlinear adaptive group LASSO approach with the adjusted BIC from Yuan and Lin (2006), 20 knots, and a second-order polynomial. Any variations to the parameters in the reference model were made while keeping all other conditions constant. found that nonlinear LASSO methods outperformed linear methods and other techniques. They also observed that two-step approaches were better than single-step methods when the data-generating process was nonlinear, indicating a preference for the nonlinear adaptive group LASSO. However, in the case of a linear data-generating process, the nonlinear adaptive group LASSO performed similarly to its linear counterpart. Therefore, it appears reasonable to permit nonlinearities. Based on the model selection performance, either of the BIC methods for the penalty parameter is recommended, as they perform equally well and significantly better than AIC or tenfold cross-validation. The number of knots parameter has a negligible effect on the model's performance, as all values demonstrate similar performance in terms of selecting the correct characteristics and omitting irrelevant ones with a high probability. At last, the second-order polynomial appears to be a slightly better choice than the step function (zero-order) due to its performance. Furthermore, using higher orders of splines leads to the selection of more characteristics than necessary. Overall, the baseline model appears to be a reasonable and justifiable choice.

Although a model may perform well in terms of model selection, it does not necessarily mean that it will perform well in out-of-sample scenarios for specific tuning parameters. As a result, to assess the performance of specific tuning parameters, Freyberger et al. (2020) estimate all alternative models to the baseline model during a specific time horizon and forecast out-of-sample returns. They measure the quality of these predictions using the root-mean-squared prediction error (RMSPE) and calculate the R^2 by regressing the realized returns on the predicted returns. Assuming a nonlinear data-generating process results in a substantial improvement in R^2 and a significant decrease in RMSPE for the nonlinear models in comparison to other model selection techniques. On the other hand, if the data-generating process is linear, the measures for the nonlinear models are almost identical to those of the linear models. Therefore, it is reasonable to use a nonlinear model a priori. The findings regarding the penalty parameters and number of knots are consistent with the discussion above: either BIC method is preferred, and changing the number of knots has minimal effect on performance, although performance deteriorates slightly with 25 or more knots. Additionally, a clear pattern emerges in the out-of-sample performance related to the polynomial order. Increasing the order of the polynomial leads to higher R^2 and lower RMSPE, but the improvement in performance from a first-order to fourth-order polynomial is small. The step function performs worse, which is logical as a fifth-order polynomial was used to simulate returns from the "true" predictors.

5 Results

5.1 Portfolio Sorting and Fama-MacBeth Regressions

To explore the possible factors affecting the expected returns of delta-hedged option portfolios, we initially analyze the role of the 130 characteristics through conventional methods and compare the outcomes with those of machine learning techniques. At the end of every month, we categorize delta-hedged option portfolios into five groups based on the value of each characteristic and calculate the return of the hedge position rebalanced on a daily basis during the subsequent month. The quintiles with the lowest and highest characteristic levels comprise the low and high quintiles, respectively. Table 5 presents the Newey and West (1994) adjusted t-statistics (t-stat) and the spreads between the average monthly delta-hedged returns of equally-weighted option portfolios in the top and bottom quintiles.

The findings are consistent with those of Brooks et al. (2018), which is not unexpected given the substantial overlap in data used. However, our analysis incorporates data up to the end of 2021, while Brooks et al. (2018) use data up to 2019. The majority of the characteristics show a significant and substantial difference in returns between the high and low option portfolios. The characteristics that exhibit the largest positive spread are consistent between call and put options and include size (mve), dollar trading volume (dolvol), industry-adjusted size (mve_ia), number of analysts covering the stock (nanalyst), the difference between historical and implied volatility (hv_iv), option's vega (vega), and dollar trading volume (dolvol). Conversely, the variables with the most negative spreads are average implied volatility (ivol_ave), implied volatility on the day prior to portfolio formation (impl_vol), R&D to sales (rd_sale), illiquidity (ill), idiosyncratic volatility (idiovol), and total volatility (tvol).

The outcomes of our research through the portfolio sorting method are in agreement with prior investigations that examine the factors that impact delta-hedged option returns, such as the divergence between implied and historical volatility, hv_iv (Goyal & Saretto, 2009), and idiosyncratic volatility, idiovol (Cao & Han, 2013). Nonetheless, portfolio sorting suffers from the"curse of dimensionality" and does not provide definite guidance on the autonomous effect of each characteristic on delta-hedged option returns, given other variables.

The use of FM regressions allows for the examination of the combined predictive ability of all characteristics. Table 5 presents the outcomes of FM regressions that analyze the relationship between delta-hedged option returns and rank-transformed characteristics. However, this approach has a disadvantage of overfitting the data in-sample, which may lead to inadequate out-of-sample performance. There are disparities in the number of predictive characteristics for delta-hedged returns between the portfolio sorting and Fama-MacBeth regression methods, and some variations in the direction of the relationship found by each approach. The Fama-MacBeth approach's contradictory findings may reflect either the independent contribution of each characteristic or noise resulting from overfitting the data. To address this issue, we employ the adaptive group LASSO method to construct and estimate a more parsimonious models that incorporate characteristics that are crucial for option expected returns.

5.2 Adaptive Group LASSO: Results

This section presents the outcomes of the model selection for both the one-step Group LASSO and two-step Adaptive Group LASSO approaches. Table 2 displays the chosen characteristics using the Group LASSO and Adaptive LASSO methods over the complete sample period from January 1996 to December 2021. The baseline model of Freyberger et al. (2020) is employed for the Adaptive Group LASSO, as described in 4.2.5. To maintain consistency between the methods, identical tuning parameter values are employed in the Group LASSO procedure.

Table 2: Selected Characteristics for Group LASSO and Adaptive LASSO

Panel A. Call Options		
Method	Group LASSO	Adaptive Group LASSO
Number Selected Variables	11	8
Characteristics	divol dy herf vega hv_iv idiovol mve zerotrade operprof rd_sale stdcf	divol dy herf vega hv_iv idiovol mve zerotrade
Panel B. Put Options		
Method	Group LASSO	Adaptive Group LASSO
Number Selected Variables	7	6
Characteristics	civpiv divol hv_iv idiovol mve roaq chfeps	civpiv divol hv_iv idiovol mve roaq

The Group LASSO and Adaptive Group LASSO methods select significantly fewer characteristics compared to the portfolio sorting method and linear Fama-MacBeth regressions. The Group LASSO selects 11 characteristics for calls and 7 for puts, while the Adaptive Group LASSO selects 8 for calls and 6 for puts. Additionally, the two-step Adaptive Group LASSO approach selects fewer characteristics than the one-step Group LASSO. Comparing the results with those of linear LASSO procedures used in Brooks et al. (2018), it is evident that the nonlinear methods have a lower number of characteristics. These results are consistent with the literature (Freyberger et al., 2020), which suggests that nonlinear methods tend to select fewer characteristics, and that the two-step approach selects fewer or the same number of characteristics compared to the one-step approach.

Additionally, there is a considerable overlap in the predictors of call option returns and put option returns. Various characteristics related to options and stocks are chosen, including implied volatility innovation (divol), deviation of implied volatility from historical volatility (hv_iv), idiosyncratic volatility (idiovol), size (mve).

Furthermore, some characteristics are found to be significant only for call options or put options but not both. For example, dividend yield (dy) and the amount of zero trading days (zerotrade) are significant predictors for call options, while the call-put implied volatility spread (civpiv) and the return on assets (roaq) are significant only for put options.

5.3 Time Variation in Predictors

If the chosen predictors contribute to returns due to mispricing or market timing rather than rational compensation for risk, then their contribution to returns would vary over time or the predictors themselves might change over time. Additionally, the predictive power of characteristics for returns could be diminished by data mining and academic publications, as found in McLean and Pontiff (2016) and Harvey et al. (2016). Freyberger et al. (2020) discovered a time variation in the set of characteristics chosen using the Adaptive LASSO shrinkage model for the cross-section of stock returns. This section explores the variation of option return predictors by utilizing model selection techniques over time.

The model is selected and estimated using ten years of data, with the first selection period being January 1996 until December 2005. Then, the selection period is rolled forward by one year while the selection window remains constant at ten years, and this process is continued until the end of the sample. In all, we perform 17 rolling model selections using the Adaptive Group LASSO. Table 3 depicts the outcomes of the rolling selection process.

Once more, it is noticeable that certain characteristics affect both call and put options, while some have a more significant impact on either calls or puts over time. Additionally, there are characteristics that impact call options but not put options, and these effects vary over time. The selection of certain characteristics over time supports the idea that these variables consistently offer relevant information for delta-hedged option returns. Additionally, the Adaptive Group LASSO method's selected model dimension varies over time when ten years of data are used. The period between 2006-2015 yields the largest set of 21 predictors for call options, while the smallest set of 5 predictors is found during 1996-2005. As for put options, the largest set of 24 predictors is obtained for 2008-2017, and the smallest set is found during 1996-2005.

Table 3: Time Variation in Selected Characteristics using Adaptive Group LASSO

	Year	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21
Panel A. Call Options																		
Variables	# Selected	5	5	14	9	10	14	15	17	19	20	21	20	17	17	17	14	10
divol	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
hv_iv	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
idiovol	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
zerotrade	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
lev	15			x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
mve	15	x	x	x	x	x	x	x	x	x	x	x	x			x	x	x
roaq	13			x	x	x	x	x	x	x	x	x	x	x	x	x		
orgcap	13					x	x	x	x	x	x	x	x	x	x	x	x	x
pctacc	12			x			x	x	x	x	x	x	x	x	x	x	x	
kurt_vol	10							x	x	x	x	x	x	x	x	x	x	
delta	10							x	x	x	x	x	x	x	x	x	x	
rd_sale	9					x	x	x	x	x	x	x	x	x				
dy	9							x		x	x	x	x	x	x	x		x
herf	8								x	x	x	x	x	x	x	x		
saleinv	8			x	x			x	x	x	x	x	x					
cashdebt	8						x	x		x	x	x	x	x				x
baspread	8			x			x	x		x	x	x	x				x	
gma	7										x	x	x	x	x	x	x	
os	7			x							x	x	x		x	x		x
std_dolvol	6										x	x	x	x	x	x		
vov	6			x	x	x		x		x							x	
bm	5			x			x	x	x								x	
ep	4							x	x	x					x			
operprof	4			x			x	x	x									
pchsaleinv	4											x		x	x	x		
Panel B. Put Options																		
Variables	# Selected	8	9	10	11	13	13	12	13	15	17	19	23	24	23	19	14	12
divol	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
hv_iv	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
idiovol	17	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
mve	16	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
zerotrade	15	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		
delta	14	x	x	x	x	x	x	x	x	x	x	x		x	x		x	
tvolume	14			x	x	x	x	x	x	x	x	x	x	x	x	x	x	
vov	13		x	x		x	x	x	x	x	x	x	x	x	x	x		
kurt_vol	13				x	x	x	x	x	x	x	x	x	x	x	x		x
std_dolvol	12				x	x	x	x	x	x	x	x	x	x	x	x		x
sfe	10	x	x	x	x	x	x	x	x								x	x
civpiv_ave	9									x	x	x	x	x	x	x	x	x
os	9									x	x	x	x	x	x	x	x	x
pctacc	8									x	x	x	x	x	x	x	x	
roaq	7								x	x	x	x	x	x	x			
cash	7										x	x	x	x	x	x	x	
fgr5yr	7	x	x	x	x	x	x											x
mom12m	6										x	x	x	x	x	x		
rd_sale	6								x	x		x	x	x	x			
agr	6					x	x	x					x	x				x
bm_ia	6										x	x	x	x	x		x	
cashdebt	6											x	x	x	x	x		x
absacc	5												x	x	x	x		x
lev	5												x	x	x	x	x	
mom36m	4												x	x	x	x		
opt_demand	4												x	x	x	x	x	

5.4 Out-of-Sample Performance

This section assesses the ability of the Adaptive Group LASSO model to predict delta-hedged option returns in out-of-sample data, and compares their performance to that of Fama-MacBeth regressions. In order to predict one-month ahead returns using the information available at the end of the previous month, we use the Adaptive Group LASSO method for model selection and estimation with a ten-year dataset. In the first instance of out-of-sample prediction, we estimate the model using delta-hedged option returns from February 1996 to January 2006, and option and stock-level characteristics from January 1996 to December 2005. Then, using the estimated coefficients and the characteristics available at the end of January 2006, we predict the delta-hedged option returns for February 2006. Subsequently, we move the sample used for selecting and estimating the model one month ahead to obtain new estimates of the coefficients. The selection and estimation window remains constant at a range of 10 years. This process is repeated until the end of the sample period. The technique of rolling selection considers the influence of time-varying predictors on option returns. I adopt a similar approach to generate Fama-MacBeth forecasts, but in contrast to the Adaptive Group LASSO method, the model remains constant over time and incorporates the entire set of predictors.

Since the Adaptive Group LASSO has a one-to-one relationship with portfolio sorts, only the predictive power of the return forecasts at portfolio level is considered (Lewellen, 2014). Table 4 evaluates the performance of decile portfolios when sorted by predicted returns, and compares the average realized delta-hedged returns across the portfolios. The Table displays various performance metrics for the predicted and realized delta-hedged returns, including the mean monthly predicted delta-hedged returns, mean monthly realized delta-hedged returns, standard deviation of the realized returns, and the annualized Sharpe ratio.

Table 4: Comparison Out-of-Sample Performance of Adaptive Group LASSO and Fama-MacBeth Regressions

	Panel A. Call Options		Panel B. Put Options	
Model	Adaptive Group LASSO	FM	Adaptive Group LASSO	FM
Pred Mean	2.324	3.149	1.728	2.614
Real Mean	2.109	2.158	1.821	1.911
t-stat	11.340	15.558	15.837	18.130
Std Dev	2.104	1.931	1.418	1.431
Sharpe	3.568	4.103	4.348	4.812

There exists a notable and statistically significant difference between the delta-hedged option portfolios in the top and bottom deciles of predicted returns. This indicates that the predicted returns correspond to the actual returns of delta-hedged portfolios. It is important to note that in analyzing the out-of-sample performance of predictive models at the portfolio level, the high-dimensional Fama-MacBeth regressions perform slightly worse than the sparse models obtained from the Adaptive Group LASSO method. Specifically, the Sharpe ratios of the Fama-MacBeth regressions are observed to be higher for both the call and put portfolios. However, it is crucial to examine if the predicted returns align with the realized returns. The Adaptive Group LASSO method predicts returns more accurately than the Fama-MacBeth regressions. Thus, the higher Sharpe ratio of the Fama-MacBeth regressions may lead to an incorrect interpretation of the outcomes. The Sharpe ratios of options strategies are significantly higher compared to the ones reported in literature on stock market anomalies. This could be attributed to the greater volatility of stochastic discount factors present in options (Cao et al., 2021).

In all, the out-of-sample performance of the sparse models selected and estimated using the Adaptive Group LASSO regularization technique in predicting delta-hedged option returns at the portfolio level is slightly better than that of the Fama-MacBeth regressions. This suggests that shrinkage methods can effectively identify the characteristics that provide essential information about the subsequent returns of delta-hedged option portfolios.

6 Conclusion

This study employs a nonparametric approach to address the issue raised by Cochrane (2011) regarding the identification of characteristics that offer incremental information for individual equity option returns. To achieve this, the Adaptive Group LASSO method is utilized for both model selection and estimation, with a dataset consisting of 130 potential predictors.

The paper demonstrates three applications of the proposed framework. Firstly, it identifies the characteristics that provide incremental information for forecasting delta-hedged option returns. Secondly, it examines whether there is time-variation in the estimation using a rolling window technique. Finally, it compares the out-of-sample performance of the Adaptive Group LASSO to that of the linear Fama-MacBeth regression.

The Adaptive Group LASSO technique selects a limited number of potential predictors for the cross-section of delta-hedged option returns when compared to other methods like portfolio sorting and Fama-MacBeth regressions. This suggests that conditioning on the characteristics has a significant impact on the results. There are predictors that affect both call and put options, while others only impact delta-hedged call or put option returns.

Moreover, the results exhibit significant time-varying behavior. The characteristics that were selected in the full sample are often also selected in a rolling window approach. Nevertheless, some other variables may sometimes be preferred to forecast the time-varying delta-hedged option returns. Furthermore, the number of predictors included in the model also changes over time.

Finally, the nonparametric and nonlinear Adaptive Group LASSO method shows a slightly better out-of-sample performance than the linear Fama-MacBeth approach. This is mainly evidenced by the smaller difference between the predicted and realized returns obtained by the former method. However, it is worth noting that the Sharpe ratios of the Fama-MacBeth regressions are higher, which could potentially lead to misleading interpretations of the results.

Future research could explore the relationship between the selected characteristics and factor exposures in delta-hedged option returns. Additionally, one could investigate the significance of the identified factors and construct investment strategies based on them.

References

Abdel-Khalik, A. R., & Chen, P.-C. (2015). Growth in financial derivatives: The public policy and accounting incentives. *Journal of Accounting and Public Policy*, *34*(3), 291–318.

Bakshi, G., & Kapadia, N. (2003). Delta-hedged gains and the negative market volatility risk premium. *The Review of Financial Studies*, *16*(2), 527–566.

Bertsimas, D., Kogan, L., & Lo, A. W. (2001). Hedging derivative securities and incomplete markets: An ϵ-arbitrage approach. *Operations research*, *49*(3), 372–397.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, *81*(3), 637–654.

Bollen, N. P., & Whaley, R. E. (2004). Does net buying pressure affect the shape of implied volatility functions? *The Journal of Finance*, *59*(2), 711–753.

Boyer, B. H., & Vorkink, K. (2014). Stock options as lotteries. *The Journal of Finance*, *69*(4), 1485–1527.

Brandt, M. W., Santa-Clara, P., & Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, *22*(9), 3411–3447.

Brooks, R., Chance, D. M., & Shafaati, M. (2018). The cross-section of individual equity option returns. *Unpublished working Paper. Lousiana State University.*

Büchner, M., & Kelly, B. (2022). A factor model for option returns. *Journal of Financial Economics*, *143*(3), 1140–1161.

Byun, S.-J., & Kim, D.-H. (2016). Gambling preference and individual equity option returns. *Journal of Financial Economics*, *122*(1), 155–174.

Cao, J., & Han, B. (2013). Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, *108*(1), 231–249.

Cao, J., Han, B., Zhan, X., & Tong, Q. (2021). Option return predictability. *Review of Financial Studies accepted, 27th Annual Conference on Financial Economics and Accounting Paper, Rotman School of Management Working Paper*, (2698267).

Cao, J., Vasquez, A., Xiao, X., & Zhan, X. (2022). Why does volatility uncertainty predict equity option returns? *Available at SSRN 3178263.*

Carr, P., & Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, *22*(3), 1311–1341.

Chinco, A., Clark-Joseph, A. D., & Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, *74*(1), 449–492.

Chinco, A., Neuhierl, A., & Weber, M. (2021). Estimating the anomaly base rate. *Journal of financial economics*, *140*(1), 101–126.

Choy, S. K., & Wei, J. Z. (2016). Liquidity risk and expected option returns. *Rotman School of Management Working Paper*, (2713915).

Christoffersen, P., Fournier, M., & Jacobs, K. (2018). The factor structure in equity options. *The Review of Financial Studies*, *31*(2), 595–637.

Christoffersen, P., Goyenko, R., Jacobs, K., & Karoui, M. (2018). Illiquidity premia in the equity options market. *The Review of Financial Studies*, *31*(3), 811–851.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, *66*(4), 1047–1108.

Coval, J. D., & Shumway, T. (2001). Expected option returns. *The journal of Finance*, *56*(3), 983–1009.

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of financial Economics*, *7*(3), 229–263.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., & Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, *33*(5), 2180–2222.

Driessen, J., Maenhout, P. J., & Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance*, *64*(3), 1377–1406.

Duan, J.-C., & Wei, J. (2009). Systematic risk and the price structure of individual equity options. *The Review of Financial studies*, *22*(5), 1981–2006.

Duarte, J., Jones, C. S., & Wang, J. L. (2019). Very noisy option prices and inference regarding the volatility risk premium. *Available at SSRN 4167308*.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, *47*(2), 427–465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3–56.

Fama, E. F., & French, K. R. (2008). Dissecting anomalies. *The Journal of Finance*, *63*(4), 1653–1678.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, *116*(1), 1–22.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, *81*(3), 607–636.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, *75*(3), 1327–1370.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Garleanu, N., Pedersen, L. H., & Poteshman, A. M. (2008). Demand-based option pricing. *The Review of Financial Studies*, *22*(10), 4259–4299.

Gibbons, M. R., Ross, S. A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, *57*(5), 1121–1152.

Goto, S., & Xu, Y. (2015). Improving mean variance optimization through sparse hedging restrictions. *Journal of Financial and Quantitative Analysis*, *50*(6), 1415–1441.

Goyal, A., & Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, *94*(2), 310–326.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, *30*(12), 4389–4436.

Han, Y., He, A., Rapach, D., & Zhou, G. (2018). What firm characteristics drive us stock returns. *Available at SSRN*, *3185335*.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Horenstein, A. R., Vasquez, A., & Xiao, X. (2020). Common factors in equity option returns. *Available at SSRN 3290363*.

Horowitz, J., Klemelä, J., & Mammen, E. (2006). Optimal estimation in additive regression models. *Bernoulli*, *12*(2), 271–298.

Huang, J., Horowitz, J. L., & Wei, F. (2010). Variable selection in nonparametric additive models. *Annals of statistics*, *38*(4), 2282.

Huang, J.-z., & Shi, Z. (2011). Determinants of bond risk premia. *Unpublished Manuscript, Pennsylvania State University*.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, *45*(3), 881–898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, *48*(1), 65–91.

Lewellen, J. (2014). The cross section of expected stock returns. *Forthcoming in Critical Finance Review, Tuck School of Business Working Paper*, (2511246).

Lintner, J. (1965). The valuation (or risk assets and the selection (or risky investments in stock portrolios and capital, budgets. *The Review of Economics and Statistics*, *47*(1), 13–37.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, *34*(3), 1436–1462.

Mossin, J. (1966). Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, 768–783.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, *61*(4), 631–653.

Ni, S. X., Pearson, N. D., & Poteshman, A. M. (2005). Stock price clustering on option expiration dates. *Journal of Financial Economics*, *78*(1), 49–87.

Ramachandran, L. S., & Tayal, J. (2021). Mispricing, short-sale constraints, and the cross-section of option returns. *Journal of Financial Economics*, *141*(1), 297–321.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the united states? *The Journal of Finance*, *68*(4), 1633–1662.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The journal of finance*, *19*(3), 425–442.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, 1040–1053.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The annals of Statistics*, *13*(2), 689–705.

Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, 590–606.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Vasquez, A. (2017). Equity volatility term structures and the cross section of option returns. *Journal of Financial and Quantitative Analysis*, *52*(6), 2727–2754.

Vasquez, A., & Xiao, X. (2020). Default risk and option returns. *Available at SSRN 2667930.*

Wang, H., & Leng, C. (2008). A note on adaptive group lasso. *Computational statistics & data analysis*, *52*(12), 5277–5286.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 49–67.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, *101*(476), 1418–1429.

Table 5: Comparative Analysis of Delta-Hedged Option Returns

		Panel A. Call Options				Panel B. Put Options			
		Portfolio sorts		Fama-MacBeth		Portfolio sorts		Fama-MacBeth	
Variable	s/o	H-L	t-stat	beta	t-stat	H-L	t-stat	beta	t-stat
absacc	s	-0.392	-6.37	-0.232	-1.69	-0.201	-3.25	-0.162	-1.22
acc	s	0.126	1.63	0.268	1.26	0.106	1.89	-0.005	-0.02
aeavol	s	0.203	3.21	0.088	1.27	0.188	2.83	0.188	2.73
age	s	0.622	6.54	-0.064	-0.72	0.265	2.61	-0.024	-0.21
agr	s	-0.052	-0.63	-0.142	-0.79	0.239	3.62	0.162	0.79
baspread	s	-0.859	-6.10	0.247	1.52	-0.552	-4.23	0.479	2.09
beta	s	-0.498	-4.27	0.023	0.27	-0.293	-2.26	0.104	0.64
betasq	s	-0.498	-4.27	-	-	-0.295	-2.29	-	-
bm_ia	s	-0.093	-1.02	-0.032	-0.28	-0.104	-1.45	-0.269	-2.02
bm	s	0.201	2.00	0.145	0.94	-0.108	-0.89	0.030	0.24
cashdebt	s	0.534	6.01	0.405	1.93	0.389	5.11	0.182	1.11
cashpr	s	-0.125	-1.93	-0.003	-0.02	0.239	3.18	0.102	0.93
cash	s	-0.629	-4,82	0.156	1.03	-0.342	-3.16	-0.175	-1.23
cfp_ia	s	0.252	3.83	0.052	0.48	0.078	1.05	0.037	0.32
cfp	s	0.824	7.68	-0.286	-1.88	0.242	2.32	-0.108	-0.52
chatoia	s	-0.064	-1.01	-0.061	-0.63	-0.105	-1.53	-0.138	-1.72
chcsho	s	-0.393	-4.99	-0.045	-0.43	-0.104	-1.13	0.042	0.57
chempia	s	0.163	2.42	0.271	2.34	0.109	2.17	-0.019	-0.19
chfeps	s	0.173	2.78	0.242	2.72	0.094	1.79	-0.084	-1.40
chinv	s	0.015	0.24	-0.152	-1.34	0.136	2.22	-0.093	-0.84
chmom	s	0.178	1.73	0.059	0.32	0.082	0.99	0.076	0.49
chnanalyst	s	-0.019	-0.33	-0.032	-0.49	0.097	1.97	0.110	1.62
chpmia	s	-0.183	-2.41	-0.045	-0.50	-0.050	-0.69	0.192	1.96
chtx	s	-0.026	-0.39	-0.042	-0.49	0.217	2.61	0.032	0.27
cinvest	s	-0.026	-0.39	-0.041	-0.49	0.212	2.58	0.029	0.23
civpiv_ave	o	-0.163	-1.79	-0.228	-1.70	0.362	6.01	0.293	2.83
convind	s	-0.103	-1.82	-0.072	-0.68	-0.072	-1.16	0.015	0.14
cpiv_hv_ratio	o	-0.742	-7.01	-	-	-0.519	-7.62	-	-
currat	s	-0.539	-5.92	0.049	0.46	-0.289	-2.96	-0.074	-0.88
dcivpiv_ave	o	-0.203	-3.45	0.004	0.05	0.156	2.53	-0.064	-0.83
depr	s	-0.532	-6.48	-0.239	-2.25	-0.176	-2.24	0.098	0.92
disp	s	-0.509	-5.49	0.082	0.86	-0.444	-4.49	0.119	1.39
divi	s	-0.108	-0.89	-0.009	-0.06	0.040	0.32	-0.272	-0.84
divol	o	-0.604	-6.72	-0.833	-7.73	-0.542	-8.01	-0.442	-6.58

Table 5: Comparative Analysis of Delta-Hedged Option Returns

		Panel A. Call Options				Panel B. Put Options			
		Portfolio sorts		Fama-MacBeth		Portfolio sorts		Fama-MacBeth	
Variable	s/o	H-L	t-stat	beta	t-stat	H-L	t-stat	beta	t-stat
divo	s	-0.293	-2.26	-0.138	-0.48	-0.100	-0.76	-0.216	-0.99
dolvol	s	1.295	8.17	-	-	1.097	8.74	-	-
dy	s	0.669	11.32	0.216	1.58	0.218	2.93	-0.079	-0.59
ear	s	0.083	1.11	-0.039	-0.49	0.121	1.78	-0.032	-0.45
egr	s	0.083	0.99	0.088	0.64	0.218	3.12	0.072	0.56
ep	s	0.832	6.72	0.128	0.70	0.392	4.53	-0.189	-1.54
fgr5yr	s	-0.327	-2.60	0.259	2.29	0.032	0.35	0.290	2.47
gma	s	0.252	2.62	0.501	2.60	0.289	3.39	0.266	1.81
grCAPX	s	0.051	0.69	0.143	1.59	0.112	1.81	0.083	0.58
grltnoa	s	0.107	1.68	-0.156	-1.61	0.190	3.29	-0.049	-0.55
herf	s	0.409	4.36	0.148	1.39	0.333	4.42	0.087	0.75
hire	s	-0.089	-1.01	0.109	0.82	0.107	1.54	0.086	0.62
hv_cpiv	o	0.934	8.32	-	-	0.604	8.63	-	-
idiovol	s	-1.132	-10.45	-2.083	-7.69	-0.699	-6.38	-1.348	-5.81
ill	s	-1.409	-9.24	-	-	-1.162	-9.85	-	-
indmom	s	0.108	0.99	-0.006	-0.05	0.129	1.30	0.001	0.02
invest	s	0.162	2.41	0.119	0.96	0.281	4.18	0.351	1.96
ivol_ave	o	-1.489	-10.23	-	-	-0.942	-7.42	-	-
kurt_vol	o	-0.229	-4.02	-0.208	-3.27	-0.168	-2.63	-0.016	-0.18
delta	o	-0.429	-6.29	-0.172	-1.73	-0.456	-5.32	-0.458	-4.40
gamma	o	-0.422	-3.42	-	-	-0.308	3.48	-	-
impl_vol	o	-1.572	-10.22	-	-	-1.018	-8.09	-	-
open_interest	o	0.358	4.11	-	-	0.350	4.51	-	-
theta	o	-0.496	-4.08	-	-	-0.245	-2.79	-	-
vega	o	1.238	8.15	0.093	0.62	0.867	8.33	0.298	2.76
volume	o	0.368	4.19	-0.029	-0.26	0.139	2.12	0.116	1.54
hv_iv	o	0.987	9.09	1.572	10.08	0.653	9.20	1.083	8.79
lev	s	0.538	4.59	0.862	3.43	0.099	0.91	0.503	2.38
lgr	s	0.034	0.36	-0.009	-0.05	0.139	2.43	-0.279	-1.61
maxret	s	-0.623	-5.73	0.091	0.86	-0.481	-4.43	-0.279	-1.91
mom12m	s	0.267	1.79	0.022	0.09	0.164	1.83	0.173	1.04
mom1m	s	0.092	0.72	-0.053	-0.51	-0.038	-0.51	0.001	0.01
mom36m	s	0.123	1.63	-0.049	-0.42	0.386	4.61	0.107	0.98
mom6m	s	0.344	2.48	0.138	0.61	0.106	1.23	-0.089	-0.49

Table 5: Comparative Analysis of Delta-Hedged Option Returns

		Panel A. Call Options				Panel B. Put Options			
		Portfolio sorts		Fama-MacBeth		Portfolio sorts		Fama-MacBeth	
Variable	s/o	H-L	t-stat	beta	t-stat	H-L	t-stat	beta	t-stat
mom6m	s	0.342	2.49	0.133	0.58	0.107	1.13	-0.089	-0.42
ms	s	0.470	6.42	0.117	1.06	0.233	2.73	-0.042	-0.47
mve_ia	s	1.045	9.91	-	-	0.881	8.77	-	-
mve	s	1.354	8.68	0.589	2.63	1.073	8.87	0.839	4.41
nanalyst	s	0.995	7.42	0.301	2.74	0.912	9.03	0.239	2.32
nincr	s	0.109	2.55	0.002	0.01	0.111	2.74	-0.042	-0.45
nopt	o	0.712	5.51	-	-	0.629	5.69	-	-
obklg	s	0.059	0.72	0.048	0.51	0.044	0.56	0.122	1.44
operprof	s	0.750	10.48	0.077	0.68	0.477	7.29	-0.187	-1.89
opt_demand	o	-0.064	-0.87	0.077	1.01	0.074	1.13	0.300	3.58
opt_dolvol	o	0.411	3.41	-	-	0.315	3.61	-	-
os	o	-0.084	-0.99	-0.155	-1.38	0.016	0.14	-0.291	-2.73
orgcap	s	-0.269	-3.00	-0.209	-1.76	-0.122	-1.43	-0.066	-0.79
pba	o	-0.301	4.11	0.228	2.15	-0.318	-5.86	0.001	0.04
pchcapx_ia	s	0.009	0.07	0.016	0.18	0.008	0.13	0.035	0.31
pchcurrat	s	-0.087	-1.37	-0.100	-1.28	-0.029	-0.54	-0.040	-0.53
pchdepr	s	0.028	0.48	0.089	1.11	-0.008	-1.57	0.069	0.91
pchgm_pchsale	s	0.009	0.13	0.047	0.53	0.063	1.10	-0.053	-0.69
pchquick	s	-0.081	-1.29	-	-	-0.039	-0.66	-	-
pchsale_pchinvt	s	-0.001	-0.03	0.228	1.56	-0.031	-0.68	0.033	0.23
pchsale_pchrect	s	0.012	0.17	0.074	1.01	-0.128	-2.09	-0.178	-2.44
pchsale_pchxsga	s	0.007	0.10	0.058	0.76	0.127	2.57	0.193	2.39
pchsaleinv	s	-0.019	-0.35	-0.200	-1.20	-0.038	-0.75	-0.048	-0.32
pcratio	o	0.338	4.84	0.033	0.37	0.229	3.71	0.015	0.19
pctacc	s	-0.223	-3.64	-0.664	-3.51	-0.084	-1.20	-0.199	-1.22
pricedelay	s	0.162	2.53	-0.092	-1.39	0.029	0.41	-0.029	-0.39
ps	s	0.313	5.95	-0.151	-1.55	0.246	5.61	0.059	0.64
quick	s	-0.505	-4.73	-	-	-0.359	-3.90	-	-
rd_mve	s	-0.538	-5.99	0.206	2.10	-0.377	-5.59	0.173	1.38
rd_sale	s	-1.872	-9.33	-0.158	-1.22	-1.222	-6.39	-0.215	-1.36
rd	s	-0.328	-3.60	0.064	0.41	-0.185	-2.39	-0.055	-0.49
realestate	s	0.053	0.69	-0.083	-1.14	0.006	0.08	-0.081	-1.07
relspread	o	-0.481	-4.75	-0.099	-0.71	-0.614	-7.34	0.088	0.92
roaq	s	0.611	8.30	0.421	2.16	0.519	7.04	0.398	1.78

Table 5: Comparative Analysis of Delta-Hedged Option Returns

		Panel A. Call Options				Panel B. Put Options			
		Portfolio sorts		Fama-MacBeth		Portfolio sorts		Fama-MacBeth	
Variable	s/o	H-L	t-stat	beta	t-stat	H-L	t-stat	beta	t-stat
roavol	s	-0.838	-10.93	0.134	1.33	-0.524	-6.45	-0.029	-0.22
roeq	s	0.644	8.74	-0.129	-0.71	0.439	7.02	-0.168	-1.11
roic	s	0.601	8.15	0.049	0.34	0.422	4.89	-0.068	-0.37
rsup	s	0.039	0.45	-0.079	-0.96	0.188	3.04	0.004	0.02
salecash	s	0.551	4.61	-	-	0.298	3.09	-	-
saleinv	s	-0.029	-0.57	-0.263	-2.45	-0.031	-0.49	-0.071	-0.69
salerec	s	-0.019	-0.21	-0.135	-1.37	0.001	0.00	0.150	1.51
sales_to_price	s	0.511	4.09	-	-	0.118	1.06	-	-
secured	s	-0.394	-5.04	0.059	0.74	-0.259	-4.21	-0.019	-0.28
sfe	s	0.542	4.98	-0.244	-2.40	0.371	3.77	0.391	3.45
sgr	s	-0.063	-0.82	-0.072	-0.41	0.177	2.61	-0.095	-0.54
shrtfee	o	-0.004	-0.09	-0.040	-0.38	-0.167	-2.73	0.229	2.34
skew	o	0.119	1.69	0.105	1.33	-0.199	-3.58	0.029	0.43
skew_vol	o	-0.253	-3.67	-0.091	-1.22	-0.129	-2.49	0.061	0.88
sratio	s	0.833	9.59	-0.199	-1.61	0.519	7.41	-0.180	-1.47
std_dolvol	s	-0.882	-9.71	-0.249	-2.81	-0.628	-8.52	-0.193	-2.25
std_turn	s	-0.442	-4.26	-	-	-0.219	-2.28	-	-
stdacc	s	-0.486	-6.01	-	-	-0.301	-3.75	-	-
stdcf	s	-0.497	-5.99	-0.061	-0.43	-0.344	-4.26	0.184	1.74
sue	s	0.089	1.31	0.164	1.91	0.030	0.58	-0.099	-1.32
tang	s	-0.429	-4.21	-0.036	-0.32	-0.191	-2.16	0.197	2.20
tb	s	0.285	4.26	-0.039	-0.54	0.249	3.21	0.102	1.21
toi	o	0.474	3.56	-	-	0.484	4.26	-	-
turn	s	-0.339	-2.64	-	-	-0.092	-0.97	-	-
tvol	s	-0.930	-6.59	-	-	-0.592	-5.05	-	-
tvolume	o	0.377	3.02	-0.217	-1.50	0.333	3.04	-0.212	-1.86
vov	o	-0.500	-5.48	-0.073	-0.84	-0.489	-8.30	-0.256	-3.71
zerotrade	s	0.162	1.24	-0.718	-5.58	0.019	0.19	-0.944	6.31