Erasmus School of Economics

Thesis Quantitative Finance

Identifying Factors of Delta-Hedged Equity Option Returns Using Adaptive Group LASSO

Sena Özdemir

450568

Supervisor: dr. Andrea A. Naghi

Second Assessor: dr. Maria Grith

March 2023

Abstract

We identify characteristics that provide incremental information on the cross-section of delta-hedged option returns using a nonparametric approach. We use the adaptive group least absolute shrinkage and selection operator (LASSO) to select characteristics and estimate the model and assess the robustness of selected characteristics over time. Many return predictors selected by conventional linear models provide no incremental information on returns when taking nonlinearities into account. After estimating the nonparametric model for various tuning parameters, only 8 characteristics are consistently selected to yield incremental information on delta-hedged option returns. The nonparametric model successfully achieves dimension reduction while also yielding improvements in out-of-sample Sharpe ratios and pricing errors compared to a conventional linear approach.

Contents

1 Introduction

Identifying factors that influence the movements of stock returns is an important question in asset pricing. A voluminous literature documents characteristics that explain the cross-section of stock returns (see, for example, Fama and French (1993, 2008, 2015), Chen et al. (2011), Hou et al. (2021), Green et al. (2017) and Harvey et al. (2016)) as well as corporate bonds (Kelly et al. (2020), Bai et al. (2019) and Bektić et al. (2019)). However, equity option returns have received less attention. Although options are generally perceived as simply being leveraged positions in the underlying assets, it is commonly accepted that options contain risk premiums that are not captured by common stock factors. Bakshi and Kapadia (2003) for example, show that delta-hedged option gains contain a nonzero volatility risk premium. Thus, an understanding of the characteristics that drive option returns provides an understanding of the cross-section of variance risk premiums.

Option analysis is traditionally done using linear Fama-Macbeth type regressions or through linear factor models. Using cross-sectional regressions Goyal and Saretto (2009a) argue that large deviations between implied volatility and historical volatility are indicative of option mispricing and that this effect cannot be captured by common risk factors. Cao and Han (2013) use a similar approach and that claim option returns decrease monotonically with idiosyncratic volatility.

Using Fama-Macbeth type linear regressions, Cao et al. (2021) propose a number of option strategies that provide significant profits even after accounting for stock factors. Initially, they document a dozen characteristics that can explain the cross-section of option returns. However, after taking idiosyncratic volatility and option illiquidity into account those profits become largely insignificant, implying that option returns can be explained through a more parsimonious model.

Jones (2006) uses a semi-parametric factor model to explain mispricing of put options in the S&P 500 index but does not succeed in explaining these abnormal excess returns. Another attempt to explain anomalies in the S&P 500 index option market using factor models is made by Constantinides et al. (2013) and they conclude that only four factors are effective at explaining the cross-section of option returns: Jump, volatility jump, volatility and liquidity. Recent papers such as Horenstein et al. (2020) employ a two-step factor model by first estimating the number of latent factors necessary to explain the cross-section of returns and then using principal component analysis (PCA) to estimate those factors. They find that four factors consisting of firm size, idiosyncratic volatility, the difference between implied and historical volatilities and market volatility risk explain the cross-section of equity option returns.

An improvement to the factor model of Horenstein et al. (2020) is made by Büchner and Kelly (2022) and Goyal and Saretto (2022), who predict option returns using instrumented principal component analysis (IPCA) which, unlike Horenstein et al. (2020), explicitly permits changes in individual asset behaviour over time. Using the factor model of Horenstein et al. (2020), Goyal and Saretto (2022) initially find that 14 portfolios constructed based on option characteristics yield significant alphas, challenging the efficiency of option markets. Their paper provides an improvement to the PCA factor model and reduces the number of significant alphas down to five. In general, the literature concerning option factor models implies that a handful of factors suffice to explain the cross-section of equity option returns (Carr and Wu, 2020; Christoffersen et al., 2018a).

Although popular in empirical asset pricing, linear factor models suffer from several shortcomings. First, factor models use portfolio sorts to approximate the conditional mean function of returns for only one characteristic. Portfolio sorts cannot be used to investigate a large number of characteristics. If we wish to sort assets into portfolios based on four different characteristics, we would already end up with $4^4 = 256$ different portfolios. Second, sorts don't allow us to easily infer information on which characteristics have incremental information on average returns (Fama and French, 2008). Third, when we sort portfolios into deciles, we must make the assumption that the expected returns in each decile are constant. Linear regressions are a possible solution to this. Although cross-sectional regressions allow us to analyze characteristics jointly, this approach has its downsides. Most importantly, we have no a priori reason to assume a linear structure in the conditional mean function of returns. In addition, linear regressions are notoriously susceptible to outliers, affecting their reliability in out-of-sample predictions.

The assumption of linearity is the Achilles' heel of literature regarding option return predictability. We contribute to the literature by introducing nonparametric estimation into the analysis of option returns. By estimating the model nonparametrically using the adaptive group least absolute shrinkage and selection operator (LASSO) of Huang et al. (2010) we allow for nonlinearities in the conditional mean function whilst also analyzing all characteristics jointly. In our empirical analysis we construct delta-hedged call option returns over a sample period January from 1996 to March 2021, as well as a multitude of option characteristics used in previous studies (Cao et al., 2021; Ou and Penman, 1989; Pontiff and Woodgate, 2008; Richardson et al., 2005; Freyberger et al., 2020; Lakonishok et al., 1994; Ang et al., 2006; Francis et al., 2004). We can group the option characteristics into four categories: (1) Past returns; (2) Contract characteristics such as bid-ask spread, open interest and volume; (3); Measures of volatility, including idiosyncratic volatility, systemic volatility and the difference between implied and historical volatility; (4) Firm characteristics such as asset growth, size and earnings per share. We include a total of 57 characteristics.

We first investigate anomalies in the call option market through the standard approach of sorting portfolios into deciles based on each characteristic and creating long-short portfolios by taking a long position in the options belonging to the highest decile and a short position in the options belonging to the lowest decile. After regressing the resulting portfolio returns on the Carhart four-factor model (Carhart, 1997), 34 out of 57 portfolios yield significant alphas. At first glance, these results imply that option markets suffer from severe inefficiencies. We also analyze all characteristics jointly using the Fama-Macbeth type regressions of Cao et al. (2021). We estimate the model on three subsamples. One where we consider all firms, another where we only consider large firms, and a final sample where we disregard the smallest decile of firms. In all three cases, the linear model leads us to believe that the number of characteristics that provide incremental information on option returns lies between 21 and 29.

Past literature teaches us that a handful of factors should suffice to explain the cross-section of option returns. So, in our next step, we estimate the adaptive group LASSO using the entire sample period from 1996 to 2021. Although widely accepted that option returns are nonlinear in nature (Jones, 2006; Constantinides et al., 2013), our results provide concrete evidence of nonlinearities in delta-hedged call option returns. When conditioning on all other characteristics, only 11 characteristics provide incremental information on option returns. Of these 11 characteristics, 8 are consistently selected when we vary our tuning parameters. Delta-hedged call option returns increase with dividend to price ratio, firm size and earnings per share and are negatively related with the number of new issues shared in the past year, maximum daily return, return on equity, share turnover and the difference between implied and historical volatility. Our main contention with current literature is that factor models' inability to incorporate nonlinearities makes them unfit to properly explain the cross-section of option returns and that by extension, the significance of most option return anomalies presented in Cao and Han (2013), Cao et al. (2021), Horenstein et al. (2020) and Büchner and Kelly (2022) are likely a result of the factor models themselves.

We examine if the model selects these characteristics consistently over time and whether the model suffers from time-varying parameters, since we implicitly assume the model's coefficients are constant over time. To answer the first question we estimate the model using a rolling window starting with data from 1996 to 2012. We slide the window forward by one year and repeat estimation until 2021. We find that dividend to price, the number of new shares issues in the past year, maximum daily return and the difference between realized volatility and implied volatility are the only characteristics that are consistently selected throughout the sample period from 2013 to 2021. We also investigate time-variation in the model's parameters by estimating the model once using data from 1996 to 2006, fixing the selected characteristics and rolling the estimation window forward by one month. We then estimate the coefficients of the fixed characteristics, roll the window forward by one month and repeat this process for the remainder of the sample. We find that the model suffers significantly from time-varying coefficients. For example, call option returns decrease with a firm's earnings to price ratio in 2010, but increase in 2016, only for the effect to become insignificant in the following years.

To alleviate concerns of in-sample overfitting, we evaluate the model's performance in out-of-sample forecasting. We estimate the nonparametric model and the linear model of Cao et al. (2021) using data from 1996 to 2018. We use the selected characteristics to predict the following month's option returns, creating a long-short portfolio by taking a long position in the ten percent of options with the highest predicted returns and a short position in the ten percent of options with the lowest predicted returns. We roll the estimation sample forward by one month and repeat until March 2021. We estimate both models separately for three different subsamples: all firms, large firms and all but the smallest firms. In all subsamples the adaptive group LASSO achieves higher Sharpe ratios and higher relative R^2's. We also compute the relative pricing error (as in Goyal and Saretto (2022) and Kelly et al. (2020)) and conclude that on average the pricing error of the nonparametric model is 8.6% lower than the linear model.

While the nonparametric model performs well with regard to out-of-sample pricing errors, this paper should not be interpreted as an attempt to deprecate existing factor models. Rather, we provide a glimpse into option analysis through a nonparametric lens. While our results show that current factor models' inability to incorporate nonlinearities is a problem, we acknowledge that the adaptive group LASSO is not a solution. We add to studies that attempt to measure the dimensionality of delta-hedged call option returns, but more specifically, we contribute to literature that attempts to incorporate nonlinearities in the analysis of option returns (Jones, 2006; Brooks et al., 2018) by identifying a subset of variables that explain the cross-section of option returns using a nonparametric model.

The remainder of this paper is outlined as follows. Section 2 summarizes our data-set and defines delta-hedged option returns. Section 3 outlines our methodological approach. Section 4 examines the cross-section of delta-hedged option returns in an empirical setting and examines time-variability in the model's estimated coefficients. Section 5 concludes. Lastly, section 5 reflects on the methods used and discusses some obstacles encountered in our empirical analysis.

2 Data

2.1 Option data

We use data from the equity options market from January 1996 to March 2021. Data on individual option prices are obtained from the IvyDB OptionMetrics database. We obtain stock returns, prices and credit ratings from the Center for Research on Security Prices (CRSP) and balance sheet data is acquired from Compustat. We consider options from all firms in the S&P 500.

We filter the options in the manner proposed by Horenstein et al. (2020). First, we exclude options with a trading volume of zero to avoid illiquid options. We also exclude options if the bid price is smaller than the ask price or if the midpoint of the bid and ask quote is less than 0.125 dollars. Third, we only consider options with moneyness between the range of $[0.8, 1.2]$. Next, we exclude options that do not adhere to the no-arbitrage condition: $\max(0, S_t - K, S_t - \sum_{t=1}^{\tau} e^{-rs} D_{t+s} - Ke^{-r\tau}) \leq C_t(\tau, K) \leq S_t$. Furthermore, options with an expiration date other than the third friday of the month are removed from the sample. We also exclude options which are not traded on the day the bid and ask quotes are reported. We filter options with a missing or an abnormal delta that is outside of the range [0,1] and lastly, we remove options from the Optionmetrics database with an implied volatility of -99.99. Optionmetrics sets implied volatility to -99.99 if the option is a special settlement, the options vega is less than half, the underlying stock price is not available on the CRSP database or because the numerical algorithm used to compute implied volatility fails to converge. We select options from the resulting list of filtered call options. Every month for each firm, we choose the option that is closest to being at the money with more than one month to maturity.

2.2 Delta-hedged option returns

Since options are derivatives of stocks, they are very susceptible to movements in the underlying asset. In this paper we analyze the returns of delta-hedged options, so that returns are independent of movements in the underlying asset. We use the approach of Cao and Han (2013) to obtain delta-hedged option returns which they define as the change in value of a self-financing portfolio consisting of a long position in a call option hedged by shorting delta shares in the underlying asset. Consider a call option that is hedged once at time t over a period $[t, t+\tau]$. The change in value of this portfolio is

$$\begin{aligned} \Pi(t, t+\tau) = C_{t+\tau} - C_t - \Delta_{C,t+\tau}[S(t+\tau) - S(t)] \\ - \frac{ar_{t+\tau}}{365}[C(t+\tau) - \Delta_{C,t} S(t+\tau)], \end{aligned} \tag{1}$$

where C_t and $\Delta_{C,t}$ are the price and delta of the call option at time t, $S(t)$ is the price of the underlying stock at time t, r_t is the annualized risk free rate at time t and a is the number of calendar days between t and $t+\tau$. Since $\Pi(t, t+\tau)$ is proportional to the initial stock price we divide (1) by the absolute value of the securities involved $(\Delta_t S_t - C_t)$. This way we can compare the option gain across stocks despite the significant difference in market prices. $\frac{\Pi(t,t+\tau)}{(\Delta_t S_t - C_t)}$) is what we refer to as delta-hedged option return.

While delta-hedging allows us to analyze option return predictability independent of movements in the underlying asset, the act of delta-hedging raises issues that make the interpretability of results difficult. A noteworthy issue is that we can compute the deltas in three different ways (Broadie et al., 2009). The first uses an option pricing model to compute the required hedging portfolio weights. Another method uses the shape of the implied volatility smile to compute hedge ratios (Bates, 2005). The third method, which is the approach we use in our analysis, is to compute deltas from the Black-Scholes model using implied volatility as a input parameter. While this is the most practical approach, one should keep in mind that the deltas obtained this way will not be the same as the ones obtained by the other methods. Formally, delta-hedging also requires rebalancing, increasing transaction costs and requires us to have more data. Therefore, we concern ourselves with the practical application of delta-hedging which involves holding the delta-hedged position for 1 month without rebalancing. The results of the buy-and-hold strategy are robust to daily rebalancing (Cao et al., 2021).

Table 1 shows summary statistics of our call option data. The data consists of 25 years of delta-hedged option returns for 489 different firms, although not all firms have call options reported in OptionMetrics throughout the entire sample period. As in Brooks et al. (2018), Cao et al. (2021) and Büchner and Kelly (2022), our delta-hedged returns are around -1% on average with an average delta of 0.51. Days to maturity ranges from 47 to 50.

Table 1
Summary of option data

	Mean	SD	Median
Delta-hedged returns (%)	-0.9075	3.2549	-1.2866
Days to maturity	48.8684	1.2856	49.0000
K/S	1.0023	0.0168	1.0009
Delta	0.5146	0.0742	0.5210

This table reports summary statistics of the option data from January 1996 to March 2021. Delta-hedged return is a self-financing portfolio going long in a call option and shorting that option's delta of the underlying stock, so as to create a portfolio that is insensitive to movements in the underlying stock. It is calculated as in Cao and Han (2013) and is defined as delta-hedged gain divided by $\delta_t S_t - C_t$ (see Equation (1)). We also report summary statistics for days to maturity, moneyness (K/S) and the options' delta.

2.3 Characteristics

There is a wide range of characteristics that might explain the cross-section of delta-hedged option returns. We will confine ourselves to characteristics commonly used in literature regarding the analysis of options (Horenstein et al., 2020; Cao et al., 2021; Brooks et al., 2018) as well as a variety of characteristics used to explain the cross-section of stock returns (Freyberger et al., 2020). The characteristics used in this study are shown in Table 2. A more elaborate explanation of each characteristic is given in the appendix. Table 3 shows descriptive statistics of the characteristics as well as their frequency. Note that all firm characteristics are reported yearly with the exception of past returns and trading frictions.

Table 2
List of firm characteristics by category

Past returns:	
r_{2-1}	Return from 2 to 1 months before prediction
r_{6-2}	Return from 6 to 2 months before prediction
r_{12-2}	Return from 12 to 2 months before prediction
r_{12-7}	Return from 12 to 7 months before prediction
r_{36-13}	Return from 36 to 13 months before prediction
Investment:	
ag	Asset growth
chinv	Change in inventory
egr	Growth in common shareholder equity
grltnoa	Growth in long-term net operating assets
invest	Property investment
issue_{1Y}	The 1 year change in shares outstanding
issue_{5Y}	The 5 year change in shares outstanding
pchsaleinv	% change in inventory
oa	Operating accruals
tef	Total external financing
Profitability:	
chtx	Change in tax expense
ep	Earnings to price
eps	Earnings per share
gma	Gross profitability
pchgm	% change in sales minus the % change in inventory
pchsale	% change in sales minus the % change in A/R
roavol	Standard deviation for 16 quarters of income
profit	Profitability
pm	Profit margin
roeq	Return on equity
s2c	Sales-to-cash
sat	Sales to total assets
sga2s	SG&A to sales
sgr	Annual percent change in sales
tb	Tax income to book income
turn	Share turnover

Intangibles:	
hire	Employee growth rate
rd_{mve}	R&D to market capitalization
rd_{sale}	R&D to sales
tan	Tangibility
Value-versus-growth:	
at	Total assets
bm	The natural logarithm of book to market equity
ch	Cash-to-assets ratio
currat	Current assets divided by current liabilities
leverage	Leverage
lgr	% change in total liabilities
s2p	Sales-to-price
sales_g	Sales growth
size	Logarithm of the market value of the firm's equity
pchcurrat	% change in current ratio
Trading Frictions:	
baspread	Bid-ask spread of the option
bidask	Ratio of the difference between the bid and ask quotes
dolvol	Natural log of trading volume times price per share
dy	Dividend to price
ivol	Stock return idiosyncratic volatility
interest	Open interest for the option contract
maxret	Maximum daily return
sratio	The ratio of systematic volatility over total volatility
std_{dolvol}	Standard deviation of daily dollar trading volume
tvol	Total volatility
volume	Option trading volume
impliedvol	Difference between implied volatility and realized volatility of the option

This table lists the characteristics considered in our option analysis sorted by category. The sample period is from January 1996 to March 2021.

Table 3
Summary Statistics of Characteristics

	Mean	Median	SD	Frequency
Past returns:				
r_{2-1}	0.02	0.01	(1.40)	m
r_{6-2}	0.11	0.04	(6.10)	m
r_{12-2}	0.18	0.09	(2.23)	m
r_{12-7}	0.09	0.05	(1.56)	m
r_{36-13}	0.37	0.14	(4.15)	m
Investment:				
ag	1.18	1.07	(0.97)	y
issue_{1Y}	21.79	0.18	(443.00)	y
issue_{5Y}	123.09	18.15	(833.71)	y
chinv	177.72	3.90	(4, 956.32)	y
grltnoa	1.14	1.07	(5.80)	y
tef	0.01	0.02	(0.32)	y
pchsaleinv	1.51	1.06	(16.84)	y
egr	701.93	214.30	(4, 311.79)	y
invest	0.07	0.04	(0.21)	y
oa	−74, 647.84	−7, 042.78	(318, 479.39)	y
Profitability:				
profit	1, 772.47	628.00	(4, 326.34)	y
chtx	24.37	7.20	(732.43)	y
pchgm	0.24	0.01	(50.16)	y
pchsale	0.47	−0.00	(42.81)	y
roeq	0.11	0.15	(2.79)	y
ep	89.09	30.75	(621.10)	y
eps	3.02	2.39	(5.30)	y
gma	7, 282.10	2, 793.10	(13, 884.72)	y
sgr	1.64	1.08	(42.81)	y
tb	0.25	0.30	(0.92)	y
turn	2, 251, 272.18	1, 759, 190.44	(855, 293.11)	y
pm	0.41	0.40	(0.58)	y
s2c	50.40	10.87	(678.23)	y
sat	0.83	0.65	(0.74)	y
sga2s	0.33	0.22	(5.20)	y
dy	0.02	0.01	(0.02)	y

	Mean	Median	SD	Frequency
Intangibles:				
tan	0.46	0.46	(0.21)	y
rd_{mve}	0.03	0.01	(0.35)	y
rd_{sale}	0.10	0.03	(0.64)	y
hire	1.08	1.03	(0.31)	y
Value-versus-growth:				
at	61, 253.58	13, 401.12	(207, 039.43)	y
size	9.60	9.56	(1.33)	y
ch	0.09	0.05	(0.10)	y
currat	1.86	1.46	(1.46)	y
bm	0.40	0.32	(1.61)	y
leverage	0.35	0.50	(14.30)	y
pchcurrat	−0.03	−0.01	(0.87)	y
lgr	1.24	1.06	(2.25)	y
s2p	514.78	162.44	(2, 751.52)	y
sales_g	1.64	1.08	(42.81)	y
Trading Frictions:				
ivol	0.017	0.015	(0.01)	m
bidask	−0.13	−0.08	(0.17)	m
baspread	0.88	0.07	(3.48)	m
dolvol	217.35	90.24	(676.97)	m
std_{dolvol}	323.30	191.50	(528.38)	m
maxret	0.00	0.00	(0.02)	m
roavol	195.65	20.14	(762.64)	m
sratio	0.54	0.56	(0.21)	m
tvol	0.02	0.02	(0.01)	m
interest	2, 991.51	562.62	(13, 823.40)	m
volume	144.44	15.72	(836.45)	m
impliedvol				m

This table reports the mean, median and standard deviation of the characteristics listed in table 2. Frequency refers to the frequency at which the characteristic changes. m is monthly and y is yearly. The sample period is from January 1996 to March 2021.

3 Methodology

3.1 The conditional mean function and nonparametric estimation

Our aim is to identify characteristics that can predict the delta-hedged option returns of firm i for $i \in \{1, \ldots, N\}$. This means we are modelling the conditional mean of the expected returns R_{it} as a function of S characteristics C_{is}. Thus, the conditional mean function of expected returns can be written as

$$m_t(x_1, \ldots, x_S) = E[R_{it}|X_{1,it-1} = x_1, \ldots, X_{S,it-1} = x_S]. \tag{2}$$

Typically, we sort assets into ten deciles for each firm characteristic and compare the mean returns across portfolios. While simplistic, portfolio sorts have their downsides. Problems with portfolio sorts include dimensionality, the assumption that returns do not vary within portfolios and that they do not let us assess predictability conditional on other characteristics.

Another approach is to assume a linear form for the conditional mean function. This implies modelling returns as

$$R_{it} = \alpha_i + \sum_{s=1}^{S} \beta_{S,i} X_{s,it-1} + \epsilon_{it}. \tag{3}$$

A benefit of linear panel regression is that it allows us to study characteristics jointly. However, we have no a priori reason to assume a linear form for the conditional mean function.

To alleviate these issues, we resort to nonparametric estimation. It is impractical to estimate the conditional mean function fully nonparametrically with this many regressors. The rate of convergence is slow and coefficients suffer from large estimation errors. As we study more characteristics, this problem becomes substantially worse. Since we want to study many characteristics jointly, a possible solution is to assume an additive conditional mean function:

$$m_t(x_1, \ldots, x_S) = \sum_{s=1}^{S} m_{ts}(x_s). \tag{4}$$

The main benefit of the additive model is that it greatly reduces the time required to reach convergence (Horowitz et al., 2006). Additive models do however require us to impose the restriction

$$\frac{\partial^2 m_t(c_1, \ldots, x_s)}{\partial x_s \partial x_{s'}} = 0 \tag{5}$$

for all $s \neq s'$. Thus, cross dependencies between characteristics are not allowed in the additive framework. Freyberger et al. (2020) try to circumvent this shortcoming by estimating the nonparametric model for small firms and then separately for large firms, but it remains time consuming to estimate the nonparametric model multiple times for different samples based on a given characteristic. We must acknowledge that additivity is a harsh assumption to make. Nevertheless, it is not as restrictive as Fama-Macbeth regressions, where we assume the conditional mean function is both additive and linear. Ultimately, the additive model provides us with econometric advantages that cannot be understated. Before we describe the nonparametric model in Section 3.3, we first provide a normalized rank transformation of the characteristics.

3.2 Normalization of characteristics

When analyzing asset returns in the cross-section, the absolute value of a firm's characteristic is usually not relevant. We care more about its rank relative to other firms in the cross-section. It is also computationally preferable to have our characteristics take values within a fixed range for reasons which will become clear in section 3.3. Because of this, we employ a normalized rank transformation to each characteristic to ensure that each of the S characteristics $X_{s,it}$ takes a value between 0 and 1. We use the transformation proposed by Freyberger et al. (2020).

The transformed characteristics are given by

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

Note that $\text{rank}(\min_{i=1,\dots,N_t} X_{s,it-1}) = 1$ and $\text{rank}(\max_{i=1,\dots,N_t} X_{s,it-1}) = N_t$ such that (6) always takes values between 0 and 1. Freyberger et al. (2020) show that in large samples, the rank transformed mean function is interchangeable with Equation (2). They also show that in smaller samples the transformed conditional mean function performs better in out-of-sample analysis in both numerical simulations as well as empirical data.

3.3 Adaptive group LASSO

We use the two-step group LASSO approach suggested by Huang et al. (2010). First we utilize the regular group LASSO procedure to reduce the model's dimensionality. The group LASSO still selects too many characteristics in the first step. In the second stage the adaptive group LASSO achieves consistent model selection. That is to say, it sets the functions $\tilde{m}_{ts}$ to 0 if that characteristic does not provide information on option returns.

Recall that we are modelling the returns R_{it} as a function of characteristics $X_{s,it}$. Suppose each characteristic $X_{s,it}$ takes values between $[a, b]$. Let $a = \xi_0 < \xi_1 < \cdots < \xi_K = b$ be a partition of $[a, b]$ into K sub intervals $I_k = [\xi_{K-1}, \xi_K)$ for $k = 0, 1, \dots, K$. In the adaptive group LASSO, we estimate functions over these K parts of the characteristic distribution. Following Freyberger et al. (2020), we will refer to the interpolation points $\xi_0, \ldots, \xi_{K-1}$ as knots and let $\xi_k = \frac{k}{K}$ for all $k = 1, \ldots, K-1$. Because the characteristics are rank normalized, the interpolation points correspond to quantiles of the characteristic distribution and in a sense we can interpret I_k as the k^{th} portfolio.

To re-iterate, we are modelling the returns

$$R_{it} = \sum_{s=1}^{S} \tilde{m}_{ts}(\tilde{X}_{s,it-1}) + \epsilon_{it}, \tag{7}$$

where $\tilde{m}_{ts}$ are mean functions for a single characteristic s and $\tilde{X}_{s,it-1}$ are the normalized rank transformed characteristics. We want to set the functions $\tilde{m}_{ts}$ to 0 for a given characteristic if that characteristic does not give us information on option returns. We've partitioned the characteristics into K subintervals so that we can estimate the functions $\tilde{m}_{ts}$ with a quadratic function for each interval. In particular, we use splines to approximate $\tilde{m}_{ts}$ (Huang et al., 2010). Using this approximation, the nonparametric components are defined as the sum of spline basis functions:

$$\tilde{m}_{ts}(\tilde{x}) \approx \sum_{k=1}^{K} \beta_{tsk} p_k(\tilde{x}), \tag{8}$$

where $p_k(\tilde{x})$ are basis functions and β_{tsk} are the parameters we wish to estimate. We can choose the amount of subintervals K ourselves. The larger K, the more precise the approximation is but it also increases the number of parameters and the variance in our estimates.

The $K \times S$ vector of coefficient estimates $\boldsymbol{\beta}_t$ is given by:

$$\tilde{\boldsymbol{\beta}}_t = \underset{\beta_{sk}:s=1,\ldots,S;k=1,\ldots,K}{\arg\min} \sum_{i=1}^{N} \left(R_{it} - \sum_{s=1}^{S} \sum_{k=1}^{K} (\beta_{sk} p_k(\tilde{X}_{s,it-1}) \right)^2 + \lambda_1 \sum_{s=1}^{S} \left(\sum_{k=1}^{K} \beta_{sk}^2 \right)^{\frac{1}{2}}, \tag{9}$$

with λ_1 being a penalty parameter. Intuitively the idea of the group LASSO procedure is simple. The first sum in Equation (9) is simply the sum of squared residuals. The second part penalizes coefficients based on their magnitude. With all characteristics being normalized between 0 and 1, the second part of (9) penalizes coefficients that do not provide incremental information on the conditional mean function. Note that unlike regular LASSO, group LASSO does not penalize the coefficients β_{sk} individually, but instead penalizes all coefficients related to a given characteristic.

While we achieve dimension reduction in the first step it still selects too many characteristics, including ones with no predictive power. The second step ensures that we only select characteristics that provide incremental information on returns. We first obtain the following weights by setting

$$w_{ts} = \begin{cases} \left(\sum_{k=1}^{K} \beta_{sk}^2\right)^{-\frac{1}{2}} & \text{if } \left(\sum_{k=1}^{K} \beta_{sk}^2\right)^{-\frac{1}{2}} > 0 \\ \infty & \text{if } \left(\sum_{k=1}^{K} \beta_{sk}^2\right)^{-\frac{1}{2}} = 0. \end{cases} \tag{10}$$

Here we define that $0 \cdot \infty = 0$, such that characteristics not selected in the first step are also not selected in the second step. The model is then solved for

$$\hat{\boldsymbol{\beta}}_t = \underset{\beta_{sk}:s=1,\ldots,S;k=1,\ldots,K}{\arg\min} \sum_{i=1}^{N} \left(R_{it} - \sum_{s=1}^{S}\sum_{k=1}^{K} \beta_{sk} p_k(\tilde{X}_{s,it-1})\right)^2 + \lambda_2 \sum_{s=1}^{S} \left(w_{ts} \sum_{k=1}^{K} \beta_{sk}^2\right)^{\frac{1}{2}}. \tag{11}$$

The choice of penalty parameters λ_1 and λ_2 in (9) and (11) determines how much we value a better fit at the cost of additional parameters. A larger penalty parameter means we are less likely to include characteristics into our model. In our empirical application we use the Bayesian information criterion for both λ_1 and λ_2 which is the number of model parameters times $\text{Log}(NT)$. Our choice of basis function is a second order spline unless specified otherwise, that means $p_1(\tilde{X}_{s,it-1}) = 1$, $p_2(\tilde{X}_{s,it-1}) = \tilde{X}_{s,it-1}$, $p_3(\tilde{X}_{s,it-1}) = \tilde{X}_{s,it-1}^2$ and $p_k(\tilde{X}_{s,it-1}) = \max(\tilde{X}_{s,it-1} - \xi_{k-3}, 0)^2$ for any $k > 3$. Lastly, while the betas in Equations (9) and (11) are time-varying, we will work under the assumption that the conditional mean function is time-invariant since this allows us to pool the observations across time. This leads to more precise parameter estimates and more reliable modelling of the conditional mean function (Gu et al., 2020). In the following section we discuss a way to create confidence bands for the conditional mean function to give us a pragmatic feeling for estimation uncertainty.

3.4 Confidence bands

Although we are usually interested in estimation uncertainty surrounding coefficient estimates, it is useful to have an idea of the estimation uncertainty surrounding the conditional mean function itself. It is important to emphasize that the following confidence bands do not affect which characteristics the adaptive group LASSO selects. They merely provide us with a feeling for the uncertainty in the estimated conditional mean function.

Remember that we assume the model

$$\tilde{m}_t(\tilde{x}_1, \ldots, \tilde{x}_S) = \sum_{s=1}^{S} \tilde{m}_{ts}(\tilde{x}_s). \tag{12}$$

However, due to the assumption of additivity the levels of the functions $\tilde{m}_{ts}(\tilde{x}_s)$ are unidentified. This is because

$$\tilde{m}_t(\tilde{x}_1, \ldots, \tilde{x}_s) = \sum_{s=1}^{S} \tilde{m}_{ts}(\tilde{x}_s) = \sum_{s=1}^{S} \Big(\tilde{m}_{ts}(\tilde{x}_s) + \alpha_s \Big) \tag{13}$$

holds for any combination of α_s that satisfies $\sum_{s=1}^{S} \alpha_s = 0$. Thus, the functions $\tilde{m}_{ts}$ are only identified up to a constant. Note that these constants are only relevant when we try to plot the conditional mean function for a given characteristic. Freyberger et al. (2020) propose the use of estimates and confidence bands for the functions $\tilde{m}_{ts} - \int \tilde{m}_{ts}(\tilde{x}_s)d\tilde{x}_s$. This means we normalize the functions $\tilde{m}_{ts}(\tilde{x}_s)$ such that they integrate to zero. These confidence bands would also allow us to test for hypotheses that are independent of the levels of the functions $\tilde{m}_{ts}$.

We briefly present uniform confidence bands as proposed in Freyberger et al. (2020). Recall that we estimate $\tilde{m}_{ts}$ using a sum of basis functions $\sum_{s=1}^{S} \hat{\beta}_{tsk} p_k(\tilde{x}_s)$. Let $\tilde{p}_k(\tilde{x}_s) = p_k(\tilde{x}_s) - \int p_k(\tilde{x}_s)d\tilde{x}_s$ and denote $\tilde{p}(\tilde{x}_s)$ as $(\tilde{p}_1(\tilde{x}_s), \ldots, \tilde{p}_K(\tilde{x}_s))'$. Define the $K \times K$ covariance matrix of $\sqrt{n}(\hat{\beta}_t s - \beta_{ts})$ as Σ_{ts}. Let $\hat{\Sigma}_{ts}$ be the heteroscedasticity consistent estimator of Σ_{ts} and define the standard error of $\sum_{k=1}^{K} \hat{\beta}_{tsk}\tilde{p}_k(\tilde{x}_s) = \sqrt{\tilde{p}(\tilde{x}_s)'\hat{\Sigma}_{ts}\tilde{p}(\tilde{x}_s)} = \sigma_{ts}(\tilde{x}_s)$. We can then write the uniform confidence band for the function $\tilde{m}_{ts}(\tilde{x}_s) - \int \tilde{m}_{ts}(\tilde{x}_s)d\tilde{x}_s$ as

$$\left[\sum_{k=1}^{K} \hat{\beta}_{tsk}\tilde{p}_k(\tilde{x}_s) - d_{ts}\sigma_{ts}(\tilde{x}_s), \sum_{k=1}^{K} \hat{\beta}_{tsk}\tilde{p}_k(\tilde{x}_s) + d_{ts}\sigma_{ts}(\tilde{x}_s) \right], \tag{14}$$

where d_{ts} is a constant. d_{ts} is chosen such that

$$P\left(\sup_{\tilde{x}_s \in [0,1]} \left| \frac{Z'\tilde{p}(\tilde{x}_s)}{\hat{\sigma}_{ts}(\tilde{x}_s)} \right| \le \hat{d}_{ts} | \hat{\Sigma}_{ts} \right) = 1 - \alpha. \tag{15}$$

The left hand side of (15) is computed using simulations. Define L_{ts} as $\left[\sum_{k=1}^{K} \hat{\beta}_{tsk}\tilde{p}_k(\tilde{x}_s) - d_{ts}\sigma_{ts}(\tilde{x}_s), \sum_{k=1}^{K} \hat{\beta}_{tsk}\tilde{p}_k(\tilde{x}_s) + d_{ts}\sigma_{ts}(\tilde{x}_s) \right]$. Assuming that the conditions in Belloni et al. (2015) are met, we get that

$$P\left(\tilde{m}_{ts}(\tilde{c}_s) - \int \tilde{m}_{ts}(\tilde{c}_s)d\tilde{c}_s \in L_{ts} \; \forall \; \tilde{c}_s \in [0,1] \right) \tag{16}$$

approaches $1 - \alpha$ as the size of the sample increases.

4 Empirical application

4.1 Portfolio sorts

We devote to section to assessing the predictive power of each characteristic listed in Table 2 using portfolio sorts. Every month we sort firms into deciles for each characteristic and create equally weighted portfolios by going long in the option returns corresponding to firms belonging to the highest decile and taking a short position in those in the lowest decile. We then regress these monthly portfolio returns on the Carhart four-factor model (Carhart, 1997) and report the resulting alphas and their corresponding t-statistics in Table 4. In total 34 out of 57 characteristics have significant alphas at a 95% confidence level. This first pass implies that option markets exhibit massive inefficiencies.

4.1.1 Portfolios based on option characteristics

Let us first take a closer look at the portfolios sorted on option characteristics. These are sorted on the ratio of the difference between the bid and ask quotes (bidask), bid-ask spread of the option (baspread), open interest of the option contract (interest) and the option trading volume (volume). Table 4 shows that there exists a premium on options with a higher absolute spread in the bid and ask quotes (baspread), a lower open interest and on options with a higher trading volume but not on options with a higher ratio between the bid and ask quotes (bidask).

These findings are not entirely in line with those in Brooks et al. (2018) since they conclude that portfolios based on the ratio between bid and ask quotes have a significant negative premium, whereas we find no significant relationship at all. The premium on illiquidity (lower open interest) is extensively documented in previous literature. Christoffersen et al. (2018b) find that the daily risk-adjusted return spread for illiquid options over liquid options is 3.4 percent. They argue that market makers hold large and risky net-long positions and that the illiquidity premium compensates for this risk. At first glance, our results seem to agree that this illiquidity premium exists.

4.1.2 Portfolios based on volatility

Volatility is arguably the most important variable in pricing options and there are numerous papers dedicated to the relation between volatility and option returns (Cao and Han, 2013; Bakshi and Kapadia, 2003; Goyal and Saretto, 2009a,b). Our analysis focuses on the effects of idiosyncratic, systemic and total volatility, as well as the difference between implied and realized volatility. Cao and Han (2013) conclude that average call option return is negative and that it decreases monotonically as idiosyncratic volatility increases. Their call options portfolios of firms belonging to the highest decile of idiosyncratic volatility earn, on average, one and a half percent less per month than portfolios of stocks in the lowest decile. Additionally, they find that the reverse is true for systemic risk. Delta-hedged option return for stocks increases monotonically as systematic risk increases. Cao et al. (2021) report the same inverse relationship between idiosyncratic volatility and option returns. Brooks et al. (2018) find a negative relationship for both idiosyncratic volatility and systemic risk. Additionally, their call option returns decrease monotonically with the difference between implied and realized volatility. A similar option pricing anomaly is presented in Goyal and Saretto (2009a). They hypothesize that large deviations between implied volatility and realized volatility can be used to predict option returns. They argue that implied volatility and historical volatility are mean-reverting by nature and thus excessive implied volatility shows an overpricing of the option which the market eventually corrects, leading to lower option returns in the near future. Returning to the effects of systemic risk, Duan and Wei (2009) show that after controlling for total volatility, higher systemic risk leads to an increase in implied volatility. If the mean-reverting nature of posited by Goyal and Saretto (2009a) holds, the results of Duan and Wei (2009) imply a negative relationship between systemic risk and option returns.

Moving on to our findings, we observe a significant negative relation between delta-hedged option return and the idiosyncratic volatility (ivol) of the underlying stock, as well as a negative alpha of -0.16 for total volatility of the underlying stock (tvol) indicating a negative premium on options on stocks with higher volatility in the past 16 months. On the other hand, we find a positive relation for systemic risk (sratio). All of these findings align with Cao and Han (2013) and Cao et al. (2021). We do not observe the negative premium for systemic risk documented in Brooks et al. (2018). When it comes to the difference between realized volatility and implied volatility, we add to the findings of Goyal and Saretto (2009a) and observe that an overestimation of implied volatility leads to lower option returns in the near future. The alpha for impliedvol is -1.36 and of the 57 characteristics we assess, the risk premium of implied volatility has the largest t-statistic at -10.67.

4.1.3 Portfolios based on past returns

Momentum is a well known anomaly in stock returns and was incorporated into the Fama French 3 factor model by Carhart (1997). A wide body of research documents that stocks which performed well in the recent past have a tendency to do so in the future and that stocks exhibit a long term (36+ months) reversal. The results of Brooks et al. (2018) imply that a momentum anomaly is also present in option returns, as they find positive returns for both 12 month and 36 month momentum portfolios. A similar conclusion is drawn in Heston et al. (2022), stating that options with high historical returns in the past 6 to 36 months continue to exhibit those returns in the near future. They find that unlike stocks, option returns are not followed by long-term reversal nor do they exhibit short-term reversal. This momentum effect in option returns is subject to dispute as Jones et al. (2020), who analyse momentum, reversal and seasonality, find that options do exhibit a significant short-term reversal. While the effect is robust over time, they argue it could be explained by other option factors with emphasis on the difference between implied and realized volatility. They note that a higher option return this month increases the implied volatility of the option relative to its historical volatility. Due to their mean-reverting nature implied volatility and thus the option's price decreases in the following month.

Although the findings of Jones et al. (2020) are robust over time and significant, we do not observe any indication of short-term reversal. While there seems to be a correlation between delta-hedged option returns and the recent momentum of the underlying stock, this is only apparent in momentum factors from 2 to 12 months before return prediction. For r_{6-2} and r_{12-2} we find significantly alphas. However, unlike Brooks et al. (2018) and Heston et al. (2022), we observe that these excess portfolio returns are in fact negative. Strong performance of the underlying asset in the past year seems to result in lower option returns. The t-statistic of 0.20 for the alpha of r_{36-12} also implies that option returns do not depend on past performance of the underlying stock from over a year ago. This disputes the existence of long-term momentum observed in Brooks et al. (2018) and Heston et al. (2022).

4.1.4 Portfolios based on firm characteristics

Finally, we explore option return predictability with regard to firm characteristics commonly used in analysis of stock returns. Both Cao et al. (2021) and Brooks et al. (2018) include firm fundamentals in their analysis of option return predictability. Cao et al. (2021) find that delta-hedged option returns decrease with cash holdings and new share issues in the past year and increase with size and profitability. Brooks et al. (2018) include 93 characteristics. Of these characteristics, 27 corresponds with firm characteristics included in our analysis.

Table 4 shows that 29 portfolios sorted on firm fundamentals yield significant alphas. As in Cao et al. (2021), our findings support the notion that delta-hedged option returns decrease with cash holdings (ch) and yearly percentage change in outstanding shares (Issue_{1Y}) and increase with size and profitability (profit). In addition, Table 4 implies that option returns also increase with earnings to price (ep), earnings per share (eps), gross profitability (gma), return on equity (roeq), sales to cash (s2c), dividend to price (dy), total assets (at) and leverage. Option returns decrease with increases in asset growth (ag), growth in long term not operating assets (grltnoa), capital expenditures and inventory (invest), percentage change in sales (pchsale), share turnover (turn), percentage growth rate in annual sales (sales_g), the annual percentage change in sales (sgr), the ratio of current assets over liabilities (currat) and with the annual percentage change in total liabilities (pchcurrat). The majority of these results correspond with Brooks et al. (2018). Aside from minor differences in significance, the only noteworthy dissimilarity is the negative risk premium that we observe for capital expenditures and inventory. This risk premium is significantly positive in their results.

Table 4
Carhart 4-factor alphas

	H-L	Alpha	t-statistic		H-L	Alpha	t-statistic
Past returns:				**Intangibles:**			
r_{2-1}	−0.00	−0.05	−0.49	tan	−0.10	−0.13	−1.64*
r_{6-2}	−0.30	−0.31	−2.52**	rd_{mve}	−0.17	−0.20	−2.04**
r_{12-2}	−0.27	−0.30	−2.35**	rd_{sale}	−0.18	−0.24	−2.32**
r_{12-7}	−0.13	−0.14	−1.17	hire	−0.32	−0.31	−3.59***
r_{36-12}	0.01	0.02	0.20				
				Value-versus-growth:			
Investment:				at	0.82	0.82	8.62***
ag	−0.21	−0.23	−2.64***	size	1.02	0.95	10.46***
issue_{1Y}	−0.31	−0.34	−4.75***	ch	−0.43	−0.42	−5.24***
issue_{5Y}	−0.04	−0.05	−0.82	currat	−0.49	−0.48	−5.38***
chinv	−0.02	−0.05	−0.73	bm	0.03	0.03	0.31
grltnoa	−0.17	−0.19	−2.16**	leverage	0.25	0.23	2.91**
tef	−0.51	−0.50	−6.30***	pchcurrat	0.10	0.07	0.83
pchsaleinv	−0.19	−0.21	−2.70	lgr	−0.13	−0.14	−1.65*
egr	−0.01	−0.00	−0.03	s2p	0.73	0.69	8.06***
invest	−0.28	−0.29	−3.95***	sales_g	−0.37	−0.39	−4.36***
oa	−0.59	−0.60	−6.32***	dy	0.71	0.73	8.87***
Profitability:				**Trading Frictions:**			
profit	0.89	0.84	9.97***	bidask	0.04	0.00	0.01
chtx	0.03	0.02	0.25	baspread	0.23	0.14	3.03***
pchgm	−0.18	−0.10	−1.18	dolvol	0.28	0.26	3.55***
pchsale	−0.15	−0.05	−0.65	std_{dolvol}	0.16	0.10	1.34
roeq	0.29	0.30	4.34***	maxret	−0.02	−0.09	−1.20
ep	0.84	0.80	8.44***	roavol	0.70	0.64	7.34***
eps	0.47	0.44	5.67***	sratio	0.15	0.11	1.69*
gma	0.89	0.84	9.30***	tvol	−0.16	−0.16	−2.02**
sgr	−0.38	−0.39	−4.36***	interest	−0.30	−0.34	2.98**
tb	0.13	0.11	1.31	volume	0.24	0.21	1.94*
turn	−0.81	−0.80	−8.90***	impliedvol	−1.40	−1.36	−10.67***
pm	0.12	0.13	1.57	ivol	−0.22	−0.22	−3.03***
s2c	0.18	0.12	1.60				
sat	−0.21	−0.28	−3.31***				
sga2s	−0.15	−0.15	−1.55				

This table reports mean returns of long-short portfolios sorted on the characteristics described in Table 2 going long in the highest decile of firms and taking a short position in the lowest decile. Monthly long-short portfolio returns are regressed on the Carhart 4-factor model. We report the resulting alphas with their corresponding t-statistic. T-statistics at a significance level of 90%, 95% and 99% are indicated with one, two and three asterisks respectively. The sample period is from January 1996 to March 2021.

4.2 Fama-Macbeth regressions

The univariate portfolio sorts in Table 4 fail to take correlations between firm characteristics into account. In this section we analyze the predictability of all characteristics jointly by performing Fama-Macbeth regressions. The resulting slope coefficients for each characteristic and their corresponding t-statistic are presented in Table 5. Since size is a well-known predictor of asset returns, the regression is performed separately for all firms, large firms and for all but the smallest firms. The univariate portfolio sorts would lead us to believe that up to 34 characteristics can be used to predict future option returns. This is close to the 23 characteristics with significant slope coefficients in the first two columns of Table 5. The most noteworthy characteristics are asset growth, change in inventory, share turnover, sales to cash, total assets, size, dividend to price, volume, idiosyncratic volatility, total volatility and the difference between implied and realized volatility.

Again the relation between delta-hedged returns and idiosyncratic volatility is negative, but now that we consider all characteristics jointly the relation is positive for total volatility. The slope coefficient for the difference between implied and realized volatility is -0.350 with the largest t-statistic of -88.01. We also find positive slope coefficients for total assets and size. Larger firms have higher delta-hedged option returns on average. For smaller firms, which tend to have a higher percentage growth in total assets (ag), the slope coefficient is negative at a 99% significance level. When analyzing all characteristics jointly, we do find a short term reversal effect (r_{2-1}) which is more in line with Jones et al. (2020). However, we no longer observe any momentum effect past 6 months.

The third column of Table 5 shows the slope coefficients when only large firms are taken into consideration. For every option month observation, we select only those firms whose size is above the mean. In this case, only 21 characteristics provide incremental information on option returns. The slope coefficients with t-statistics larger than 15 in the first column are still significant at a 99% confidence level and their signs do not change. When we only consider large firms asset growth no longer yields incremental information on option returns. Since asset growth tends to vary more among smaller firms this indicates it is perhaps not asset growth itself that is a relevant characteristic, so much as its correlation with size. There are only thirteen characteristics that selected in each of the three samples based on firm size: chinv, invest, pchsale, gma, at, size, dy, ep, eps, s2p, volume, tvol and impliedvol. This is notably less than the 23 characteristics we select when we use all firms in the sample. Still, we know the linear model has significant drawbacks. Most importantly, there is no reason to assume that the conditional mean function is linear. In the next section we circumvent this by estimating the model nonparametrically using the adaptive group LASSO.

Table 5
Slope Coefficients of Fama-Macbeth Regressions

	All firms		Large firms		No tiny firms	
	Slope	t-statistic	Slope	t-statistic	Slope	t-statistic
ag	−0.140***	−14.097	−0.031	−0.983	−0.142***	−15.338
issue_{1Y}	−0.042	−1.236	−0.025	−0.624	−0.026	−0.499
issue_{5Y}	0.017	0.208	−0.006	−0.038	0.004	0.011
chinv	0.121***	10.531	0.050**	2.559	0.138***	14.475
grltnoa	0.045	1.468	0.041*	1.706	0.101***	7.735
tef	−0.031	−0.700	−0.014	−0.185	−0.052**	−2.031
pchsaleinv	−0.048*	−1.652	−0.005	−0.025	−0.049*	−1.839
egr	−0.021	0.327	−0.014	−0.188	−0.029	−0.628
invest	−0.082***	−4.810	−0.074***	−5.556	−0.095***	−6.890
profit	0.049*	1.705	0.025	0.642	−0.029	0.650
chtx	−0.019	0.246	0.014	0.189	0.016	0.184
pchgm	−0.053*	1.972	0.026	0.710	0.046*	1.621
pchsale	−0.101***	−7.245	−0.048**	−2.352	−0.093***	−6.551
gma	−0.070***	−3.461	−0.242***	−59.601	−0.177***	−23.877
sgr	−0.017	−0.215	−0.005	−0.023	0.009	0.063
tb	−0.011	0.086	0.026	0.702	0.012	0.101
turn	−0.097***	−6.759	−0.024	−0.587	−0.065***	−3.213
pm	0.037	0.962	−0.061***	−3.812	−0.005	−0.016
s2c	0.119***	10.100	0.011	0.120	0.098***	7.317
sat	0.023	0.366	0.102***	10.635	0.045	1.517
sga2s	−0.005	−0.014	0.158***	25.612	0.074***	4.192
ch	0.032	0.715	−0.024	−0.577	0.033	0.827

Table 5
Slope Coefficients of Fama-Macbeth Regressions (Continued)

	All firms		Large firms		No tiny firms	
	Slope	t-statistic	Slope	t-statistic	Slope	t-statistic
at	0.172***	21.298	0.237***	57.389	0.189***	27.089
size	0.194***	26.881	0.101***	10.365	0.227***	39.185
currat	−0.012	−0.097	−0.023	−0.545	−0.024	−0.422
pchcurrat	0.027	0.520	0.016	0.255	0.009	0.065
bm	−0.046	−1.504	−0.013	−0.171	−0.011	−0.084
oa	−0.002	−0.003	0.017	0.293	−0.009	−0.059
leverage	−0.024	−0.400	−0.021	−0.466	−0.024	−0.426
dy	0.286***	58.508	0.109***	12.006	0.232***	40.946
rd_{mve}	0.038	1.030	−0.064***	−4.120	−0.034	−0.898
rd_{sale}	−0.036	−0.920	0.021	0.433	0.004	0.014
roeq	−0.067***	−3.261	−0.034	−1.155	−0.062***	−2.883
ep	−0.055**	−2.140	0.166***	28.156	0.114***	9.958
eps	0.078***	4.378	0.075***	5.709	0.126***	12.124
hire	0.029	0.609	0.002	0.005	0.024	0.430
lgr	0.076***	4.097	0.026	0.672	0.059***	2.626
salesg	0.017	0.215	−0.005	−0.026	0.009	0.063
tan	0.073***	3.862	0.015	0.233	0.025	0.491
s2p	−0.052*	−1.961	−0.166***	−28.011	−0.133***	−13.523
volume	0.244***	42.539	0.058***	3.431	0.189***	27.167
bidask	0.040	1.119	0.094***	9.027	0.072***	3.907
interest	0.013	0.121	−0.21	−0.427	−0.003	−0.005
r_{2-1}	−0.049*	−1.713	0.002	0.003	−0.036	−0.999
r_{6-2}	−0.060**	−2.558	−0.029	−0.837	−0.072***	−3.907
r_{12-2}	0.003	−0.005	−0.041*	−1.720	0.009	0.057
r_{12-7}	−0.035	−0.895	0.019	0.372	0.002	−0.002
r_{36-12}	0.032	0.721	0.003	0.007	0.011	0.089
baspread	−0.078***	−4.385	−0.024	−0.604	−0.054**	−2.228
dolvol	0.062***	2.746	0.021	0.445	0.035	0.907
ivol	−0.096***	−6.618	−0.077***	−5.979	−0.040	−1.200
tvol	0.104***	7.762	0.095***	9.125	0.056**	2.338
sratio	−0.008	−0.041	0.005	0.028	0.012	0.105
std_{dolvol}	−0.006	−0.041	−0.036	−1.307	−0.022	−0.374
maxret	−0.053**	−2.000	0.029	−0.856	−0.043	−1.403
roavol	−0.050*	−1.773	0.004	0.017	−0.033	−0.828
impliedvol	−0.350***	−88.010	−0.165***	−27.809	−0.286***	−62.247

This table presents Fama-Macbeth slope coefficients of the characteristics in Table 2 with their corresponding t-statistic. 90%, 95% and 99% significance are denoted by one, two and three asterisks respectively. Large firms refers to firms with a market value above average. Tiny firms are defined as the smallest 10% of firms based on market value. The sample period is from January 1996 to March 2021.

4.3 Selected characteristics in the adaptive group LASSO

Now we examine characteristic selection of the adaptive group LASSO for a variety of sample periods and firm sizes. A comparison between the out of sample performance of the adaptive group LASSO and the linear model is done in Section 4.5. The purpose of this section is to provide some insights into the characteristics that the nonparametric model selects and the relation between those characteristics and delta-hedged call option returns.

Table 6
Selected Firm Characteristics in Nonparametric Model

Firms		All	All	All	All	All	Large firms	No tiny firms
Sample		Full	Full	Full	2015-2021	1990-2015	Full	Full
Knots		15	10	5	10	10	10	10
Sample Size		65618	65618	65618	27181	38437	32985	58942
# Selected		11	11	11	11	13	6	10
Sharpe Ratio		0.0992	0.0727	0.0112	-0.4311	0.3631	0.4019	0.2013
Characteristics	# Selected	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ag	1						ag	
baspread	3	baspread	baspread			baspread		
dy	7	dy	dy	dy	dy	dy	dy	dy
ep	2					ep		ep
eps	4	eps	eps	eps		eps		
interest	1				interest			
issue_{1Y}	5	issue_{1Y}	issue_{1Y}	issue_{1Y}		issue_{1Y}		issue_{1Y}
maxret	5	maxret	maxret	maxret		maxret		maxret
pchgm	1				pchgm			
pchsale	3		pchsale		pchsale			pchsale
profit	3	profit	profit		profit			
r_{6-2}	1					r_{6-2}		
r_{12-2}	2	r_{12-2}					r_{12-2}	
r_{12-7}	3			r_{12-7}	r_{12-7}			r_{12-7}
rdsale	1				rdsale			
roeq	4	roeq	roeq	roeq		roeq		
s2c	3			s2c		s2c		s2c
sga2s	2				sga2s	sga2s		
size	6	size	size	size		size	size	size
sratio	1				sratio			
std_{dolvol}	1			std_{dolvol}				
turn	7	turn	turn	turn	turn	turn	turn	turn
impliedvol	7	impliedvol	impliedvol	impliedvol	impliedvol	impliedvol	impliedvol	impliedvol

Never selected: at, bidask, bm, ch, chinv, chtx, currat, dolvol, gma, grltnoa, hire, invest, ivol, issue_{5Y}, leverage, lgr, oa, pchcurrat, pchsaleinv, pm, r_{2-1}, r_{36-12}, rd_{mve}, roavol, s2p, sales_g, sat, sgr, tan, tb, tef, tvol and volume. This table reports selected characteristics in the nonparametric model (11) as well as in-sample Sharpe ratio for different amounts of interpolation points (knots), sample periods and different selections of firms. Large firms refers to firms with an above average market value, whereas tiny firms are the smallest 10% of firms based on market value. The sample period is from January 1996 to March 2021.

We estimate the model in Equation (11) with 15 knots using the entire sample period. The selected characteristics and in-sample Sharpe ratio are shown in Table 6. As expected the group LASSO selects less characteristics than the linear model. Model (1) shows that only 11 characteristics provide incremental information on option returns, as opposed to the 23 characteristics selected by the Fama-Macbeth regression. With the exception of issue_{1Y} and r_{12-2}, each of the 11 selected characteristics is also significant at a 99% confidence level in the linear model.

The characteristics the adaptive group LASSO selects varies slightly depending on the number of interpolation points. With 15 interpolation points the group LASSO selects the bid-ask spread of the option, dividend to price ratio, earnings per share, yearly percentage change in outstanding shares, maximum daily return, profitability, return from 12 to 2 months before prediction, return on equity, size, share turnover and the deviation of the option's implied volatility from its historical volatility. Using 10 knots, return from 12 to 2 months before prediction is replaced by the percentage change in sales minus the percentage change in A/R. While there are minor differences in models (1), (2) and (3), it is reasonable to assume that the number of interpolation points is not extremely important.

To illustrate why the nonparametric model reduces the number of selected characteristics, the unconditional mean functions of asset growth, total assets and tangibility are plotted alongside their conditional mean functions in Figure 1. We initially found that each of these characteristics yield significant return premiums in univariate portfolio sorts. Even after considering all characteristics jointly in the linear model, we found that asset growth, total assets and tangibility had significant slope coefficients. On the left panels of Figure 1 we see the unconditional mean functions of asset growth, total assets and tangibility. Firms with low asset growth and tangibility and high total assets have higher expected returns. This is in line with the results of the univariate portfolio sorts, where we found a equally weighted monthly portfolio return of -0.21 for asset growth, -0.10 for tangibility and a significant positive return of 0.82 for total assets. The Fama-Macbeth regression in Table 5 also shows a negative premium of -0.140 for asset growth and a positive premium of 0.172 for total assets, although the option premium for tangibility is positive at 0.073 when all characteristics are considered in a linear regression. When we condition on all other characteristics in the nonparametric model we obtain the conditional mean function plotted on the right hand side of the panel. We now observe that firms with low asset growth do not yield higher option returns than firms with high asset growth. Total assets appears to have no influence on option returns at all and if tangibility were to be a selected characteristic, its relation with delta-hedged option returns is seemingly opposite when we condition on other characteristics. Firms with high tangibility actually have slightly higher returns than firms with low tangibility.

To illustrate the effects selected characteristics have when we allow for nonlinear relationships we plot the mean functions of dividend to price (dy), size and earnings per share (eps) in Figure 2. Although the conditional plots on the right hand side are not identical to the unconditional mean function on the left hand side. Their relation with respect to option returns does not change drastically. Average returns still increase when dividend to price increases, regardless of whether we condition on other characteristics or not. We observe a similar effect for size. It appears that not every characteristic has a significant nonlinear interaction with delta-hedged returns as both of their conditional mean functions could be approximated by a somewhat linear function. However, the conditional mean function of earnings per share shows the need for nonlinearities in the model. Expected returns increase steadily with earnings to price up to 0.75, after which subsequent increases in earnings to price actually lead to a decrease in average returns.

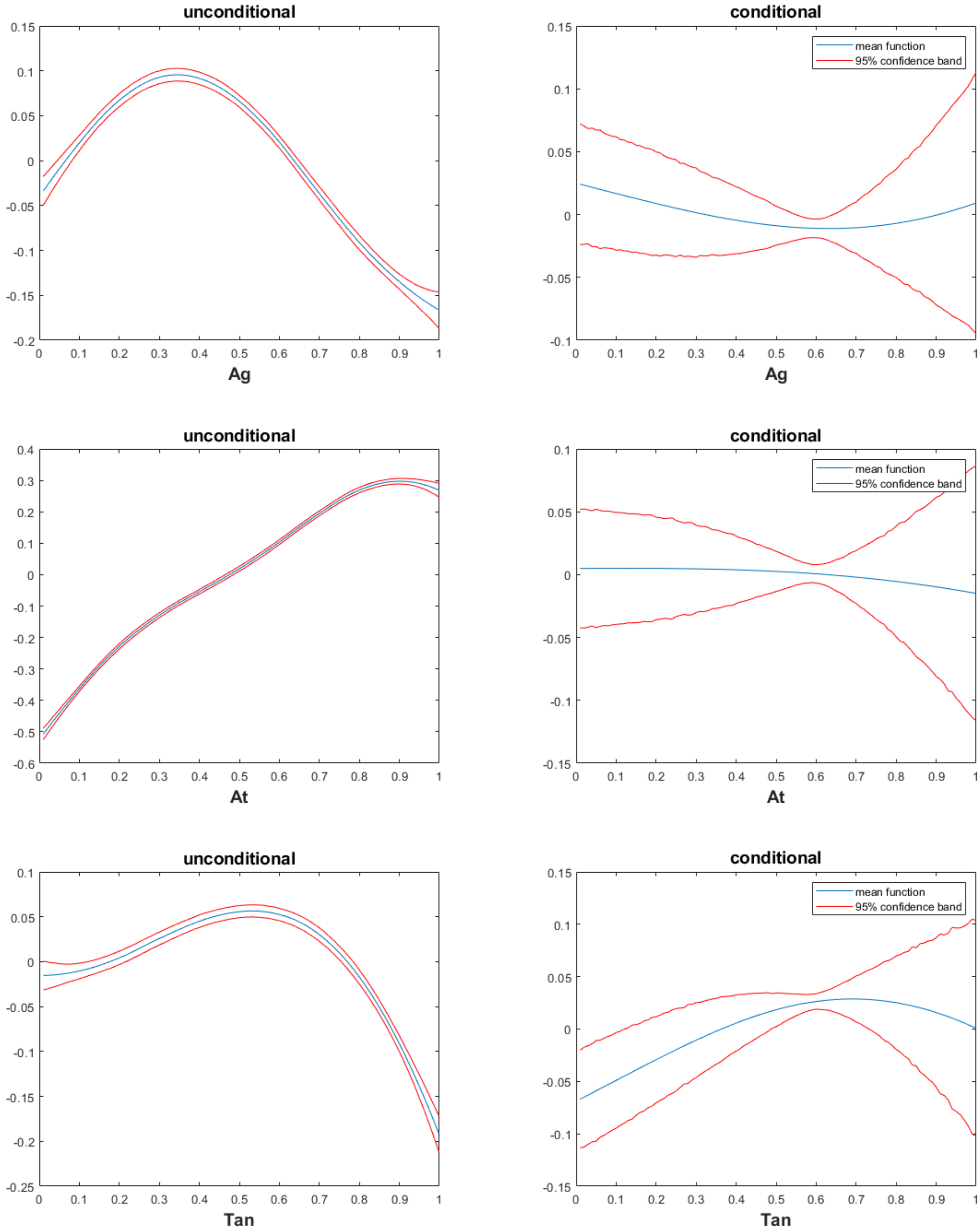

Figure 1
Unconditional and conditional mean function: Asset growth (ag), total assets (at) and Tangibility (tan)
Effect of normalized asset growth (ag), total assets (at) and tangibility (tan) on expected returns. The left figures present the unconditional relation between the given characteristic and option returns. The right figures report the association after conditioning on every characteristic. The 95% confidence bands are shown in red. The sample period is from January 1996 to March 2021.

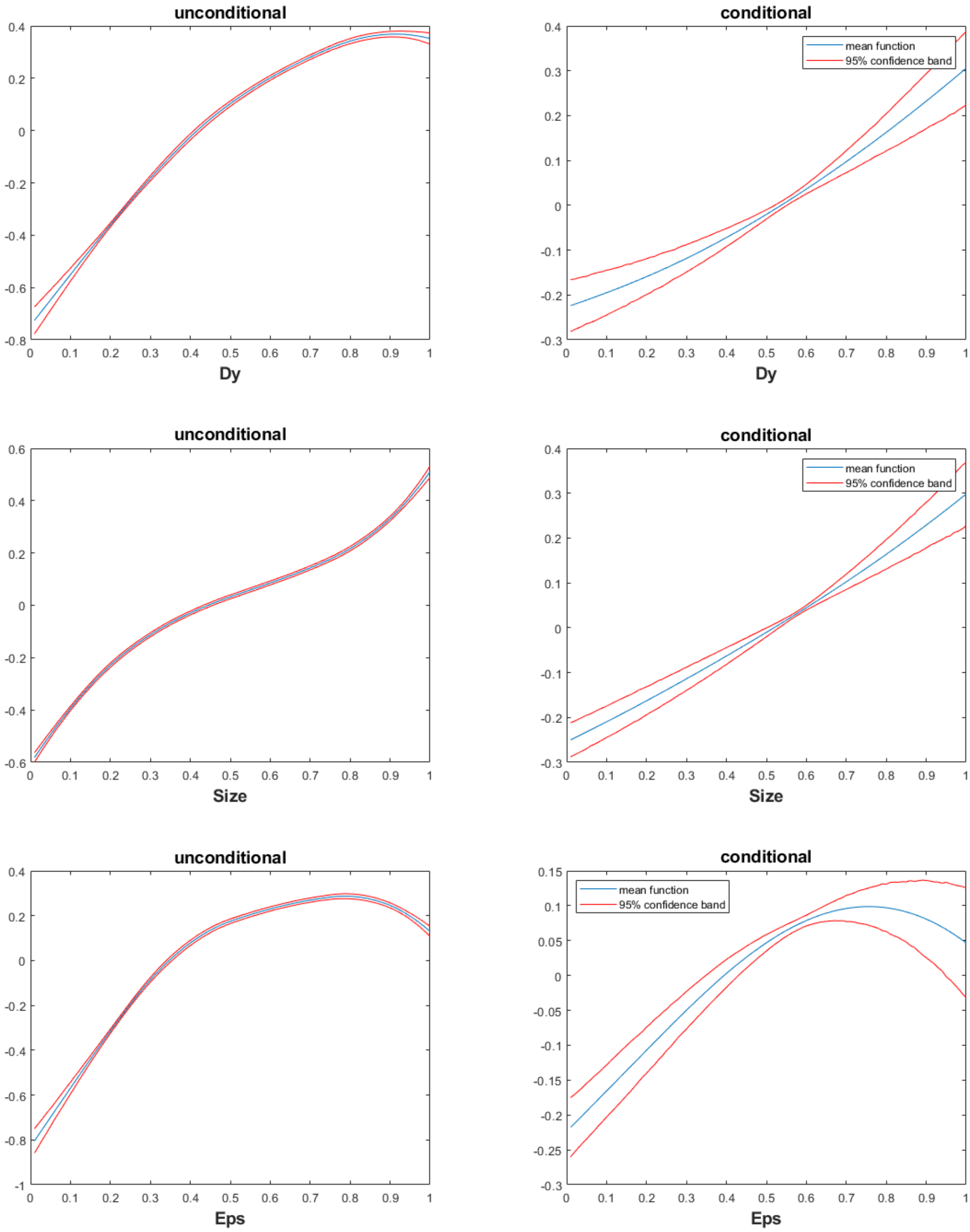

Figure 2
Unconditional and conditional mean function: Dividend to price (dy), size and earnings per share (eps)
Effect of normalized Dividend to Price (dy), size and earnings per share (eps) on expected returns. The left figures present the unconditional relation between the given characteristic and option returns. The right figures report the association after conditioning on every characteristic. Their 95% confidence bands are shown in red. The sample period is from January 1996 to March 2021.

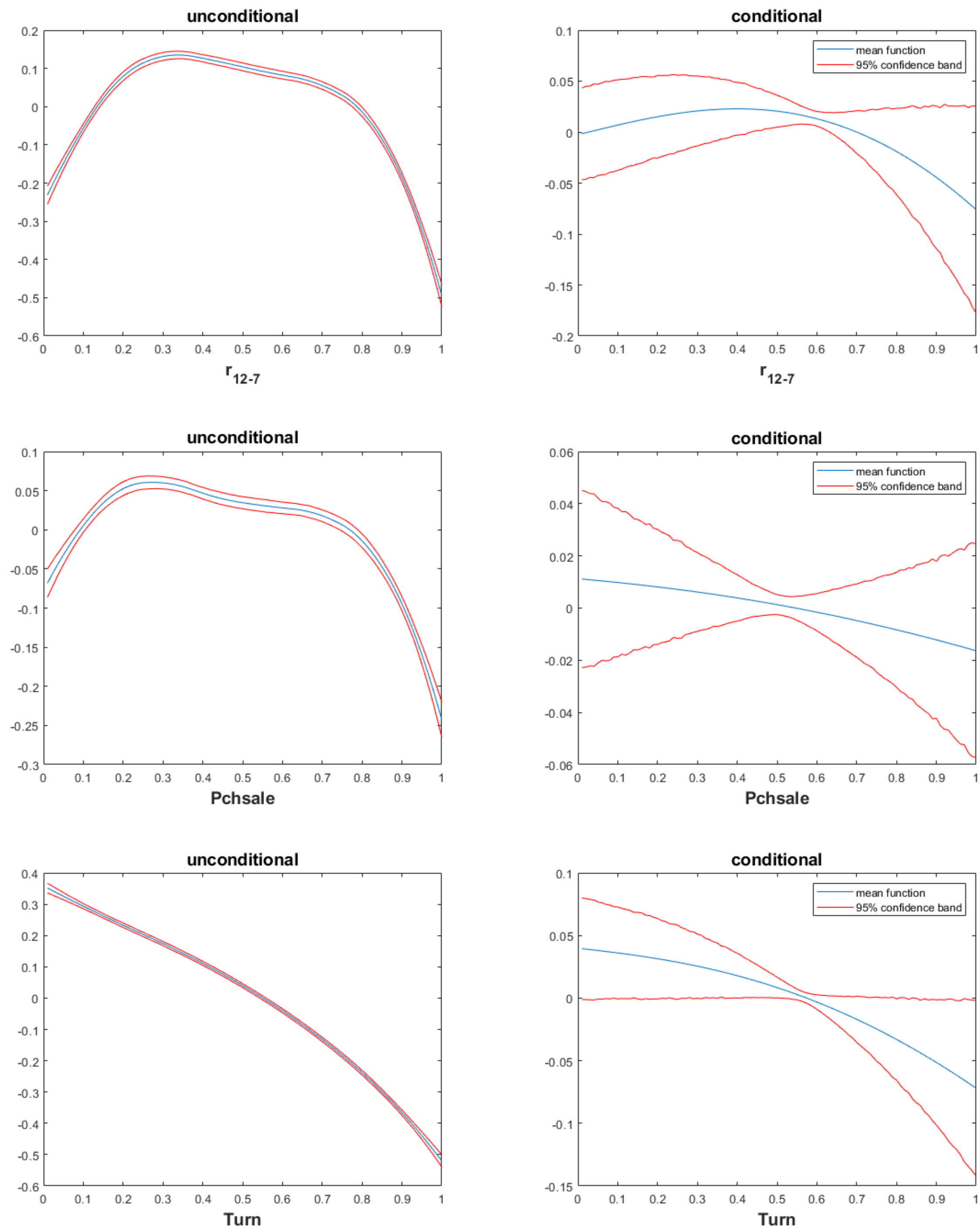

Figure 3

Unconditional and conditional mean function: Return from 12 to 7 months before prediction (r_{12-7}), % change in sales minus the %change in A/R (pchsale) and share turnover (turn)

Effect of normalized return from 12 to 7 months before prediction (r_{12-7}), % change in sales minus the % change in A/R (pchsale) and share turnover (turn) on expected returns. The left figures present the unconditional relation between the given characteristic and option returns. The right figures report the association after conditioning on every characteristic. Their 95% confidence bands are shown in red. The sample period is from January 1996 to March 2021.

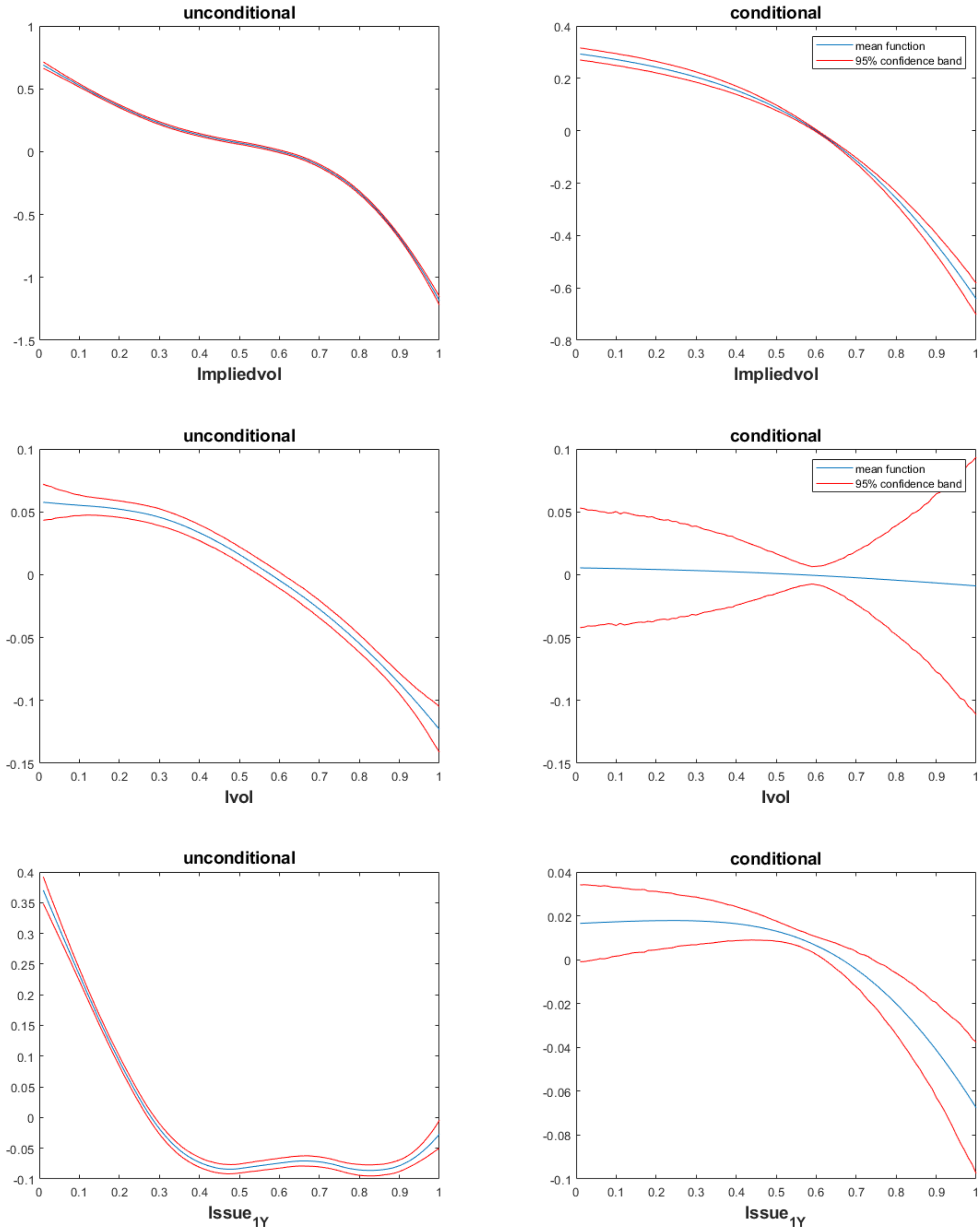

Figure 4

Unconditional and conditional mean function: The difference between implied and realized volatility (impliedvol), idiosyncratic volatility (ivol) and the 1 year change in shares outstanding (issue$_{1Y}$)

Effect of normalized 1 year change in shares outstanding (Issue$_{1Y}$), idiosyncratic volatility (ivol) and implied volatility (impliedvol) on expected returns. The left figures present the unconditional relation between the given characteristic and option returns. The right figures report the association after conditioning on every characteristic. Their 95% confidence bands are shown in red. The sample period is from January 1996 to March 2021.

Figure 3 plots mean functions for returns from 12 to 7 months before prediction (r_{12-7}), the percentage change in sales minus the percentage change in A&R (pchsale) and share turnover (turn). Surprisingly, the unconditional plots of r_{12-7} and pchsale are almost identical. Both characteristics were previously not selected by the portfolio sort method with t-statistics of -1.17 and -0.65 respectively. However, pchsale is selected by the Fama-Macbeth regression with a t-statistic of -7.245. In the adaptive group LASSO both characteristics are selected in three out of the seven models. The plots in Figure 3 show the unconditional and conditional mean function of model (1), which does not select either characteristic. When we condition on all other characteristics, the relation between r_{12-7} and pchsale with call option returns does not change. Firms with low past returns have lower expected delta-hedged option returns in the future. As past returns increase, the expected option returns increase as well. This effect persists until we reach firms in the top 30%-40% of past earners. After this point, further increases in past returns vastly decrease option returns. The top percentile of past earners also achieve the lowest delta-hedged option returns. We see the same relationship, albeit smoother, in the conditional mean function but the effect both characteristics have on option returns is now insignificant. Since we observe no short term reversal (r_{2-1}) in any of the nonparametric models and any long-term momentum (r_{36-12}) can also be ruled out, it is safe to say our results disagree with the conclusions of Heston et al. (2022) and Jones et al. (2020). The only momentum we observe is between 12 and 2 months before prediction. Except for model (2), every model selects one of three momentum characteristics: r_{6-2}, r_{12-2} and r_{12-7}. It appears these three momentum characteristics are somewhat interchangeable. In this sense, our results do agree with the aforementioned papers in that momentum can be used as a predictor of option returns. Unfortunately, we cannot conclusively determine which momentum characteristic is the true predictor since our selection is inconsistent. Since correlation between these characteristics is extremely high, it is probable that each of them will suffice as a predictor of call option returns. Turning our attention to the plots for share turnover, we notice that the slope coefficient of -0.097 of the linear model is not entirely inaccurate. This may be because effect of share turnover on returns is close to linear even in the nonparametric model. Although share turnover is not generally considered as a return predictor in the literature, it is selected in all seven nonparametric models.

Lastly, we direct our attention to Figure 4, where we plot the mean functions for yearly percentage change in outstanding shares (issue_{1Y}), idiosyncratic volatility (ivol) and the difference between implied and realized volatility (impliedvol). The alpha of impliedvol had a t-value -10.67 and the slope coefficient of impliedvol was -0.35 with a t-value of 88.01. It is no surprise that this characteristic is selected in all seven nonparametric models. We do however observe that the linear model heavily underestimates the magnitude of the slope coefficient. On average, firms with the highest difference between implied and realized volatility (a higher implied volatility relative to realized volatility) have call option returns that are 1% lower than firms with the lowest difference. The difference between implied and realized volatility was already found to be a predictor of delta-hedged returns in Goyal and Saretto (2009a). Horenstein et al. (2020) also incorporate it as one of the four factors in their model. Our results agree that options with an overestimated implied volatility are overpriced and yield lower delta-hedged returns in the near future. Although the relation implied by the group LASSO does not seem entirely linear, it is in line with the conclusions drawn from both the portfolio sorts and Fama-Macbeth regression. The difference between a call option's implied volatility and its realized volatility in the past month appears to be a strong indicator of a firm's expected delta-hedged returns. Next, we take a look at idiosyncratic volatility. Recall that Cao and Han (2013) claim delta-hedged returns decrease as idiosyncratic volatility increases. These findings are supported by Cao et al. (2021). Both studies rely on Fama-Macbeth regressions. In Table 4 we found a significant negative alpha of -0.22. The Fama-Macbeth coefficient for idiosyncratic volatility in Table 5 is also negative and significant at -0.096. The left hand side of Figure 4 shows the negative relation found in the literature. Only when we condition on all other characteristics and account for nonlinearities do we see that idiosyncratic volatility has no effect on delta-hedged returns. While we cannot conclusively determine why idiosyncratic volatility appears as an empirical irregularity in portfolio sorts, it is likely that the significant alpha is a result of the factor model's failure to take into account some form of aggregate risk. The last characteristic we examine is the yearly percentage change in outstanding shares. It is the only characteristic selected in the adaptive group LASSO that is not selected in the linear model, despite having a significant alpha. Although conditioning on all other characteristics in the linear model led us to believe that the effect of issue_{1Y} is insignificant, the nonparametric conditional mean function shows that returns are not effected by changes in the range of [0, 0.4] but they decrease as the number of newly issued shares increases beyond 0.4.

We briefly return to the objective of this paper, which is to find characteristics that provide incremental information on delta-hedged option returns. We sorted options into deciles based on each characteristic and created long-short portfolios with a long position in the highest decile and a short position in the lowest decile. 34 of these trading strategies have returns that cannot be explained by a traditional factor model. Our contention with these results and by extension the conclusions in Horenstein et al. (2020) and Büchner and Kelly (2022) is that the significance of most of these anomalies is most likely a result of the factor models in question and not of inefficiencies in the market. After accounting for nonlinearities and modelling the conditional mean function directly we find that only eight characteristics consistently provide incremental information on option returns: Dividend to price ratio, earnings per share, percentage change in outstanding shares, maximum daily return, return on equity, size, share turnover and the difference between implied and historical volatility.

4.4 Time-variation in selected characteristics

In Section 3.3 we opted to pool the observations across time to decrease estimation uncertainty (Gu et al., 2020). By doing this, we implicitly assume that the betas in Equations (9) and (11) do not vary over time. However, literature regarding asset returns documents that the conditional mean function is time-varying for equities in the cross-section (see, for example, Ghysels (1998) and Dangl and Halling (2012)). We devote this section to examine if the estimated coefficients of the models in Table 6 are in fact time-varying by nature.

Ideally, we would estimate Equation (9) and (11) separately for each time observation t. Unfortunately we do not have enough firms in the cross-section to achieve reliable coefficient estimates this way. Instead, we will estimate the adaptive group LASSO using a rolling window with 100 time observations, which is equivalent to around 10 years of data. We estimate the adaptive group LASSO using an initial estimation sample from January 1996 to June 2006 and fix the selected characteristics. Subsequently, we roll the window forward by one month and run the adaptive group LASSO again using only those characteristics. This is repeated until the entire sample has been exhausted. Using the initial sample period, the adaptive group LASSO selects seven characteristics: Size, issue_{1Y}, dy, ep, r_{36-12}, maxret and impliedvol. The resulting time-varying conditional mean functions for each selected characteristic are shown in Figures 5, 6 and 7.

In Figure 5 we see that the conditional mean function for size and the one year change in outstanding shares is not constant throughout time. Small firms have lower expected option returns but this effect seems most prominent during 2006 to 2012 and dissipates over time. Towards the year 2020 size has a much smaller effect on expected option returns. The bottom panel plots the conditional mean function for the change in outstanding shares, which actually displays opposite developments over time. In 2008, firms at the extreme ends of the spectrum exhibit the lowest average option returns, with the highest returns obtained by firms in the middle. Over time, firms that issued more shares obtained significantly lower returns, with an all time low of -0.15 in 2018, whereas firms which decreased their outstanding shares achieved returns of more than 0.05 from 2014 till 2021.

Figure 6 shows that the conditional mean function with regard to dividend to price does not vary over time. During the entire sample period, firms earn higher option returns when their dividend to price ratio increases relative to other firms in the cross-section. We see time-variation in the effect of earnings to price. During the start of the sample period, option returns are negatively correlated with a firm's earnings to price ratio. From 2006 to 2014 firms with the highest earnings to price ratio consistently achieve the lowest

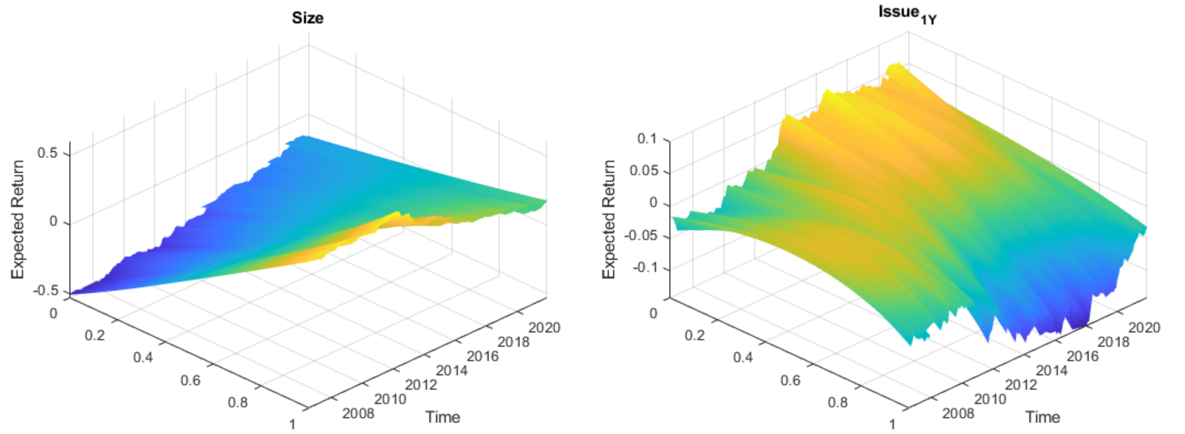

Figure 5
Time varying conditional mean function of size and the one year change in outstanding shares
Influence of size and the one year percentage change in outstanding shares (issue$_{1Y}$) on expected returns over time. The sample period is from December 2006 to March 2021.

expected returns. It is not until 2016 that the curve flattens and the effect of earnings to price on returns becomes less noticeable. After 2018, firms with a low earnings to price ratio no longer yield higher option returns than firms with a high earnings to price ratio.

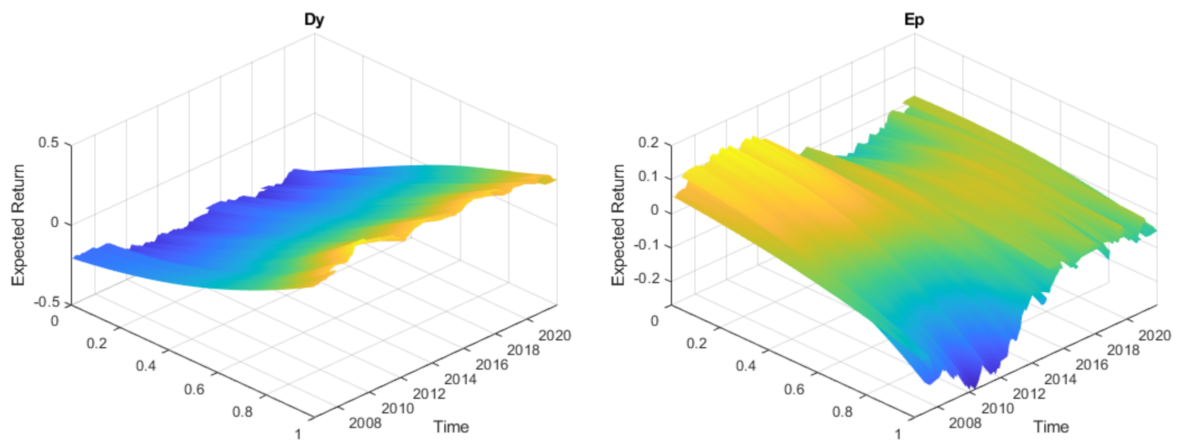

Figure 6
Time varying conditional mean function of dividend to price and earnings to price
Influence of dividend to price (dy) and earnings to price (ep) on expected returns over time. The sample period is from December 2006 to March 2021.

Figure 7 plots the conditional mean function of returns from 36 to 12 months before prediction, maximum daily return and the difference between implied and realized volatility. The surface plot for r_{36-12} shows that, at the start of the sample, firms in the upper quantile yield higher option returns. Option returns decrease as long-term momentum decreases. Over the next six years the effect long-term momentum has on delta-hedged option returns levels and firms with high momentum start to yield similar option returns

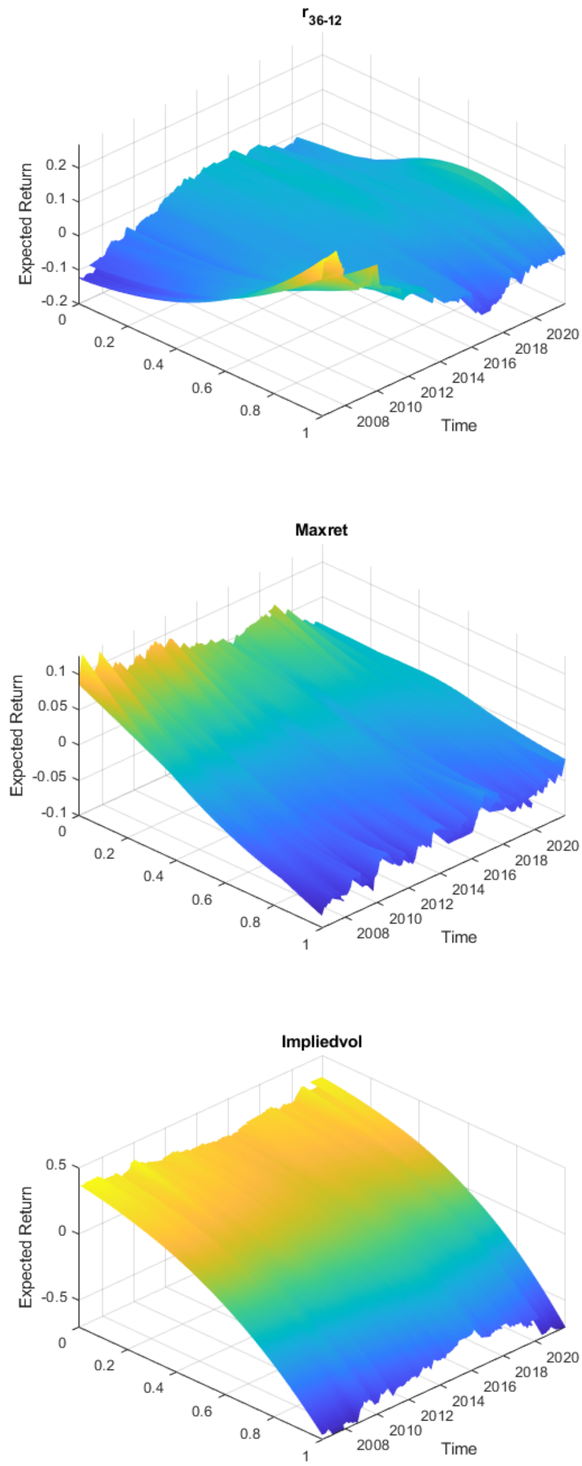

Figure 7
Time varying conditional mean function of maximum daily return, return 36 to 12 months before prediction and the difference between implied and realized volatility
Influence of maximum daily return (maxret), return 36 to 12 months before prediction (r_{36-12}) and difference between implied and realized volatility (impliedvol) on expected returns over time. The sample period is from December 2006 to March 2021.

to firms with low momentum. In the following four years momentum actually decreases a firm's expected option returns. In Section 4.1.3 we discussed that both Brooks et al. (2018) and Heston et al. (2022) find a positive long-term effect of momentum on option returns. While we observe this effect in the beginning of our sample, the positive relation between long-term momentum and option returns is not robust over time. Next we look at the conditional mean function of maximum daily return which remains fairly constant throughout the entire sample period. Firms with low maximum daily return in the previous month consistently yield higher delta-hedged call option returns in the following month and those option returns appear to decrease linearly with increases in maximum daily return. This effect persists throughout the entire sample period but is less pronounced during the later stages. Moving on to the effect of implied volatility, more specifically, the difference between realized volatility and implied volatility, hardly any changes can be seen during the entire 13 year sample. Options whose implied volatility is significantly higher than their realized volatility yield far lower option returns than options with implied volatility below their historical volatility. Again, the explanation for this phenomenon is that an option's implied volatility tends to mean-revert to its realized volatility. An option with a higher implied volatility is thus overpriced and with time the market corrects that price back until the implied volatility equals the options realized volatility. This decrease in the option's price is inevitably associated with lower option returns. On the other end of the spectrum, options with an underestimated implied volatility will likely increase in price in the following months. Based on magnitude, size, dividend to price and the difference between implied and realized volatility have the strongest effect on average expected call option returns.

Although the decision to pool the observations of Equation (9) across time is a near necessity, we must conclude that the assumption of constant betas in the nonparametric model does not hold. To further investigate whether time varying coefficients are negatively impacting our model, we assess if model selection is consistent throughout time. We estimate the nonparametric model using data from January 1996 to December 2011, roll the sample forward by one year and select characteristics again. We repeat this till the end of the sample. Characteristics selection over time is shown in Figure 8. Ideally we would see consistent model selection over the entire ten year estimation period from 2012 to 2021. This is not the case. Results are only partially in line with the selected characteristics in Table 6. For instance, issue_{1Y}, profit, dy, maxret and impliedvol are selected in at least five out of ten years. These characteristics were also selected consistently when we considered the entire sample period in Table 6. however, share turnover, which was selected in all seven nonparametric models, is only selected in two out of ten years and return on equity is selected in just four years. What is most surprising is that size, eps, ep and baspread are not selected at all. Even though they were chosen in seven, four, two and three out of the seven models in Table 6 respectively. Figure 8 implies that profit, gma, bm, dy, maxret and impliedvol are consistently selected by the nonparamet-

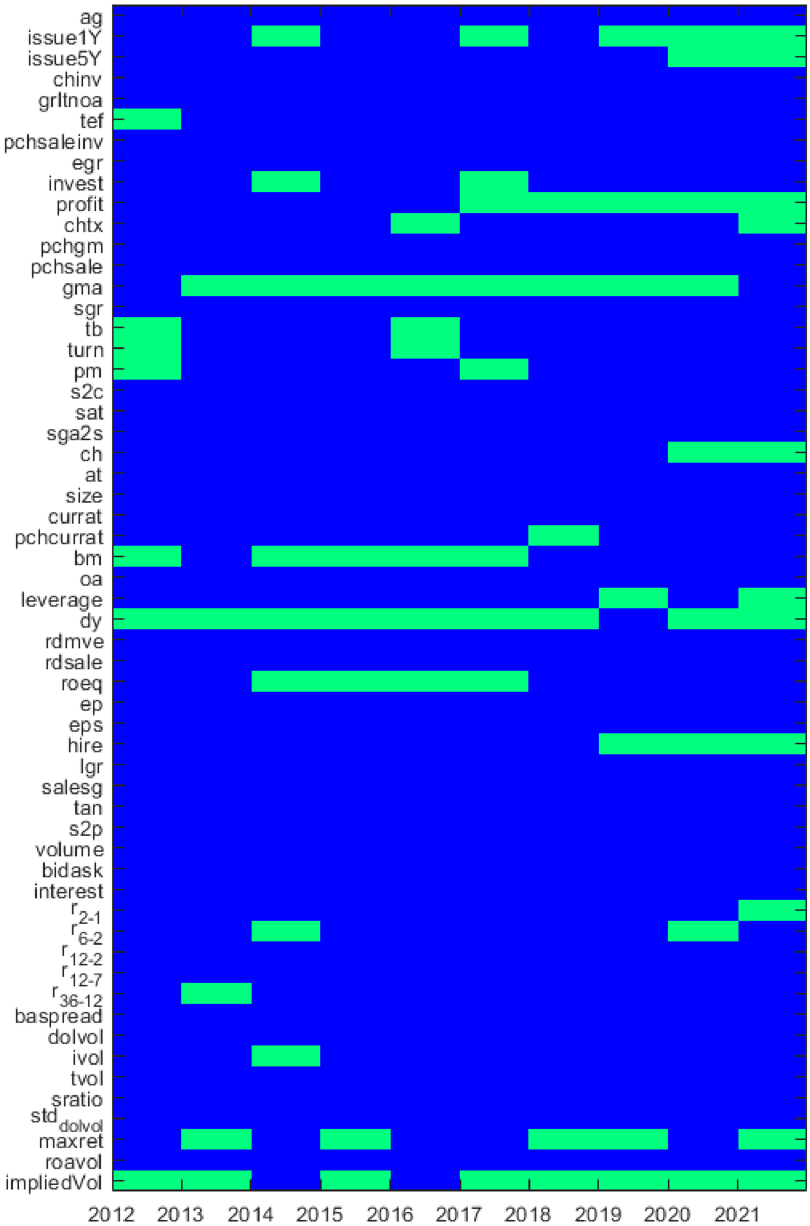

Figure 8
Selected characteristics in nonparametric model with rolling window selection
Characteristics are selected using the adaptive group LASSO with a sample period from January 1996 to December 2011. The sample is then repeatedly rolled forward by one year. Selected characteristics in each year are shown in green.

ric model over time, while Table 6 indicates that gma and bm have no predictive power on returns when we consider the entire sample. We do observe that the nonlinear model consistently selects less characteristics than the linear model shown in Figure 9, although the linear model obtains more consistent model selection in general.

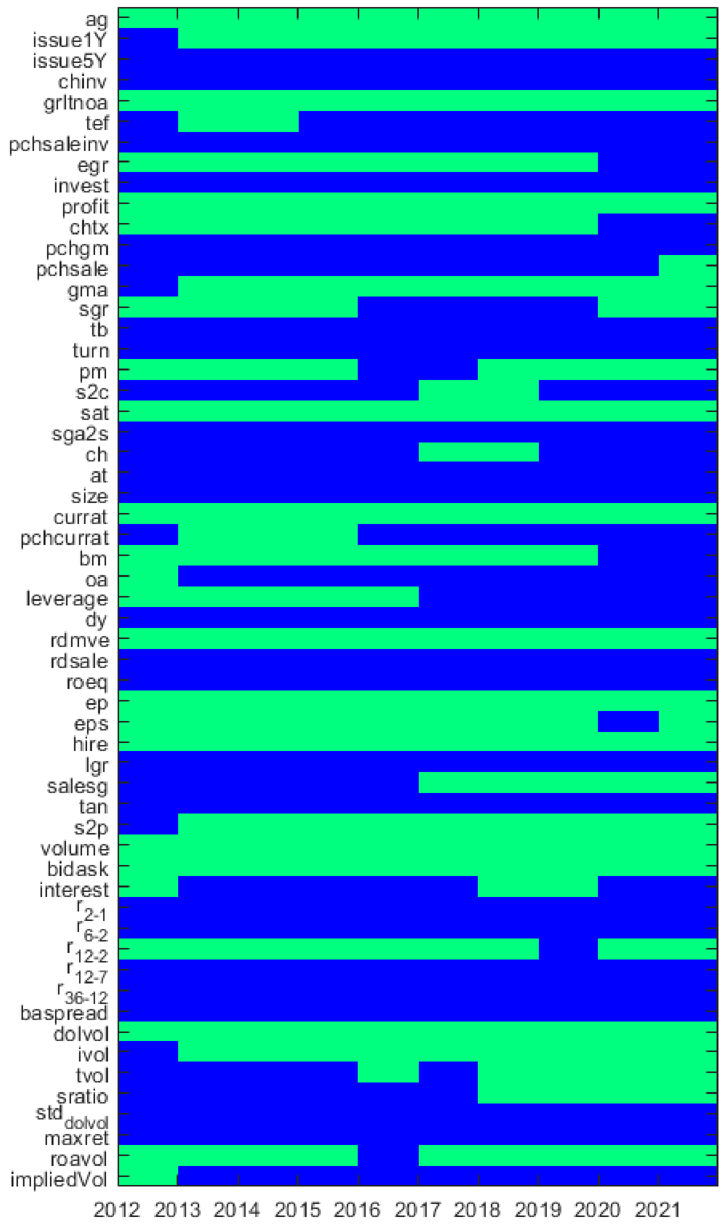

Figure 9
Selected characteristics in linear model with rolling window selection
Characteristics are selected using a linear model with a sample period from January 1996 to December 2011. The sample is then repeatedly rolled forward by one year. Selected characteristics in each year are shown in green.

4.5 Out-of-sample analysis

While we provided arguments for benefits of the nonparametric model over linear models, it is crucial to study the efficacy of each approach in practice. We split the sample into two parts, using data from 1996 to 2018 to select characteristics. The selected characteristics of each model are then used to predict the next month's option returns. We create a hedged long-short portfolio using the realized option returns of that month, going long in the ten percent of firms with the highest predicted returns and shorting the ten percent of firms with the lowest predicted option returns. We roll the estimation sample forward by one month and repeat the process for the remainder of the sample. We report average returns, standard deviation, Sharpe ratio, skewness and kurtosis of the out-of-sample hedge portfolios. We follow Lewellen (2014) to assess predictability by regressing the monthly realized returns on monthly predicted returns. We report the average β's and R^2s of this regression. Lastly we report the average pricing error as it is defined in Goyal and Saretto (2022).

Table 7
Out-of-sample return prediction of the adaptive group LASSO for different knots

Model	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO	LASSO
Firms	All	All	All	All	All	All	All	All
Sample	1996-2018	1996-2018	1996-2018	1996-2018	2012-2018	2012-2018	2012-2018	2012-2018
Knots	20	15	10	5	20	15	10	5
# Selected	12	12	13	12	7	9	11	8
				Long-Short Portfolio				
Mean Return	0.17	0.24	0.17	0.14	0.06	0.07	0.17	0.13
SD	1.50	1.46	1.19	1.40	1.17	1.34	1.23	1.40
Sharpe Ratio	0.40	0.58	0.51	0.35	0.18	0.19	0.47	0.32
Skewness	1.09	2.36	1.10	1.46	0.20	0.21	0.45	-0.40
Kurtosis	6.99	12.83	7.11	8.11	5.40	4.33	4.94	5.53
β	0.14	0.18	0.13	0.12	-0.03	0.06	0.15	0.13
R^2	-0.22%	0.26%	0.49%	0.10%	-0.18%	-0.12%	-0.23%	-0.13%
Pricing error	77.76%	78.44%	79.20%	80.56%	82.25%	81.66%	81.00%	81.91%
				Long Leg				
Mean Return	-1.00	-0.85	-0.93	-0.94	-1.02	-1.13	-1.12	-1.08
SD	1.98	2.21	2.04	2.10	1.94	1.83	1.85	1.88
Sharpe Ratio	-1.75	-1.33	-1.58	-1.56	-1.93	-2.13	-2.10	-1.99
Skewness	1.90	1.86	1.76	1.77	2.20	1.83	2.11	1.96
Kurtosis	9.30	7.91	8.16	8.26	10.29	12.05	11.59	10.48
β	3.02	1.90	2.44	2.66	2.56	3.43	3.41	3.10
R^2	-0.76%	-0.92%	-0.96%	-0.80%	-0.27%	-0.27%	-0.26%	-0.25%
Pricing error	92.02%	92.41%	92.26%	92.95%	98.27%	97.35%	97.69%	97.78%
				Short Leg				
Mean Return	1.17	1.09	1.11	1.09	1.08	1.20	1.29	1.21
SD	1.81	1.75	1.73	1.76	1.94	1.61	1.58	1.69
Sharpe Ratio	2.25	2.17	2.21	2.14	1.83	2.58	2.82	2.47
Skewness	-2.48	-2.52	-2.63	-2.72	-2.67	-0.76	-0.83	-0.99
Kurtosis	12.13	11.39	11.87	12.73	13.44	5.90	6.38	5.85
β	0.85	0.76	0.78	0.77	0.95	1.17	1.25	1.19
R^2	0.53%	1.16%	1.20%	0.87%	-1.31%	-1.35%	-1.70 %	-1.08%
Pricing error	81.53%	81.56%	80.45%	78.79%	79.06%	75.96%	75.21%	75.67%

This table reports out of sample results for hedge portfolios going long in the decile of firms with the highest predicted returns and taking a short position in the lowest decile of predicted returns. The out-of-sample period is from January 2019 to March 2021. Characteristics to predict returns are selected using the adaptive group LASSO with data from 1996 to 2018 or from 2012 to 2018. The selected characteristics are then fixed and the estimation window is rolled forward by one month to estimate the model again and create hedge portfolios for that month. We repeat this for the remaining out-of-sample period. The average return, standard deviation, Sharpe ratio, skewness and kurtosis of the resulting hedge portfolios are reported, as well as the average β and R^2 from regressions of realized returns on predicted returns. We also report the model's relative pricing error as its defined in Goyal and Saretto (2022).

We would expect that more knots allow for a better approximation of the conditional mean function and thus better out-of-sample return prediction. Table 7 shows that this is not the case in practice. In regard to average returns and Sharpe ratios the nonparametric model with 20 knots yields worse results than with 15 and 10 knots. Due to the potential of time-variation in the models parameters as discussed in Section 4.4 the model is estimated using two different samples. One with a sample period from 1996 to 2018 and a more recent sample period from 2012 to 2018. Reducing the sample period aims to minimize the estimation error from time variability in the parameters with the obvious downside of a smaller sample size. This drawback is most noticeable for the models with 15 and 20 knots. The more knots we use the more parameters we need to estimate and logically we require a larger sample to achieve precise estimates. Unfortunately the results for the models with 5 knots and 10 knots do not improve when we consider the shorter sample period. Sharpe ratios as well as R-squared's are equal or lower compared to when the entire sample is used. if the shorter sample period has any benefit in regard to parameter consistency over time, it is entirely offset by the increased estimation uncertainty due to a lack of observations.

In Table 6 we saw that in-sample model selection is not entirely consistent across different numbers of knots. This is no different out of sample and Table 7 shows that the ideal number of interpolation points appears to be between 10 and 15. In Section 4.3 we considered the adaptive group LASSO with quadratic splines. While intuitive, there are endless orders of splines to choose from. We estimate the nonparametric model for five different orders of splines and present the results in Table 8. Although the highest order spline yields the highest Sharpe ratio of 1.28, Sharpe ratios do not increase monotonically with an increase in spline order. In fact, the fourth order spline yields worse out-of-sample Sharpe ratios than the third, second and even first order spline. Surprisingly, the first order spline, which equates to a linear conditional mean function, yields competitive Sharpe ratios to the quadratic spline. This stands in contrast with the negative R^2 of -1.15% that it obtains for the long-short hedge portfolio, while every higher order yields R^2's of at least 0.30%. It is difficult to draw definitive conclusions from the R^2 values since out-of-sample predictability is very low even in the most optimistic scenario. In any case, a higher order spline does not necessarily yield better out-of-sample results. Because of this, we will restrict ourselves to a quadratic spline with ten knots when comparing out-of-sample prediction between the nonparametric and the linear model.

Turning to the primary purpose of this section, we examine whether the nonparametric model provides exclusive benefits in out-of-sample prediction. Table 9 reports the results for the adaptive group LASSO with 10 knots and a quadratic spline alongside results of the linear model. In columns 1 and 2 we consider out-of-sample performance of the nonlinear and the linear model using the entire sample period from 1996 to 2018 and taking all firms into consideration. Regarding long-short portfolios, the nonparametric model consistently yields higher Sharpe ratios, has betas closer to 1 and achieves better R^2s than the linear model. The nonparametric model achieves an annual Sharpe ratio of 0.51, which is 76% higher than the linear model. We see the linear model's tendency to

Table 8
Out-of-sample return prediction of the adaptive group LASSO for different orders

Model	LASSO	LASSO	LASSO	LASSO	LASSO
Firms	All	All	All	All	All
Sample	1996-2015	1996-2015	1996-2015	1996-2015	1996-2015
Knots	10	10	10	10	10
Order	1	2	3	4	5
# Selected	11	10	9	9	11
			Long-Short Portfolio		
Mean Return	0.35	0.33	0.28	0.21	0.40
Standard Deviation	1.40	1.24	1.21	1.12	0.99
Sharpe Ratio	0.87	0.92	0.82	0.64	1.39
Skewness	0.83	-0.30	-0.16	-0.41	1.08
Kurtosis	6.14	5.77	6.28	7.36	5.73
β	0.27	0.26	0.23	0.18	0.29
R^2	-1.15%	0.30%	0.33%	0.62%	0.30%
Pricing error	97.59%	96.44%	95.87%	94.71%	98.85%
			Long Leg		
Mean Return	-0.36	-0.43	-0.48	-0.47	-0.42
Standard Deviation	0.99	0.98	0.94	0.95	0.98
Sharpe Ratio	-1.27	-1.50	-1.77	-1.69	-1.47
Skewness	0.38	0.59	0.58	0.58	0.51
Kurtosis	2.90	3.24	3.27	3.19	2.96
β	1.04	1.36	1.61	1.55	1.32
R^2	-5.55%	-3.45%	-2.84%	-1.62%	-2.06%
Pricing error	92.61%	93.76%	92.40%	91.94%	92.40%
			Short Leg		
Mean Return	0.72	0.75	0.76	0.67	0.77
Standard Deviation	1.02	1.05	1.03	0.91	0.86
Sharpe Ratio	2.45	2.49	2.58	2.58	3.08
Skewness	-0.44	-0.07	-0.03	-0.72	-0.42
Kurtosis	2.88	4.24	4.41	3.99	3.09
β	0.48	0.49	0.50	0.45	0.52
R^2	1.36%	1.40%	1.34%	1.76%	1.77%
Pricing error	105.69%	102.37%	106.53%	104.84%	108.21%

This table reports out of sample results for hedge portfolios going long in the decile of firms with the highest predicted returns and taking a short position in the lowest decile of predicted returns. The out-of-sample period is from January 2016 to December 2018. Characteristics to predict returns are selected using the adaptive group LASSO with data from 1996 to 2015. The selected characteristics are then fixed and the estimation window is rolled forward by one month to estimate the model again and create hedge portfolios for that month. We repeat this for the remaining out-of-sample period. The average return, standard deviation, Sharpe ratio, skewness and kurtosis of the resulting hedge portfolios are reported, as well as the average β and R^2 from regression of realized returns on predicted returns. We also report the model's relative pricing error as its defined in Goyal and Saretto (2022).

overfit to the data. The linear model selects 23 characteristics while the nonparametric model only requires 13 characteristics to predict the cross-section of returns. The LASSO obtains higher average returns as well as a lower standard deviation. The portfolio returns of the nonlinear model have a much higher kurtosis than the linear model. This may indicate that the nonparametric model is better in predicting the extreme end of returns. When we consider the long and short legs separately we find that the nonlinear model performs very well at predicting the short leg of returns. Its Sharpe ratio is 46% higher than the linear model and the R^2 is 1.20%, whereas the linear model has an R^2 of -1.30%, which implies that the linear model predicts worse than just using a simple mean. However, the opposite is true for the long leg of returns. Here the nonlinear

Table 9
Out-of-sample comparison between the nonparametric and linear model

Model	LASSO	Linear	LASSO	Linear	LASSO	Linear	LASSO	Linear
Firms	All	All	size $> d_5$	size $> d_5$	size $> d_1$	size $> d_1$	All	All
Sample	1996-2018	1996-2018	1996-2018	1996-2018	1996-2018	1996-2018	2006-2018	2006-2018
Knots	10	-	10	-	10	-	10	-
# Selected	13	29	8	18	11	30	11	30
Long-Short Portfolio								
Mean Return	0.17	0.13	0.22	0.23	0.03	-0.43	0.19	0.19
SD	1.19	1.53	1.70	2.61	1.74	1.33	0.88	1.14
Sharpe Ratio	0.51	0.29	0.44	0.31	0.07	-1.12	0.75	0.57
Skewness	1.09	-0.22	-0.92	0.62	-0.34	-0.59	-0.05	-0.23
Kurtosis	7.11	4.26	5.13	4.25	4.53	5.83	4.53	6.93
β	0.13	0.01	0.23	0.06	0.02	-0.05	0.19	0.02
R^2	0.49%	-0.24%	0.34%	-0.88%	0.02%	-0.73%	0.14%	0.06%
Pricing error	79.20%	86.84%	78.51%	82.83%	82.35%	93.38%	78.44%	84.45%
Long Leg								
Mean Return	-0.93	-0.71	-1.03	-0.64	-1.02	-1.11	-0.98	-0.96
SD	2.04	2.35	2.06	1.83	1.71	2.15	1.70	1.70
Sharpe Ratio	-1.58	-1.04	-1.74	-1.22	-2.08	-1.79	-2.00	-1.97
Skewness	1.76	2.09	0.43	1.56	0.26	2.97	2.84	0.88
Kurtosis	8.16	10.00	6.10	7.76	3.73	15.56	14.88	5.20
β	2.44	0.10	4.46	0.67	3.54	0.15	3.86	0.16
R^2	-0.96%	1.27%	-3.95%	0.96%	-0.95%	0.90%	-3.38%	1.79%
Pricing error	92.26%	82.65%	87.53%	76.61%	93.96%	83.06%	92.37%	83.14%
Short Leg								
Mean Return	1.11	0.84	1.25	0.87	1.06	0.68	1.17	1.15
SD	1.73	1.92	1.88	1.92	1.46	2.39	1.71	1.56
Sharpe Ratio	2.21	1.51	2.31	1.57	2.51	0.99	2.38	2.56
Skewness	-2.63	-2.80	-1.11	-0.30	-0.81	-2.46	-2.69	-0.91
Kurtosis	11.87	14.81	7.86	4.75	4.92	11.05	12.76	5.57
β	0.78	0.81	1.20	0.28	0.84	0.45	0.95	0.72
R^2	1.20%	-1.30%	0.50%	-0.11%	0.27%	-1.30%	0.63%	-4.31%
Pricing error	80.45%	97.23%	76.68%	83.22%	83.00%	119.11%	78.78%	98.47%

This table reports out of sample results for hedge portfolios going long in the decile of firms with the highest predicted returns and taking a short position in the lowest decile of predicted returns. The out-of-sample period is from January 2019 to March 2021. Additionally, the model is ran separately for large firms (size $> d_5$) and for all except the smallest firms (size $> d_1$). Characteristics to predict returns are selected using either the adaptive group LASSO or a linear model with data from 1996 to 2018 or from 2012 to 2018. Those characteristics are then fixed and the estimation window is rolled forward by one month to estimate the model again and create hedge portfolios for that month. We repeat this for the remaining out-of-sample period. The average returns, standard deviations, Sharpe ratios, skewness and kurtosis of the resulting hedge portfolios are reported, as well as the average β and R^2 from regressions of realized returns on predicted returns. We also report the model's relative pricing error as its defined in Goyal and Saretto (2022).

model does not yield accurate predictions. On average, delta-hedged returns during the out-of-sample period are negative at -0.88. At a reasonable level of predictability, we expect the mean returns of the long leg to be higher than -0.88, instead the nonlinear model has an average return of -0.93 in the long leg. We could expect to outperform this by randomly selecting firms each month. While a beta of 1 is ideal, we obtain a beta of 2.44, which indicates that the nonlinear model heavily underestimates the expected return dispersion (Lewellen, 2014). The linear model actually outperforms the nonlinear model in the long leg, with an average monthly return of -0.71 and an R-squared of 1.27%. This is reaffirmed by the pricing errors which are 10% higher for the LASSO.

We also compare out of sample performance large firms (size $> d_5$) and all but the smallest firms (size $> d_1$) in columns 3 to 6. In general, the results are consistent with our previous findings. The nonlinear model gives better Sharpe ratios, betas closer to 1 and higher R^2s but is unable to predict accurately in the long leg. It appears that nonlinearities are important when predicting call option returns, but the adaptive group LASSO is only able to achieve this in the extreme negative end of returns. Columns 7 and 8 consider an sample period from 2006 to 2018 rather than the full period of 1996 to 2018. The results are similar to columns 1 and 2. To summarize, the long-short portfolio returns of the nonparametric model consistently have higher Sharpe ratios, higher R^2s than the linear model. The adaptive group LASSO also yields lower pricing errors.

5 Conclusion

We apply the adaptive group LASSO to find characteristics that provide incremental information on expected call option returns. (1) We analyze whether characteristics have incremental predictive power on expected returns. (2) We document whether the influence and selection of characteristics varies over time and lastly (3) we compare the performance of the adaptive group LASSO with the conventional linear approach in out-of-sample forecasting.

A summary of the results is as follows: (1) Considering a total of 57 characteristics, the number of characteristics required to explain the cross-section of expected delta-hedged call option returns is in the range of 6 and 13 depending on the sample period, number of interpolation points, order of spline and which firms are taken into consideration. (2) The influence of characteristics on expected returns, as well as the selection of those characteristics varies considerably over time. (3) In every sub-sample, the amount of characteristics the nonlinear model selects is less than half that of its linear counterpart. The nonparametric model obtains out-of-sample Sharpe ratios that are between 37% and 75% higher than those of the linear model. this effect is robust with respect to the number of interpolation points and the order of spline that we choose. Results are also robust with regard to firm size. Characteristics that are consistently selected to have incremental information on option returns are: The one year percentage change in outstanding shares, maximum daily return, the difference between implied and realized volatility, size, dividend to price ratio and share turnover. Although we find some level of predictability in option returns, our results show that much of the profitability in option returns documented in the extant literature can be explained through a nonparametric model. The implication is that, although imperfect, option markets still appear to exhibit an extraordinary level of efficiency and that pricing options using a framework based on the efficient market hypothesis is not entirely unreasonable.

References

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The journal of finance*, 61(1):259–299, 2006.

Jennie Bai, Turan G Bali, and Quan Wen. Common risk factors in the cross-section of corporate bond returns. *Journal of Financial Economics*, 131(3):619–642, 2019.

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

Sati P Bandyopadhyay, Alan G Huang, and Tony S Wirjanto. The accrual volatility anomaly. *Unpublished Manuscript, University of Waterloo*, 2010.

David S Bates. Hedging the smirk. *Finance Research Letters*, 2(4):195–200, 2005.

Demir Bektić, Josef-Stefan Wenzler, Michael Wegener, Dirk Schiereck, and Timo Spielmann. Extending fama–french factors to corporate bond markets. *The Journal of Portfolio Management*, 45(3):141–158, 2019.

Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics*, 186(2):345–366, 2015.

Derrick A Bennett. How can i deal with missing data in my study? *Australian and New Zealand journal of public health*, 25(5):464–469, 2001.

Mark Broadie, Mikhail Chernov, and Michael Johannes. Understanding index option returns. *The Review of Financial Studies*, 22(11):4493–4529, 2009.

Robert Brooks, Mobina Shafaati, and Don M Chance. The cross-section of individual equity option returns. *Available at SSRN*, 2018.

Matthias Büchner and Bryan Kelly. A factor model for option returns. *Journal of Financial Economics*, 143(3):1140–1161, 2022.

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

Jie Cao, Bing Han, Xintong Zhan, and Qing Tong. Option return predictability. In *Review of Financial Studies accepted, 27th Annual Conference on Financial Economics and Accounting Paper, Rotman School of Management Working Paper*, 2021.

Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.

Peter Carr and Liuren Wu. Option profit and loss attribution and pricing: A new framework. *The Journal of Finance*, 75(4):2271–2316, 2020.

Long Chen and Lu Zhang. A better three-factor model that explains more anomalies. *Journal of Finance*, 65(2):563–595, 2010.

Long Chen, Robert Novy-Marx, and Lu Zhang. An alternative three-factor model. *Available at SSRN 1418117*, 2011.

Tarun Chordia, Richard Roll, and Avanidhar Subrahmanyam. Market liquidity and trading activity. *The journal of finance*, 56(2):501–530, 2001.

Peter Christoffersen, Mathieu Fournier, and Kris Jacobs. The factor structure in equity options. *The Review of Financial Studies*, 31(2):595–637, 2018a.

Peter Christoffersen, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui. Illiquidity premia in the equity options market. *The Review of Financial Studies*, 31(3):811–851, 2018b.

George M Constantinides, Jens Carsten Jackwerth, and Alexi Savov. The puzzle of index option returns. *Review of Asset Pricing Studies*, 3(2):229–257, 2013.

Thomas Dangl and Michael Halling. Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181, 2012.

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

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

Eugene F Fama and Kenneth R French. Dissecting anomalies. *The Journal of Finance*, 63(4):1653–1678, 2008.

Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

Jennifer Francis, Ryan LaFond, Per M Olsson, and Katherine Schipper. Costs of equity and earnings attributes. *The accounting review*, 79(4):967–1010, 2004.

Joachim Freyberger, Andreas Neuhierl, and Michael Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

Eric Ghysels. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573, 1998.

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

Amit Goyal and Alessio Saretto. Option returns and the cross-sectional predictability of implied volatility. *Journal of Financial Economics*, 95:310–326, 2009b.

Amit Goyal and Alessio Saretto. Are equity option returns abnormal? ipca says no. *IPCA Says No (August 19, 2022)*, 2022.

Jeremiah Green, John RM Hand, and X Frank Zhang. The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, 30(12):4389–4436, 2017.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

Campbell R Harvey, Yan Liu, and Heqing Zhu. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016.

Steven L Heston, Christopher S Jones, Mehdi Khorram, Shuaiqi Li, and Haitao Mo. Seasonal momentum in option returns. *Available at SSRN 4167231*, 2022.

Alex R Horenstein, Aurelio Vasquez, and Xiao Xiao. Common factors in equity option returns. *Available at SSRN 3290363*, 2020.

Joel Horowitz, Jussi Klemelä, and Enno Mammen. Optimal estimation in additive regression models. *Bernoulli*, 12(2):271–298, 2006.

Kewei Hou, Haitao Mo, Chen Xue, and Lu Zhang. An augmented q-factor model with expected growth. *Review of Finance*, 25(1):1–41, 2021.

Jian Huang, Joel L Horowitz, and Fengrong Wei. Variable selection in nonparametric additive models. *Annals of statistics*, 38(4):2282, 2010.

Christopher S Jones. A nonlinear factor analysis of s&p 500 index option returns. *The Journal of Finance*, 61(5):2325–2363, 2006.

Christopher S Jones, Mehdi Khorram, and Haitao Mo. Momentum, reversal, and seasonality in option returns. *Available at SSRN 3705500*, 2020.

Bryan T Kelly, Diogo Palhares, and Seth Pruitt. Modeling corporate bond returns. *Available at SSRN 3720789*, 2020.

Robert Kohn and Craig F Ansley. Estimation, prediction, and interpolation for arima models with missing data. *Journal of the American statistical Association*, 81(395): 751–761, 1986.

Josef Lakonishok, Andrei Shleifer, and Robert W Vishny. Contrarian investment, extrapolation, and risk. *The journal of finance*, 49(5):1541–1578, 1994.

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

Jane A Ou and Stephen H Penman. Financial statement analysis and the prediction of stock returns. *Journal of accounting and economics*, 11(4):295–329, 1989.

Jeffrey Pontiff and Artemiza Woodgate. Share issuance and cross-sectional returns. *The Journal of Finance*, 63(2):921–945, 2008.

Scott A Richardson, Richard G Sloan, Mark T Soliman, and Irem Tuna. Accrual reliability, earnings persistence and stock prices. *Journal of accounting and economics*, 39 (3):437–485, 2005.

Aurelio Vasquez and Xiao Xiao. Default risk and option returns. *Available at SSRN 2667930*, 2020.

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

Appendix

A.1 Characteristics

1. *ag*: Asset growth, computed as the year-on-year percentage change in total assets (Cao et al., 2021).
2. *at*: Total assets.
3. *baspread*: Bid-ask spread of the option.
4. *bidask*: The ratio of the difference between the bid and ask quotes of option to the midpoint of the bid and ask quotes at the end of previous month.
5. *bm*: The natural logarithm of book equity for the fiscal year-end in a calendar year divided by market equity at the end of December of that year (Cao et al., 2021).
6. *ch*: Cash-to-assets ratio defined as the value of corporate cash holdings over the value of the firm's total assets. Cao et al. (2021).
7. *chinv*: Change in inventory.
8. *chtx*: Change in tax expense.
9. *currat*: Current assets divided by current liabilities (Ou and Penman, 1989).
10. *dolvol*: Natural log of trading volume times price per share from month $t-2$ (Chordia et al., 2001).
11. *dy*: Dividend to price.
12. *egr*: Growth in common shareholder equity.
13. *ep*: Earnings to price.
14. *eps*: Earnings per share.
15. *gma*: Gross profitability.
16. *grltnoa*: Growth in long-term net operating assets.
17. *hire*: Employee growth rate.
18. *impliedvol*: Difference between implied volatility and realized volatility of the option, where realized volatility is calculated as the historical volatility over the last trading month.

19. *interest*: The open interest for the option contract on the trading date preceding the portfolio formation date.

20. *invest*: Annual change in gross property, plant, and equipment, plus annual change in inventories, all scaled by lagged total assets (Chen and Zhang, 2010).

21. $issue_{1Y}$: The change in shares outstanding from 11 months ago as in Pontiff and Woodgate (2008).

22. $issue_{5Y}$: The 5-year real change in outstanding shares (Cao et al., 2021).

23. *ivol*: Stock return idiosyncratic volatility (Ang et al., 2006).

24. *leverage*: Computed as the sum of total debt and the par value of the preferred stock minus deferred taxes and investment tax credit, divided by market equity as in Vasquez and Xiao (2020).

25. *lgr*: Annual percent change in total liabilities (Richardson et al., 2005).

26. *maxret*: Maximum daily return.

27. *oa*: Operating accruals (Bandyopadhyay et al., 2010).

28. *pchcurrat*: Percent change in current ratio.

29. *pchgm*: Percent change in sales minus the percent change in inventory.

30. *pchsale*: Percent change in sales - the percent change in A/R.

31. *pchsaleinv*: Percent change in inventory.

32. *pm*: Sales minus costs of goods sold to sales (Freyberger et al., 2020).

33. $profit$: Profitability, calculated as earnings divided by book equity in which earnings are defined as income before extraordinary items.

34. r_{2-1}: Return from 2 to 1 months before prediction (Freyberger et al., 2020).

35. r_{6-2}: Return from 6 to 2 months before prediction (Freyberger et al., 2020).

36. r_{12-2}: Return from 12 to 2 months before prediction (Freyberger et al., 2020).

37. r_{12-7}: Return from 12 to 7 months before prediction (Freyberger et al., 2020).

38. r_{36-13}: Return from 36 to 12 months before prediction (Freyberger et al., 2020).

39. rd_{mve}: R&D to market capitalization.

40. rd_{sale}: R&D to sales.

41. *roavol*: Standard deviation for 16 quarters of income before extraordinary items divided by average total assets (Francis et al., 2004).

42. *roeq*: Return on equity.

43. *s2c*: Sales-to-cash is the ratio of net sales to Cash and Short-Term Investments (Freyberger et al., 2020).

44. *s2p*: Sales-to-price is the ratio of net sales (SALE) to the market capitalization as of December (Freyberger et al., 2020).

45. $sales_g$: Sales growth is the percentage growth rate in annual sales following Lakonishok et al. (1994) (Freyberger et al., 2020).

46. *sat*: Sales to total assets (Freyberger et al., 2020).

47. *sga2s*: SG&A to sales is the ratio of selling, general and administrative expenses to net sales.

48. *sgr*: Annual percent change in sales (Lakonishok et al., 1994).

49. *size*: The Natural Logarithm of the market value of the firm's equity (Cao et al., 2021).

50. *sratio*: The ratio of systematic volatility over total volatility. Total volatility (tvol) is the standard deviation of daily stock returns over the previous month and systematic volatility is calculated by $\sqrt{tvol^2 - ivol^2}$

51. std_{dolvol}: Monthly standard deviation of daily dollar trading volume (Chordia et al., 2001).

52. *tan*: We follow Freyberger et al. (2020) and define tangibility as (0.715 $\times$ total receivables + 0.547 $\times$ inventories + 0.535 $\times$ (property, plant and equipment) + cash and short-term investments) / total assets (at).

53. *tb*: Tax income to book income.

54. tef: Total external financing, computed as net shares issuance plus net debt issuance minus cash dividends, scaled by total assets (Cao et al., 2021).

55. *turn*: Share turnover.

56. *tvol*: Total volatility is the standard deviation of the residuals from a regression of excess returns on a constant as in Ang et al. (2006) (Freyberger et al., 2020). We use one month of daily data and require at least fifteen non-missing observations.

57. *volume*: Option trading volume on the previous trading date.

A.2 Discussion

There are several limitations when dealing with the cross-section of option returns. This section provides a extensive summary of the potential roadblocks one encounters when trying to identify characteristics with predictive power, as well as suggestions to circumvent these problems.

Data

Unlike equity stock prices, it is notoriously difficult to find a reliable database of equity option prices. As in this paper, most academic papers resort to the IvyDB OptionMetrics database. Unfortunately, it is not a database without flaws. To start, option data in OptionMetrics available to academics does not begin until 1996, whereas we have access to stock prices from 1960 onwards. Second, OptionMetrics does not provide a list of all historical option prices. For various months, numerous firms have no reported option prices listed in the OptionMetrics database despite having options listed at those times. Third, OptionMetrics applies a price correction to account for the early exercise premium of American style options. Although not hugely problematic, for options that are not deeply out of the money, this correction algorithm is mostly a black box.

OptionMetrics mostly provides us with the dependent variable y. A more pressing issue in the data-collection of cross-sectional analysis is the construction of our design matrix X. When we consider 57 characteristics, we will inevitably have observations with missing values in our design matrix. The number of missing characteristics determines the severity of the problem, but technically one missing observation introduces unnecessary bias into the model. There are several ways to approach this problem. The simplest and probably most intuitive solution is the one applied in this paper, which is to remove any y observations with missing x values. While this allows for smooth regression, it does force us to throw away a substantial portion of our option observations. Another approach could be to run the regression despite missing variables. Bennett (2001) argues that if more than 10% of data is missing, we are introducing significant bias into estimation. I would argue that the benefit of included dependent variables with missing x values, solely for the purpose of adding more data into the model, is completely overshadowed by the bias it introduces. Not to mention the effects it has on the interpretability of results. Another idea is to use some form of interpolation to fill in the missing data entries in the design matrix, although this is rarely done and almost exclusively in the case of missing y variables (Kohn and Ansley, 1986).

Models in conventional literature

This paper compares the nonparametric model with individual portfolio sorts and linear Fama-Macbeth regressions, both of which are still commonly used in the analysis of equity option returns. However, factors models are by far the preferred model of choice for most academic papers. Although factor models are notoriously bad in out-of-sample prediction. If we truly want to evaluate the efficacy of a nonparametric model it is only fair to assess out-of-sample prediction alongside the predictions of a well-known factor model. I've constrained myself to a linear model for three reasons. Because of its simplicity, ease of interpretation and because there are countless adaptations of factor models in the known literature. With each proposed method claiming exclusive upsides that only their factor structure provides. In that sense, the simple linear regression approach carries less baggage and provides a clear comparison between a nonlinear and a linear approach. Generally, linear models tend to perform comparable to factor models, but we cannot draw fair conclusions about the adaptive group LASSO without comparing it to factor models used in recent years. The performance of the adaptive group LASSO should be compared to a factor model as proposed in Büchner and Kelly (2022), Horenstein et al. (2020) or Goyal and Saretto (2022).

Runtime

I have refrained from mentioning runtime in this paper because in part it is a function of the hardware at someones disposal and their level of programming skill. However, I do find it important to mention the runtime of the nonparametric model here to put the results into perspective. A numerical solver or the algorithm of Yuan and Lin (2006) is required to find a solution to Equation (9) and (11). Matlab R2020b was used to estimate the models in this paper. Matlab is known to have a slow compiler compared to intermediate-level languages such as c++ and even compared to high-level languages such as python and R. However, writing your code in any of these languages will not solve the core problem. The runtime of any numerical solver will increase exponentially with the amount of variables it has to estimate. A huge benefit of the linear model is that with 57 characteristics, we only have to estimate 58 coefficients. In the case of the adaptive group LASSO with 20 knots we are forced to estimate $20 \times 57 = 1140$ characteristics. To put this into perspective, estimating the linear model in column 2 of Table 9 took 3 seconds, while the nonlinear model in column 1 took over 18 hours. Estimating the nonlinear model with 20 knots using a numerical solver takes between two to four days. The rate of convergence of the algorithm of Yuan and Lin (2006) is generally faster but extremely inconsistent.

Limitations of empirical analysis and suggestions

The results in section 4.5 imply that there is a nonlinear structure in the extreme negative end of delta-hedged option returns. We were unable to outperform the simple linear model in the long leg of returns. There is a strong possibility that this is caused by the unusual out-of-sample period of January 2019 to March 2021. The World Health Organization announced the global pandemic on the 30th of January 2020, which means around half of our out-of-sample period is during an exceptional crisis. It is safe to say this has significant impact on the accuracy of our model's forecasts. Two simple suggestions follow: (1) Perform out-of-sample analysis over numerous sample periods to assess robustness of empirical results. The downside is that the sample period post-2018 contains a relatively large percentage of the total number of observations. (2) A more relevant approach would be to incorporate a time-varying element into the adaptive group LASSO since we find that coefficients vary significantly over time. Our iteration of the adaptive group LASSO is unable to capture this effect. For completeness, a simulation study should be performed to not only evaluate the predictability of the model in a controlled environment, but also to make an informed choice of tuning parameters. I have only employed the Bayesian information criterion to select characteristics because of runtime complications. A model with different penalty parameters would likely select different characteristics.