The Information in Option Strike Price Introductions

Hojoon Lee *

February 27, 2026

Abstract

I study the information content of option strike price introductions. Stocks with options introduced at strike prices above the prevailing maximum outperform those with options introduced below the minimum by 4% over the following 12 months. This return spread is not explained by stock-price momentum or other stock- and option-level predictors, but is larger when new options trade actively and informed trading proxies are high. New high-leverage out-of-the-money options attract abnormal demand and earn long-run returns. Option introductions are difficult to predict using machine learning, suggesting that new options allow investors to incorporate stock-specific information through additional, cost-effective leverage.

Keywords: Asset prices, Option strike prices, Information flow, Leverage, Informed trading

*Texas Christian University, Neeley School of Business, Finance Department, email: hojoon.lee@tcu.edu. I thank Rui Albuquerque, Michele Andreolli, Simcha Barkai, Vincent Bogousslavsky, Ronnie Chen (discussant), Yong Chen, Ryan Davies, Yevgeny Frenkel, Andreas Fuster, Vincent Grégoire (discussant), Samuel Hartzmark, Mathias Hasler, Trevor Haynes, Burton Hollifield, Shuaiqi Li, Angela Ma, Andrey Malenko, Nadya Malenko, Alan Marcus, Gregor Matvos, Dmitriy Muravyev, Gideon Ozik, Paola Pederzoli (discussant), Jeffrey Pontiff, Ronnie Sadka, Daniel Smith, Tommaso Tamburelli, Annette Vissing-Jorgensen, Jiaqi Zhang, Alexei Zhdanov, participants at the FMA Derivatives & Volatility 2025, NFA 2025, FMA 2025, Boston College Brown Bag 2024, Boston College Eagles Finance Conference 2024, Florida State University, Nanyang Technological University in Singapore, Texas Christian University, and University of Cincinnati for insightful comments. All errors are my own.

Conflict-of-interest disclosure statement

Hojoon Lee

I have nothing to disclose.

1 Introduction

One of the most important questions in finance is how private information is incorporated into asset prices. While private information is known to be gradually reflected in prices (Grossman and Stiglitz (1980)), a major challenge in understanding this process is that the information set of informed traders remains unobservable. Driven by economic incentives, informed traders seek to trade in markets where their informational advantage can be maximized. Specifically, the options market may attract informed traders by offering cost-effective leverage (Black (1975)). As a result, options trading may convey incremental information about the underlying stock that is not yet reflected in stock prices.

In this paper, I propose a novel mechanism through which the options market leads the stock market: the introduction of new option strike prices. The intuition is that options introduced at strikes outside of the prevailing range offer investors cost-effective leverage, which may help incorporate information into stock prices. Consistent with this intuition, I find that stocks with new options introduced at strikes above the prevailing maximum outperform those with strikes below the prevailing minimum by at least 4% for the following 12 months. The findings suggest that stock-specific information is being incorporated into stock prices through cost-effective leverage provided by new out-of-the-money options.

To the best of my knowledge, this is the first paper to examine whether strike price introductions contain information about the underlying stock. The existing literature focused on first-time option listings for stocks that previously had no listed options and their effects on the underlying stock's price and volatility (Conrad (1989), Detemple and Jorion (1990), among others). However, this paper focuses on option strike price introductions for stocks that already have listed options. I document a channel by which information is incorporated into stock prices through the introduction of new option strike prices.

Option exchanges introduce options with new strike prices. The rulebooks of major exchanges explicitly state that they introduce new options in response to either significant stock price changes or demand from institutional, corporate, or individual investors. To explore whether such introductions are mechanically predictable versus demand-driven, I use two methods to predict future option

introductions based on stock- and option-level characteristics: panel regression and an Extreme Gradient Boosting (XGBoost, Chen and Guestrin (2016)) model. The adjusted R^2's are less than 20% for panel regressions and the out-of-sample precision rates for the XGBoost model are below 33%, suggesting that option introductions are difficult to predict using observable information alone.

To capture the direction of strike price introductions, I construct a monthly variable $INTRO_q3$, which equals 1 (-1) if the stock has relatively more options introduced above (below) than below (above) the prevailing strike price range. This summarizes the overall direction of option introductions toward upside or downside, as only about 5 stocks per month have an equal number of options introduced on both sides.

Using this measure, I document that option introductions at extreme strike prices predict future stock returns. Event-time analyses show that stocks with $INTRO_q3 = 1$ outperform those with $INTRO_q3 = -1$ by at least 4% for the following 12 months, and the return spread does not reverse for up to 24 months. The return spread increases to over 8% when newly introduced options trade on the day of listing, suggesting that the option introduction event itself, rather than how options are introduced, is what drives the predictability. The results are robust to controlling for characteristic-based benchmark portfolios (Daniel et al. (1997)). Moreover, the Sharpe ratio of the baseline portfolio using $INTRO_q3$ with a 12-month holding horizon is sizable compared to the anomaly return spreads over the same sample period (1996-2023). It ranks in the top 21% among the 180 value-weighted anomaly portfolios and in the top 32% among the 213 anomaly portfolios with a 12-month holding horizon using top minus bottom decile sorts in Chen and Zimmermann (2022).

Fama and MacBeth (1973) regressions further support the predictability of option introductions. Controlling for a broad set of stock- and option-level predictors, including momentum and implied volatility, $INTRO_q3$ has a coefficient of 0.10 (t-statistic 2.59), which translates to a 0.20% monthly return spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ over the next 12 months. A monthly long-short portfolio based on $INTRO_q3$ earns a statistically significant alpha of 0.25% (t-statistic 2.24) against the Fama and French (2018) six-factor model plus the short-term reversal factor. The monthly alpha increases to 0.55% (t-statistic 2.92) when using a subset of stocks

whose newly introduced options have positive volume on the day of introduction.

Consistent with stock-level results, option strategies based on option introductions also yield economically and statistically significant returns. When options are introduced at strike prices above the prevailing maximum, call options are high-leverage out-of-the-money contracts and put options are low-leverage in-the-money contracts, and vice versa. Based on this intuition, I construct option strategies that buy the new out-of-the-money call options for stocks with $INTRO_q3 = 1$ and buy the new out-of-the-money put options for stocks with $INTRO_q3 = -1$, respectively. Event-time analysis shows that both option strategies earn positive and statistically significant returns excess of the market for the following 12 months. Furthermore, long-maturity options perform particularly well in the long-run, suggesting that option introductions reflect long-horizon information.

To further distinguish whether option introductions reflect investor demand or enable informed trading, I examine predicted versus realized introductions from the machine learning model. True positive option introductions—those predicted by the model and subsequently realized—exhibit the largest return spread. The return spread for true positives is larger than that for false positives (4.59% versus 1.53% over 12 months), where false positives are option introductions predicted by the model but not realized. This suggests that option introduction events themselves enable investors to obtain additional leverage. Consistent with this interpretation, long-short portfolios based solely on model-implied option introductions generate a weaker return spread (2.02% over 12 months) than portfolios based on true positives. These results indicate that option introductions convey information beyond the model's predictions.

I present three pieces of evidence that are consistent with option introductions enabling leverage for informed investors.

First, analysts tend to underpredict the earnings of stocks that have options introduced at extreme strikes. The earnings forecast error spread is positive and statistically significant between stocks with relatively more options introduced above the prevailing strike price range and those with relatively more options introduced below the range. This spread is larger when restricting the sample to stocks whose newly introduced options have positive trading volume on the day of introduction. These results are consistent with analysts not fully incorporating information contained in option

introductions.

Second, newly introduced high-leverage options have a higher call-to-total open interest ratio than options with the previous maximum or minimum strike. As mentioned earlier, calls are high-leverage out-of-the-money contracts and puts are low-leverage in-the-money contracts when they are introduced above the prevailing maximum strike price. If investors seek additional leverage, newly listed options should exhibit a higher call-to-total open interest ratio than existing options with the highest leverage, since open interest reflects the number of outstanding contracts. Consistent with this hypothesis, the results show that, at the end of the month, the call-to-total open interest ratio is higher for new options introduced on the upside and lower for those introduced on the downside. For example, for the upside, the call-to-total open interest ratio is 0.89 for newly introduced options above the prevailing maximum strike price, compared with 0.85 for existing options at the previous maximum strike, implying increased demand for high-leverage options.

Finally, the results are stronger for stocks with higher proxies for informed trading: Informed trading intensity (Bogousslavsky et al. (2024)), option-to-stock volume ratio (Easley et al. (1998), Johnson and So (2012)), and idiosyncratic volatility (Ang et al. (2006)). First, I use the informed trading intensity measure from Bogousslavsky et al. (2024) that captures the institutional trading volume of a given day based on 13F filings. Next, the option-to-stock volume ratio proxies for informed trading based on the intuition that informed traders might disguise their trades in more liquid markets (Easley et al. (1998)). More informed trading is expected when the options market is relatively more liquid than the stock market. Finally, I use idiosyncratic volatility as a proxy for informed trading, based on the idea that stock prices driven more by firm-specific information provide greater opportunities for informed traders to act (Roll (1988), Stambaugh et al. (2015)). Using the three proxies, I double-sort stocks by high and low values of the informed trading proxies and by $INTRO_q3 = -1, 1$. The resulting return spread is consistently larger when proxies of informed trading are high. Daily analysis further shows that informed trading intensity is highest on the day of option introduction.

In contrast, option delistings, unlike option introductions, do not predict future stock returns, consistent with rulebooks stating that options are delisted when open interest is 0 and there are no

investor requests to retain them. Similarly, index option introductions do not predict future index returns, consistent with private information being unlikely to predict market-wide movements and with the general difficulty of predicting aggregate returns (Campbell and Thompson (2008), Welch and Goyal (2008), among others). Moreover, option straddle portfolios, consisting of both call and put options, generate positive and significant returns for $INTRO_q3 = -1$, suggesting that downward option introductions contain information about future volatility. Finally, the results are robust across various specifications using short- or long-maturity options, stocks with high or low borrowing fees, different subperiods, and alternative event-study specifications.

The rest of the paper is organized as follows. Section 2 reviews the related literature. Section 3 provides institutional details on how exchanges introduce new strike prices. Section 4 describes the data and variables used in the empirical analyses. Section 5 presents the main empirical results. Section 6 presents additional empirical results. Section 7 concludes.

2 Related Literature

The existing literature primarily focused on the impact of first-time option introductions on stocks that previously had no listed options. Theoretically, options are redundant assets that can be replicated with stocks and bonds in a frictionless market (Black and Scholes (1973)). However, markets are incomplete in the real world due to trading costs, short-sale constraints, and imperfect information (An et al. (2014)). Thus, option introductions may affect underlying stock prices. Several papers explore the consequences of first-time stock option introductions. Skinner (1989) shows that when options are first introduced to an underlying stock, its volatility decreases and volume increases. Conrad (1989) shows that stock prices permanently increase around the date of first-time option introductions. Detemple and Jorion (1990) find that option introductions are associated with increased stock price and decreased volatility, but these effects weakened over time. Detemple and Selden (1991) theoretically shows that the underlying asset price increases when options are introduced. Also, Skinner (1990) documents that less information is contained in earnings announcements after options are introduced to a stock, consistent with more private information being produced. The in-

formation content about expected returns is also explored by Lee and So (2017), where they show abnormally high analyst coverages predict firms' fundamental performance. Adding to this literature, I explore the information content of new option introductions for stocks that already have options.

Many papers provide theoretical evidence that informed traders trade in the options market before trading in the stock market. Black (1975) argues that informed investors might trade in the options market to exploit the leverage options provide. Back (1993) proposes a model that shows trading in options provides stronger signals than trading in stocks, and thus, options cannot be entirely hedged with stocks. Consistent with the literature, I show evidence of a new channel through which informed traders trade in the options market when they have private information.

The existing literature also provides empirical evidence of the options market leading the stock market. Easley et al. (1998) develop a model where buying a call or selling a put contains information about the underlying stock. Roll et al. (2010) explore the determinants of the option-to-stock volume ratio and provide evidence that it relates to trading costs, institutional holdings, and earnings announcement returns. Johnson and So (2012) show that option-to-stock volume negatively predicts future stock returns at the week-level (and reverse quickly after) because traders with negative information can trade on options more easily than short-selling the underlying stock. Hu (2014) decomposes stock imbalances into option-induced imbalances and option-independent imbalances, and shows that option-induced imbalances contain information on the underlying stock. More recently, Muravyev et al. (2022) argue that option volatility and volume predict future stock returns because they reflect stock borrowing fees. Pan and Poteshman (2006) show that put-call option volume ratios negatively predict future stock returns at the daily and weekly levels. Xing et al. (2010) documents that stock with steeper volatility smirks have low future expected returns without reversals, consistent with informed traders trading out-of-the-money put options. Cremers and Weinbaum (2010) provide evidence that based on put-call parity, weighted-average implied volatility differences between calls and puts predict future stock returns at the 1-week and 1-month level, which are not explained by short selling costs. Fodor et al. (2011) show that changes in call (put) open interest positively (negatively) predict future stock returns at the daily level. Jin et al.

(2012) provide evidence that supports these results by showing that volatility smirks have stronger predictability before firm-specific information events. Muravyev et al. (2013) use put-call parity to estimate the disagreement between implied stock prices and actual stock prices, and find that option prices do not contain information on future returns of underlying stocks using 39 stocks. An et al. (2014) show that changes in call (put) option implied volatility positively (negatively) predict future stock returns for up to 6 months. Blau et al. (2014) show that put-call ratios predict stock returns at the daily level, whereas option-to-stock volume ratios predict stock returns at the weekly and monthly levels. Bergsma et al. (2020) show that volume- and open-interest-weighted average moneyness positively predicts future stock returns at the daily level. Kang et al. (2022) show that the put-to-call ratio of out-of-the-money options using daily volume and open interest negatively predicts future stock returns at the 1-month level but reverses quickly after. In this paper, I provide suggestive evidence that informed investors with strong private information drive the introduction of options with new strike prices at the extreme.

Recent papers have documented how mispricing affects stock or option prices. Goncalves-Pinto et al. (2020) show that not only informed trading but price pressure in the stock market leads to temporary mispricing where options variables become informative. On the other hand, Eisdorfer et al. (2022) studies option mispricing where investors seem to underprice 5-week options compared to 4-week options due to inattention to maturities when rolling over monthly positions.

Finally, several studies have examined inventory risk faced by option market makers and their risk management. Muravyev (2016) provides evidence that option prices are affected more by market makers' inventory risk induced by order imbalances than asymmetric information. Hu et al. (2023) document evidence from the Korean Stock Exchange that option market makers manage risk by inventory management but rarely delta-hedge their positions, contrary to prior beliefs. The results in this paper suggest the information in the introduction of new strike prices is most likely acquired by investors instead of option market makers.

3 Institutional Details

In this section, I explore how option exchanges introduce new strike prices for individual stock options. The main goal of exchanges is to maximize trading volume in the options market. To achieve this, they follow a set of rules for adding option strike prices, which are stated almost identically across the rulebooks of major option exchanges (CBOE, NYSE, and NASDAQ). Understanding these rules helps clarify the mechanics behind option introductions.

> "Additional series of options of the same class may be opened for trading on the Exchange when the Exchange deems it necessary to maintain an orderly market, to meet customer demand or when the market price of the underlying stock moves more than five strike prices from the initial exercise price or prices. The opening of a new series of options on the Exchange will not affect any other series of options of the same class previously opened." (Rule 4.5. Series of Option Contracts Open for Trading, CBOE Rulebook)

Other statements related to adding series of options include the following:

> "The Exchange may also open additional strike prices of (Weekly, Monthly, Quarterly) Option Series that are more than 30% above or below the current price of the underlying index or security, provided that demonstrated customer interest exists for such series, as expressed by institutional, corporate, or individual customers or their brokers. Market-makers trading for their own account will not be considered when determining customer interest under this provision." (CBOE Rulebook)

The following statement explains how exchanges delist options:

> "When there is no open interest in a series, the Exchange may delist such series. Delisting shall be preceded by a notice to TPH organizations concerning the delisting." (Rule 4.4. Withdrawal of Approval of Underlying Securities, CBOE Rulebook)
>
> "...the Exchange will, on a monthly basis, review series that are outside a range of five strikes above and five strikes below the current price of the underlying ETF, and

delist series with no open interest in both the put and the call series having a: (i) strike higher than the highest strike price with open interest in the put and/or call series for a given expiration month; and (ii) strike lower than the lowest strike price with open interest in the put and/or call series for a given expiration month." (CBOE Rulebook)

"Notwithstanding the above referenced delisting policy, customer requests to add strikes and/or maintain strikes in Quarterly Options Series in ETF options in series eligible for delisting shall be granted." (CBOE Rulebook)

Based on the statements above, exchanges appear to follow a largely passive process, but there seems to be a demand component in strike price introductions and delistings. New strike prices may be introduced, or existing options may not be delisted, in response to investor demand, suggesting that these decisions are driven not only by price changes but also by investor demand. Moreover, the rulebooks explicitly state that "market-makers will not be considered when determining customer interest", partially ruling out the alternative explanation that hedging demand drives option introductions. Delistings, in contrast, seem to occur only when an option has no open interest.

Overall, the rulebooks of major exchanges suggest a demand component in the introduction of strike prices, but details on the decision process are ambiguous. Exchanges do not specify the timing of the option introductions following significant stock movements, how such movements are measured (over which time interval), or what qualifies an "orderly market". It is also unclear which timeframe exchanges use when calculating the 30% threshold relative to the current price. Furthermore, the rulebooks do not provide details on how exchanges respond to investor demand. In this paper, I show that strike price introductions contain information on the underlying stock, whereas delistings do not.

4 Data and Variables

4.1 Data

Stock returns and firm characteristics are obtained from the Center for Research in Security Prices (CRSP) covering January 1996 to December 2023. Individual stock options and their option-level characteristics are obtained from OptionMetrics covering January 1996 to August 2023. Standard risk factors and the T-bill rate are obtained from Kenneth R. French's data library[1]. Analyst earnings forecasts are obtained from the I/B/E/S Summary History dataset.

The characteristic-based benchmark portfolios are constructed following Daniel and Titman (1997) using the replication code provided by Drechsler (2024). Following the literature, I incorporate CRSP delisted returns for each stock after it is delisted. For stocks delisted from NYSE and AMEX, I set delisted returns to -35%, and for stocks delisted from NASDAQ, I set delisted returns to -55% (Shumway (1997), Shumway and Warther (1999), Chen and Zimmermann (2022)).

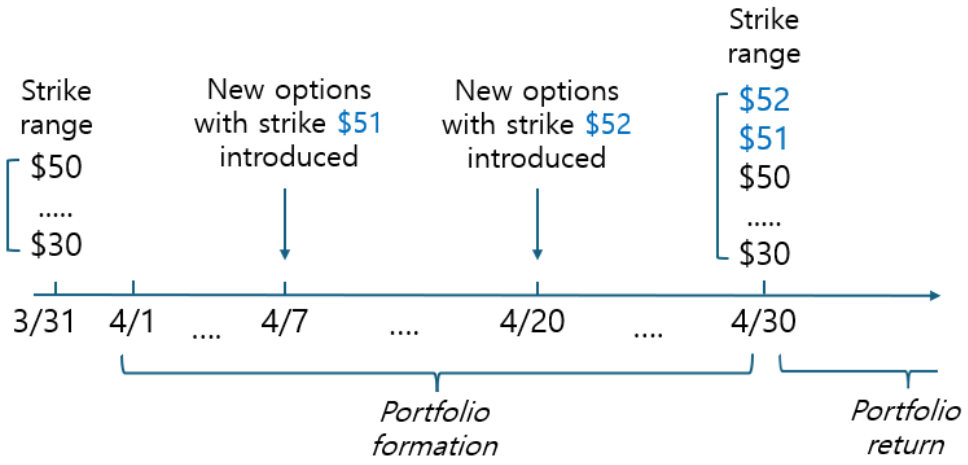
I exclude stocks with prices below $1 at the end of each month to mitigate microstructure concerns. Options are required to have standard contract sizes (multiplier 100) and strike price information available in the OptionMetrics database at the end of each month to avoid issues related to non-standard contracts and stock splits. Stocks are required to have share codes 10 or 11 and at least one listed option in the OptionMetrics, which excludes first-time option introductions. All individual stock options are American options, where holders can exercise the option at any time before expiration. All index options are European options, where holders can exercise the option only at expiration.

4.2 Main Variable: $INTRO_q3$

The main variable is constructed by counting the number of options introduced above and below the prevailing range of strike prices for a given stock at the end of each month. The goal is to capture the overall direction of option introductions. The options are counted across different option types,

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

strike prices, and expiration dates as long as they are introduced outside of the current strike price range. Note that in almost all cases, the call and put options are introduced at the same time for a given strike price.



For each stock at the end of month t, I define $INTRO_t$ as the number of options introduced during month t with strike prices above the highest strike available at the end of month $t-1$, minus the number of options introduced below the lowest strike available at the end of month $t-1$.

Using $INTRO_t$, I construct the main variable $INTRO_q3$, which sorts stocks into three groups. $INTRO_q3$ equals 1, 0, -1 if $INTRO_t$ is positive, 0, negative, respectively. For example, $INTRO_q3 = 1$ indicates that, in a given month, the stock had more options introduced above than below the prevailing strike price range at the end of the previous month. This definition accounts for stocks with options introduced both above and below the prevailing strike prices by taking the difference. If an equal number of options are introduced both above and below the prevailing strike prices, then $INTRO_q3 = 0$. On average, only about 5 stocks per month in the full sample have same number of option introductions on both sides of the prevailing strike prices. Figure C.1 illustrates how $INTRO$ and $INTRO_q3_t$ are measured.

$$
\begin{aligned}
&= 1 \quad && (INTRO_t > 0) \\
INTRO_q3_t &= 0 \quad && (INTRO_t = 0) \\
&= -1 \quad && (INTRO_t < 0)
\end{aligned}
\tag{1}
$$

Figure 1 shows the number of stocks with positive $INTRO$ ($INTRO_q3 = 1$) and negative $INTRO$ ($INTRO_q3 = -1$), respectively, in each month. Summary statistics of stocks sorted by $INTRO_q3$ are reported in Table C.1. On average, there are around 293 stocks with $INTRO_q3 = 1$ and 156 stocks with $INTRO_q3 = -1$ per month. The asymmetry may arise for the following reasons: First, aggregate stock returns are positively correlated with strike price introductions. When market returns are positive, more stocks have positive returns on average, mechanically increasing the number of options introduced above the prevailing maximum strike price, and vice versa. For example, many stocks had options introduced below the prevailing minimum strike price during market declines, such as the 2008 Global Financial Crisis and 2020 COVID crisis. Second, strike prices are bounded below at zero but unbounded above. When the stock's price is low and existing strike prices are already near 0, the number of options that can be introduced below the minimum strike price is limited.

5 Empirical Results

5.1 Predicting Option Strike Price Introductions

In this section, I explore whether option introductions are predictable. If option introductions in the next month are predictable by using information from the current month, such as past returns and volume, this would suggest that exchanges largely make passive decisions to introduce options. Conversely, if option introductions are not fully explained by stock or option characteristics, this would suggest a discretionary component, reflecting potential investor demand. The degree of predictability may provide insight about the relative importance of passive and discretionary factors.

I use two different methods to examine whether option introductions are predictable: panel regressions and a machine learning algorithm called Extreme Gradient Boosting (XGBoost). While panel regression results are more straightforward and easier to interpret than those from XGBoost, they are less effective in capturing nonlinearities and interactions across variables. In each method, I separately predict whether or not the stock is $INTRO_q3 = 1$ and whether or not the stock is $INTRO_q3 = -1$, respectively. This allows me to cleanly analyze each direction of option introductions, especially when price changes are likely to be the largest driver of introductions.

First, I run panel regressions with stock and month fixed effects to explore whether option introductions are predictable within a given stock at a given month. The standard errors are clustered at the stock level to account for consecutive option introductions within a given stock.

The panel regressions with two different dependent variables are expressed as follows:

$$added_above_{i,t} = \alpha_i + \gamma_t + stock_variables_{i,t-1} + option_variables_{i,t-1} + e_{i,t} \tag{2}$$

$$added_below_{i,t} = \alpha_i + \gamma_t + stock_variables_{i,t-1} + option_variables_{i,t-1} + e_{i,t} \tag{3}$$

The dependent variable $added_above_{i,t}$ equals 1 if $INTRO_q3_t = 1$ and 0 otherwise and $added_below_{i,t}$ equals 1 if $INTRO_q3_t = -1$ and 0 otherwise. The stock and month fixed effects are denoted as α_i and γ_t, respectively. The independent variables include 1-month return, 11-month return skipping the most recent month, maximum strike price to stock price ratio, minimum strike price to stock ratio, and change in out-of-the-money call and put option implied volatility. The full set of variables used for predicting option introductions are described in Appendix A.

Column 1 and 2 of Panel A in Table 1 show that the within R^2 are 6.0% and 9.8%, respectively, suggesting it is difficult to predict whether a given stock will have options introduced in either direction. While there is no clear benchmark to compare these numbers with, the stock and option variables that are most likely to be considered by exchanges when they decide to introduce options explain less than 10% of the variation in option introductions. The results are consistent with the demand component of option introductions being sizable, although it is likely an overstatement to claim that it accounts for over 90%.

Next, I train an XGBoost model (Chen and Guestrin (2016)) to predict option introductions. The main purpose of training a machine learning model is to capture potential nonlinearities and interactions between variables, addressing the limitation of linear regressions. The decision to use the XGBoost algorithm among others, and how it is trained is detailed in Appendix B.

The XGBoost model is trained on two separate models that predicts 'Added below' and 'Added above', respectively, with the same independent variables in panel regressions. The dependent variable 'Added below (above)' is an indicator variable that equals 1 if $INTRO_q3 = -1$ ($INTRO_q3 = 1$) and 0 otherwise. The XGBoost model generates estimates of option introduction probabilities between 0 and 1. The probabilities are converted to binary variables 1 and 0 based on whether it is higher or lower than 0.5. I focus on two metrics to evaluate the model's performance: precision and recall. Precision is defined as the number of true positives divided by the sum of true positives and false positives, whereas recall is defined as the number of true positives divided by the total number of actual positives. These metrics take into account that the model could achieve high accuracy in predicting option introductions simply because the unconditional probabilities are 12% and 6% for a stock being classified as $INTRO_q3 = 1$ and $INTRO_q3 = -1$, respectively. A naive model that always predicts no introductions would achieve an accuracy of 88% and 94%.

Panel B of Table 1 shows that the precision for predicting 'Added below' and 'Added above' are both less than 33% out-of-sample. The precision is lower for the model predicting $I(INTRO_q3 = -1)$ than for the model predicting $I(INTRO_q3 = 1)$. Recall is over 85% for both models. These results imply that the model predicts a high number of option introductions but less than 33% are correct, and correctly predicts 85% of the actual option introductions.

Overall, the results suggest that predicting which stocks will have options introduced in the following month is difficult using information in the current month. Although we do not observe the exact criteria used by exchanges, the results consistently indicate that option introductions are not easily predictable, leaving room for investor demand to influence them.

5.2 Single-sorted Stock Portfolios

In this section, I test whether option introductions outside of the prevailing strike price range predict future stock returns. If new options were introduced purely mechanically by exchanges, we would expect little or no predictability, as their introduction would be anticipated.

The main result of this paper can be summarized by Figure 2. The portfolio with $INTRO_q3 = 1$ significantly outperforms the portfolio with $INTRO_q3 = -1$ for up to 12 months after portfolio formation. The return difference does not reverse up to 24 months, implying it is less likely to be driven by short-term hedging activities but more likely to reflect information.

Table 2 reports event-time cumulative stock portfolio returns (in %) up to 12 months after portfolio formation sorted by $INTRO_q3$. The cumulative returns in excess of the market return are negative for $INTRO_q3 = -1$ and positive $INTRO_q3 = 1$. The event-time cumulative return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is 4.39% (t-statistic 4.51) 12 months after portfolio formation, although it is 0.51% (t-statistic 1.73) and insignificant in the first month. Both long and short portfolios contribute to the return spread, with a stronger magnitude observed for $INTRO_q3 = -1$. When using returns in excess of the characteristic-based benchmark portfolios (Daniel et al. (1997)), the cumulative returns are reduced to 2.38% (t-statistic 3.98) but remain statistically significant.

The magnitude of the return spreads is nontrivial. The Sharpe ratio of the return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is in the top 21% against the 180 value-weighted long-short anomaly portfolios and in the top 32% against the 213 anomaly portfolios using a 12-month holding horizon, based on the data from Chen and Zimmermann (2022).

When new options are introduced, in almost all cases, the calls and puts are introduced at the same time. If the options are introduced above the prevailing maximum strike price, the call option is the high-leverage out-of-the-money option and the put option is the low-leverage in-the money option, and vice versa if the options are introduced below the prevailing minimum strike price. If new out-of-the-money options are traded immediately after their introduction, this may reveal actual, realized investor demand for leverage. Consistent with this intuition, the return spread increases to 8.40% when considering only out-of-the-money options that have positive volume on

introduction day. This filter strengthens the magnitude of the return spread, although it reduces the average number of monthly stocks included in $INTRO_q3 = -1$ and $INTRO_q3 = 1$ to 28 and 72, down from 156 and 293.

Several factors may explain why the return spread is weak in the first month. First, the delayed stock price response might be consistent with private information slowly being incorporated through future news (Grossman and Stiglitz (1980)). In Section 5.7.1, evidence supports that analysts might be underpredicting earnings for stocks with options introduced outside of the prevailing strike range. Second, it may be partly driven by short-term reversals associated with hedging activities following option introductions. Table C.2 reports significantly positive return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$ on the introduction day, but also a partial reversal in subsequent days. This is also illustrated in Figure 3. The event-time portfolios exclude option introduction days but capture short-term reversals, which persists for at least three days and may reduce first-month returns. The short-term reversal likely reflects hedging activity associated with newly introduced options. For example, when newly introduced call options are sold, purchasing the underlying stocks can serve as a hedge, creating temporary, non-informative price pressure that subsequently reverses. Finally, option introductions are noisy signals, as they also include passive introductions. Although using new out-of-the-money options with positive trading volume on introduction day lead to larger return spreads, it remains difficult to accurately identify which options attract informed demand.

Overall, the results are consistent with option introductions with extreme strike prices containing information about the underlying stock. The results are stronger when options are traded immediately after introduction, which reflects realized investor demand. Given that option exchanges are unlikely to have superior information on multiple stocks at the same time, and their primary objective is to maximize option volume, the existence of a return spread is consistent with newly introduced options containing information about future stock returns.

5.3 Single-sorted Option Portfolios

In this section, I explore whether option portfolios formed from newly introduced options with extreme strike prices generate profits and aligned with the returns of corresponding stock portfolios.

Recall that out-of-the-money options provide additional leverage over in-the-money options for a given strike price. Based on this intuition, two option portfolios are formed at the end of each month: one that buys newly introduced out-of-the-money call options with strike prices above the prevailing maximum strike price, and one that buys newly introduced out-of-the-money put options with strike prices below the prevailing minimum strike price. These portfolios capture investors' demand for higher leverage through newly introduced options.

Table 3 shows the event-time cumulative returns for buying new out-of-the-money put options for stocks with $INTRO_q3 = -1$ and call options for stocks with $INTRO_q3 = 1$, respectively. Both portfolios exhibit positive and statistically significant returns up to 12 months after portfolio formation. Notably, the magnitude of the option returns is more than 10 times larger than the stock returns, consistent with options providing greater leverage over stocks. The results are based on assuming an effective spread of 50%, where an option is bought half-way from the bid towards the bid-ask midpoint and sold half-way from the ask towards the bid-ask midpoint. This may be a relatively conservative assumption, as Muravyev and Pearson (2020) report that the effective spread of out-of-the-money options for an S&P500 stock is less than 10% and might be even smaller in practice.

Table 3 reports event-time cumulative returns of portfolios consisting of options with short and long times to maturity. The rationale behind the maturity split is that if newly introduced options have long maturities, the predictability of returns is likely to be long-term, whereas short-maturity options are expected to exhibit short-term predictability. The newly introduced options are divided into short-term and long-term maturities at 100 days, as the median time to maturity is approximately 98 days. Short-maturity portfolios generate positive short-term returns (20.13% and 5.13% in the first month for out-of-the-money puts and calls, respectively), although these are statistically insignificant, whereas the long-maturity portfolios show negative short-term returns but significantly positive long-term returns (60.55% and 47.19% in the sixth month for out-of-the-money puts and calls, respectively).

However, it is important to note that the options chosen by informed traders are not directly observable. As a result, it is unclear whether they trade short-maturity options that require fre-

quent rolling or long-maturity options to exploit long-term information. Short-maturity options require frequent rolling, which can increase costs from bid-ask spreads and theta decay, but they provide immediate leverage when the underlying stock moves. In contrast, long-maturity options are less sensitive to theta decay, but they typically provide lower realized leverage compared to short-maturity options when the underlying stock moves (Frazzini and Pedersen (2022)). Overall, if some informed traders partially adjust their portfolios by purchasing newly introduced options, the above results can reveal these choices.

A natural question is whether the out-of-the-money options offer greater leverage than existing options and whether this comes at a liquidity cost. Table C.3 compares the month-end leverage and bid-ask spreads of newly introduced out-of-the-money options and existing options, averaged across months. The out-of-the-money puts and calls offer 27% and 19% greater leverage than existing options, although their bid-ask spreads are larger than existing options. Despite the leverage-liquidity trade off, Table 3 shows that option strategies might generate positive profits, assuming 50% effective spreads.

5.4 Fama-MacBeth Regressions

In this section, I explore whether $INTRO_q3$ is explained by stock and option characteristics proposed in the literature as predictors of future stock returns, using Fama and MacBeth (1973) regressions. Despite the low correlation between $INTRO_q3$ and existing variables reported in Table C.4, Fama and MacBeth (1973) regressions test whether $INTRO_q3$ contributes to explaining the cross-section of expected stock returns after controlling for these variables.

Table 4 shows Fama and MacBeth (1973) regressions of 1-month and 12-month returns, respectively, on $INTRO_q3$ and stock characteristics. Each month, future 1-month (or 12-month average) are regressed on $INTRO_q3$ and stock characteristics in the cross-section. The reported coefficients are the time-series averages of these monthly cross-sectional regressions, along with their corresponding t-statistics.

Table 4 shows that when the dependent variable is the future 1-month return, $INTRO_q3$ has a positive but insignificant coefficient. However, when the dependent variable is the future

12-month average return, $INTRO_q3$ has a positive and statistically significant coefficient of 0.10 (t-statistic 2.59) and significantly explains the cross-section of stock returns after controlling for stock- and option-level variables. This is consistent with $INTRO_q3$'s long-term predictability observed in Section 5.2 and Section 5.3. $INTRO_q3$ has the largest coefficient among all variables in the regression. The coefficient of 0.10 translates into an average monthly return spread of 0.20% (0.10 $\times$ 2) over the next 12 months between stocks with $INTRO_q3 = 1$ and those with $INTRO_q3 = -1$.

While the set of stock and option variables in Table 4 does not cover all the variables documented in the literature, it includes those that are likely to be relevant for explaining the variation in option introductions. For example, the out-of-the-money put-call volume ratio (Kang et al. (2022)) captures how much volume the out-of-the-money call options had over the past month relative to out-of-the-money put options, which relates to the open interest or volume spillovers that may lead to new strike price introductions. Overall, the result suggests that this variable is unlikely to explain option introductions.

5.5 Time-series Regressions

In this section, I construct monthly portfolios based on option introductions and examine their risk exposures to the Fama-French six-factor portfolios, along with the short-term reversal factor. I consider two time-series portfolios: one that sorts stocks by $INTRO_q3 = -1, 0, 1$, and another that also sorts by $INTRO_q3 = -1, 0, 1$ but only includes option introductions that have positive trading volume on the day of introduction. The first portfolio includes all option introductions, whereas the second portfolio focuses on options with realized investor demand.

Table 5 shows monthly regressions of portfolio returns with 12-month holding horizons on the six-factor returns of Fama and French (2018) plus the short-term reversal factor. The first three columns report returns excess of the 1-month T-bill rate, the alpha against the Fama and French (2018) and Fama and French (2018) plus the short-term reversal factor, respectively. The next seven columns report coefficient estimates from regressing monthly portfolio returns on the factor returns. In a given month, all portfolios are formed by holding the 12 portfolios constructed in the current and previous 11 months, and rolled forward each month by adding new portfolios and

closing those that were included 12 months earlier (Jegadeesh and Titman (1993)). The 12-month holding periods are consistent with the long-term predictability of option introductions presented in Section 5.2.

The monthly alpha of the portfolio spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$ are reported in the fourth row of Table 5. It generates an alpha of 0.25% (t-statistic 2.24) per month against the Fama and French (2018) plus the short-term reversal factor. This alpha increases to 0.55% (t-statistic 2.92) when considering only options with positive volume on the introduction day. Although both portfolio returns load positively on momentum and negatively on short-term reversal, the significant alphas suggest that the returns are driven by information beyond these factors.

5.6 Option Introductions and Investor Demand

In this section, I revisit the XGBoost prediction model trained in Section 5.1 and explore whether new options are introduced in response to demand or whether their introductions attract demand. I examine the event-time cumulative returns for portfolios formed from the subset of stocks identified by the model. I focus on two groups of stocks: true positives and false positives. True positives consist of stocks where the model predicted an option introduction and an actual option introduction occurred. False positives consist of stocks where the model predicted an option introduction but no actual option introduction occurred.

Table 6 reports event-time cumulative return spreads between $INTRO_q3 = 1$ and $INTRO_q3 = -1$ up to 12 months after portfolio formation. The true positive stocks generate the largest return spread with a magnitude 4.59% at 12 months. In contrast, the false positives generate smaller spreads than true positives. This implies that option introductions events themselves enable information to be incorporated into stock prices. Consistent with this interpretation, long-short portfolios based solely on model-implied option introductions generate a weaker return spread of (2.02% over 12 months) than portfolios based on true positives. These results indicate that option introductions convey information beyond the model's predictions.

5.7 Evidence of Informed Trading

In this section, I provide suggestive empirical evidence that the return spread between stocks with options introduced above and below the prevailing strike price is driven by informed trading. I present three pieces of evidence consistent with informed trading: earnings forecast errors, call-to-total open interest ratio, and portfolios double-sorted on proxies for informed trading.

5.7.1 Analyst Earnings Forecast Errors

In this section, I explore whether analysts incorporate information contained in option introductions into their earnings forecasts. Analysts are information producers where they aim to learn all available information about the target firm (Givoly and Lakonishok (1979), among others). However, if the information from new option introductions is not considered, it might not be immediately incorporated into the earnings forecasts. Therefore, I test whether analysts incorporate information contained in option introductions.

Table 7 shows the earnings forecast errors for the portfolios sorted by $INTRO_q3$, as well as for the subset of stocks with positive trading volume in the newly introduced out-of-the-money options. The earnings forecast error is positive (negative) and statistically significant for the portfolio of stocks with $INTRO_q3 = 1$ ($INTRO_q3 = -1$). The earnings forecast error spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is positive and statistically significant, and is larger in magnitude among stocks with positive trading volume in newly introduced out-of-the-money options. The magnitude of earnings forecast errors is larger for the short-leg than the long-leg portfolio, consistent with the baseline results reported in Table 2. These results are consistent with analysts not fully incorporating the information contained in option introductions.

5.7.2 Call-to-total Open Interest Ratio after Option Introductions

In this section, I explore whether the high-leverage options attract greater investor demand than low-leverage options following their introduction. Specifically, I compare the call-to-total open interest ratio between newly introduced options and existing options. Table 8 reports call-to-total open

interest ratios of newly introduced options above and below the existing strike range, along with those of existing options at the previous maximum and minimum strike prices. This focuses on whether newly introduced, high-leverage options attract greater demand than existing options.

I first aggregate the call and put open interest across all newly introduced options outside of the existing strike range at the end of each month, and then calculate the call-to-total open interest ratio as the ratio of call open interest to the sum of call and put open interest. Table 8 reports that the call-to-total open interest ratio of options introduced below the existing strike range is 0.38, compared with 0.41 for existing options with previous minimum strikes. This implies the newly introduced, high-leverage out-of-the-money put options exhibit higher open interest than call options. Conversely, the call-to-total open interest ratio of options introduced above the existing strike range is 0.89, compared with 0.85 for existing options with previous maximum strikes. In this case, out-of-the-money call options exhibit higher open interest than put options.

5.7.3 Double-sorted Portfolios using Proxies for Informed Trading

In this section, I examine whether the option introductions are associated with proxies for informed trading. The analysis focuses on three proxies: Informed Trading Intensity (ITI, Bogousslavsky et al. (2024)), option-to-stock volume ratio (Easley et al. (1998), Johnson and So (2012)), and idiosyncratic volatility (Ang et al. (2006)).

First, the ITI is a firm-day measure predicting the intensity of institutional trading, trained on Schedule 13D filings. For each firm-month, ITI is averaged across the firm's daily observations. Second, the option-to-stock volume ratio proxies for informed trading based on the intuition that informed traders might disguise their trades in more liquid markets (Easley et al. (1998)). More informed trading is expected when the options market is relatively more liquid than the stock market. Finally, I use idiosyncratic volatility as a proxy for informed trading based on the idea that stock prices driven more by firm-specific information provide greater opportunities for informed traders to act (Roll (1988), Stambaugh et al. (2015)). Monthly idiosyncratic volatility is measured as the standard deviation of residuals from regressing daily returns on the Fama and French (1993) 3-factor model. More details on how variables are constructed is shown in Appendix A.

Stocks are first split into high and low groups based on the median value of each proxy, and then independently, stocks are sorted by $INTRO_q3$. The event-time cumulative return spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is reported separately for the high- and low-proxy groups. The results are reported in Table 9. For all three proxies, the return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is larger when the proxy is high than when it is low. The results are consistent with option introductions associated with informed trading.

Furthermore, I use the firm-day ITI measure to analyze daily ITI around option introduction dates. I estimate the following regression model:

$$ITI_{i,t} = \alpha_i + \gamma_t + \sum_{d=-5}^{10} \beta_d D_{i,t}^d + \epsilon_{i,t} \tag{4}$$

where $ITI_{i,t}$ is the daily ITI for each stock i at day t. Day -9 to -6 serves as the reference period. Stock and day fixed effects are included as α_i and γ_t. β_d is the coefficient of interest that represent deviations of ITI from the reference period. $D_{i,t}^d$ is an event-time indicator that equals 1 if date t corresponds to relative day d for stock i. The regression is estimated through previous day's market capitalization. The results are robust to using OLS.

The estimated β_d's are shown in Figure 4. The coefficients sharply increase around option introduction days for both stocks with $INTRO_q3 = -1$ and $INTRO_q3 = 1$. In both cases, the largest coefficient is on the option introduction day. Based on undocumented summary statistics, around 18% of option introductions occur on the first trading day after the third Friday of the month, suggesting that elevated ITI near the third Friday might be related to option expiration dates. But ITI remains persistent for a few days after introduction, particularly for stocks with $INTRO_q3 = -1$, implying that introductions might be associated with informed trading.

6 Additional Tests

6.1 Option delistings

In this section, I examine whether option delistings, as opposed to option introductions, contain information about the underlying stock. The rulebooks of option exchanges state that options are delisted when they have zero open interest and there are no investor requests to retain them. This implies delistings have a demand component, but it is likely stale and less likely to be reflected in trades.

To test this hypothesis, stocks are sorted by $DELIST_q3 = -1, 1$ that represent stocks with options delisted above and below the current strike prices over the past month. $DELIST_q3 = 1$ includes stocks with options delisted above the current month's maximum strike price; $DELIST_q3 = -1$ includes those below the current month's minimum strike price. The spread is the long-short return spread between stocks with $DELIST_q3 = -1$ and $DELIST_q3 = 1$ to align the direction with $INTRO_q3$.

Consistent with this notion, the results reported in Table 10 show that option delistings do not predict future stock returns. The cumulative return spreads is 0.96% (t-statistic 1.33) at 12 months following portfolio formation. Furthermore, option delistings might be associated with option introductions on the other side of the strike range. Considering only delistings without corresponding at the opposite extreme, the return spread between delistings above and below the prevailing strike price range is closer to zero at 0.42% (t-statistic 0.55). These results suggest option delistings are more likely to be driven by mechanical factors than investor demand, unlike option introductions.

6.2 Index Options

In this section, I explore whether option introductions in index options contain information about the underlying index. The literature documents that predicting 1-month market returns is notoriously difficult (Welch and Goyal (2008), Campbell and Thompson (2008), among others). Also, it is unlikely that informed traders have private information about market-wide movements. If option

introductions contain information about the future movements of underlying index, they should predict future index returns. To test this hypothesis, I regress future 1-month index returns on $INTRO_q3$ and $INTRO$ variables for each index in the time-series, following Pan and Poteshman (2006). I consider the following indexes: Dow Jones Industrial Average, NASDAQ 100, S&P 500, S&P 100, S&P Midcap 400, S&P Smallcap 600, Russell 2000, and AMEX, based on their data availability in OptionMetrics.

Table 11 shows the results of time-series regressions of future 1-month index returns on current month's $INTRO_q3$ and $INTRO$ for each corresponding index. The first and second set of results show that $INTRO_q3$ and $INTRO$, which incorporates the number of options, do not predict future index returns. All coefficients are insignificant with t-statistics less than 2. These results suggest that index option introductions are less likely to convey information about the future index movements, whereas individual stock option introductions contain stock-specific information.

6.3 Information on Volatility

In this section, I explore whether information in option introductions are related to volatility. One major difference between stocks and options is that options may increase in value with rising stock volatility, even when the stock price remains unchanged. To explore whether volatility information is contained in newly introduced options, I construct a straddle portfolio that consists of pairs of calls and puts for a given stock, month, strike price, and expiration date, and then average returns across stock-months.

Table 12 shows the event-time cumulative returns for straddle portfolios using on options from stocks sorted by $INTRO_q3$. The straddle portfolio for $INTRO_q3 = -1$ is positive and statistically significant, with a 12-month cumulative return of 36.31%, whereas the portfolio for $INTRO_q3 = 1$ is positive but statistically insignificant, with a 12-month cumulative return of 7.32%. The results are consistent with option introductions below the prevailing minimum strike containing information about the underlying stock's volatility, whereas those above the prevailing maximum contain little such information.

6.4 Robustness Tests

In this section, I conduct robustness tests for baseline results reported in Table 2.

6.4.1 Option Maturities

Table C.5 shows the cumulative event-time return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$, 12 months after portfolio formation, is robust to using options with short or long maturities: less than 100 days and more or equal to 100 days (the median maturity is 98 days for new options). In both cases, the cumulative returns are positive and statistically significant (4.11% and 4.92% for short- and long-maturity options), although long-maturity options generate slightly larger return spreads.

As mentioned in Section 5.3, options chosen by informed traders are not directly observable. As a result, it is unclear whether they trade short-maturity options that require frequent rolling or long-maturity options to exploit long-term information. The observed similarity in return spreads could be explained by the fact that options with varying maturities are typically introduced at the same time.

6.4.2 Stock Borrowing Fees

Table C.5 shows that the return spread from option introductions is robust to using stocks with high and low borrowing fees. The cumulative event-time return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$, 12 months after portfolio formation, is positive and statistically significant whether or not stocks have estimated borrowing fees above or below 1% (Muravyev et al. (2022)). The return spread 12 months after portfolio formation is 4.39% for easy-to-borrow stocks and 4.23% for hard-to-borrow stocks. This is consistent with option introductions unlikely to be related to short-selling constraints. The stock borrowing fees are estimated following Muravyev et al. (2022). See Appendix A for details.

6.4.3 Subperiod Analysis

Table C.5 shows that the return spread from option introductions is robust across different subperiods. The cumulative event-time return spread between $INTRO_q3 = 1$ and $INTRO_q3 = -1$, 12 months after portfolio formation, is positive and statistically significant in both of the subperiods: 1996-2009 and 2009-2023. The split is motived by Bondarenko and Muravyev (2022), where they show that the predictability of option variables on future stock returns deteriorates after October 2009, following a massive crackdown on insider trading through options. I split the portfolio formation months into two periods at September 2009. The results imply that return spreads associated with option introductions are unlikely to be driven by illegal insider information.

6.4.4 Event Study Specifications

Table C.5 shows that the baseline results in Table 2 are robust to different specifications of event studies. First, the results are robust to using Newey and West (1987) standard errors with twelve lags for event-time cumulative returns. This mitigates the concern that autocorrelation in overlapping cumulative returns across events is driving the results. Second, the results are robust to using only stocks that had no options introduced in the prior 12 months. This mitigates the concern that option introductions and returns might not be independent events (Lyon et al. (1999)). For example, if a given firm has multiple option introductions within the past 12 months, the subsequent 12-month cumulative returns after each option introduction might be correlated due to overlapping months. Overall, the baseline results are robust to different event-time specifications.

6.4.5 Number of Options Introduced

The number of options introduced above or below the existing strike range, across different strike prices and maturities, may reflect the magnitude of investor demand. Based on this idea, I split $INTRO_q3 = -1$ and $INTRO_q3 = 1$ into two groups each, respectively, by the median number of options introduced.

Stocks are first split into $INTRO_q3 = 1$ and $INTRO_q3 = -1$. Within each group, the stocks are

split into high and low groups based on the median number of options introduced. Table C.6 shows the event-time cumulative return spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ for the high- and low-number groups. The return spread is larger for stocks with a high number of options introduced than for those with a low number of options introduced, with event-time 12-month cumulative returns of 6.50% and 2.48%, respectively.

7 Conclusion

In this paper, I propose a novel mechanism through which the options market leads the stock market: the introduction of new option strike prices. Exchanges introduce options with new strike prices for individual stock options not only in response to price changes but to investor demand. If introductions were purely passive, they would not contain any information beyond what is reflected in stock prices. However, stocks with options introduced above the prevailing maximum strike price outperform those with options introduced below the prevailing minimum strike price by at least 4% over the following 12 months. This return spread is not explained by stock-price momentum or other stock- and option-level predictors, but is stronger when new options are traded actively and when proxies for informed trading are high. Analysts tend to underpredict earnings for the firms with option introductions, and the results are consistent with new high-leverage out-of-the-money options attracting abnormal demand. Overall, these findings suggest that options introduced at strikes outside of the prevailing range offer investors cost-effective leverage, which may facilitate stock-specific information to be incorporated into stock prices.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

An, Byeong-Je, Andrew Ang, Turan G Bali, and Nusret Cakici. The joint cross section of stocks and options. *The Journal of Finance*, 69(5):2279–2337, 2014.

Ang, Andrew, 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.

Back, Kerry. Asymmetric information and options. *The Review of Financial Studies*, 6(3):435–472, 1993.

Bergsma, Kelley, Vivien Csapi, Dean Diavatopoulos, and Andy Fodor. Show me the money: Option moneyness concentration and future stock returns. *Journal of Futures Markets*, 40(5):761–775, 2020.

Black, Fischer. Fact and fantasy in the use of options. *Financial Analysts Journal*, 31(4):36–41, 1975.

Black, Fischer and Myron Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

Blau, Benjamin M, Nga Nguyen, and Ryan J Whitby. The information content of option ratios. *Journal of Banking & Finance*, 43:179–187, 2014.

Bogousslavsky, Vincent, Vyacheslav Fos, and Dmitriy Muravyev. Informed trading intensity. *The Journal of Finance*, 79(2):903–948, 2024.

Bondarenko, Oleg and Dmitriy Muravyev. What information do informed traders use? *Working Paper*, 2022.

Campbell, John Y and Samuel B Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531, 2008.

Chen, Andrew Y. and Tom Zimmermann. Open source cross-sectional asset pricing. *Critical Finance Review*, 27(2):207–264, 2022.

Chen, Tianqi and Carlos Guestrin. Xgboost: A scalable tree boosting system. *Cornell University*, 2016.

Conrad, Jennifer. The price effect of option introduction. *The Journal of Finance*, 44(2):487–498, 1989.

Cremers, Martijn and David Weinbaum. Deviations from put-call parity and stock return predictability. *Journal of Financial and Quantitative Analysis*, 45(2):335–367, 2010.

Daniel, Kent and Sheridan Titman. Evidence on the characteristics of cross sectional variation in stock returns. *The Journal of Finance*, 52(1):1–33, 1997.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers. Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, 52(3):1035–1058, 1997.

Detemple, Jerome and Philippe Jorion. Option listing and stock returns: An empirical analysis. *Journal of Banking & Finance*, 14(4):781–801, 1990.

Detemple, Jerome and Larry Selden. A general equilibrium analysis of option and stock market interactions. *International Economic Review*, pages 279–303, 1991.

Drechsler, Qingyi (Freda). Python programs for empirical finance. https://www.fredasongdrechsler.com, 2024.

Eames, Michael J and Steven M Glover. Earnings predictability and the direction of analysts' earnings forecast errors. *The accounting review*, 78(3):707–724, 2003.

Easley, David, Maureen O'hara, and Pulle Subrahmanya Srinivas. Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2):431–465, 1998.

Eisdorfer, Assaf, Ronnie Sadka, and Alexei Zhdanov. Maturity driven mispricing of options. *Journal of Financial and Quantitative Analysis*, 57(2):514–542, 2022.

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

Fama, Eugene F and Kenneth R French. Choosing factors. *Journal of Financial Economics*, 128(2): 234–252, 2018.

Fama, Eugene F and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.

Fodor, Andy, Kevin Krieger, and James S Doran. Do option open-interest changes foreshadow future equity returns? *Financial Markets and Portfolio Management*, 25:265–280, 2011.

Frazzini, Andrea and Lasse Heje Pedersen. Embedded leverage. *The Review of Asset Pricing Studies*, 12(1):1–52, 2022.

Givoly, Dan and Josef Lakonishok. The information content of financial analysts' forecasts of earnings: Some evidence on semi-strong inefficiency. *Journal of Accounting and Economics*, 1(3): 165–185, 1979.

Goncalves-Pinto, Luis, Bruce D Grundy, Allaudeen Hameed, Thijs Heijdenvan der , and Yichao Zhu. Why do option prices predict stock returns? the role of price pressure in the stock market. *Management Science*, 66(9):3903–3926, 2020.

Grossman, Sanford J and Joseph E Stiglitz. On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408, 1980.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, and others. The elements of statistical learning, 2009.

Hu, Jianfeng. Does option trading convey stock price information? *Journal of Financial Economics*, 111(3):625–645, 2014.

Hu, Jianfeng, Antonia Kirilova, and Dmitriy Muravyev. Options market makers. *Available at SSRN*, 2023.

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

Jin, Wen, Joshua Livnat, and Yuan Zhang. Option prices leading equity prices: Do option traders have an information advantage? *Journal of Accounting Research*, 50(2):401–432, 2012.

Johnson, Travis L and Eric C So. The option to stock volume ratio and future returns. *Journal of Financial Economics*, 106(2):262–286, 2012.

Kang, Chang-Mo, Donghyun Kim, Junyong Kim, and Geul Lee. Informed trading of out-of-the-money options and market efficiency. *Journal of Financial Research*, 45(2):247–279, 2022.

Lee, Charles MC and Eric C So. Uncovering expected returns: Information in analyst coverage proxies. *Journal of financial economics*, 124(2):331–348, 2017.

Lyon, John D, Brad M Barber, and Chih-Ling Tsai. Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance*, 54(1):165–201, 1999.

Muravyev, Dmitriy. Order flow and expected option returns. *The Journal of Finance*, 71(2):673–708, 2016.

Muravyev, Dmitriy and Neil D Pearson. Options trading costs are lower than you think. *The Review of Financial Studies*, 33(11):4973–5014, 2020.

Muravyev, Dmitriy, Neil D Pearson, and John Paul Broussard. Is there price discovery in equity options? *Journal of Financial Economics*, 107(2):259–283, 2013.

Muravyev, Dmitriy, Neil D Pearson, and Joshua M Pollet. Why does options market information predict stock returns. *Available at SSRN*, 2022.

Newey, Whitney K. and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

Pan, Jun and Allen M Poteshman. The information in option volume for future stock prices. *The Review of Financial Studies*, 19(3):871–908, 2006.

Richardson, Scott A, Siew Hong Teoh, and Peter D Wysocki. Tracking analysts' forecasts over the annual earnings horizon: Are analysts' forecasts optimistic or pessimistic? *Available at SSRN 168191*, 1999.

Roll, Richard. Presidential address: R2. *Journal of Finance*, 43(2):51–566, 1988.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam. O/s: The relative trading activity in options and stock. *Journal of Financial Economics*, 96(1):1–17, 2010.

Shumway, Tyler. The delisting bias in crsp data. *The Journal of Finance*, 52(1):327–340, 1997.

Shumway, Tyler and Vincent A Warther. The delisting bias in crsp's nasdaq data and its implications for the size effect. *The Journal of Finance*, 54(6):2361–2379, 1999.

Skinner, Douglas J. Options markets and stock return volatility. *Journal of Financial Economics*, 23 (1):61–78, 1989.

Skinner, Douglas J. Options markets and the information content of accounting earnings releases. *Journal of Accounting and Economics*, 13(3):191–211, 1990.

Stambaugh, Robert F, Jianfeng Yu, and Yu Yuan. Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948, 2015.

Welch, Ivo and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508, 2008.

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao. What does the individual option volatility smirk tell us about future equity returns? *Journal of Financial and Quantitative Analysis*, 45(3):641–662, 2010.

Table 1: Predicting Option Introductions

This table shows the results of predicting next month's option introduction using information from the current month. Panel A shows results from panel regressions of indicator variables for option introductions on stock and option characteristics. Coefficient estimates for a subset of characteristics are reported. The dependent variable, $I(INTRO_q3 = 1)$ (respectively, $I(INTRO_q3 = -1)$), equals 1 if $INTRO_q3 = 1$ (respectively, $INTRO_q3 = -1$) and 0 otherwise. Panel B shows the out-of-sample precision and recall of the XGBoost algorithm in predicting next month's option introduction based on information from the current month. Precision is defined as the number of true positives divided by the sum of true positives and false positives, whereas recall defined as the number of true positives divided by the total number of actual positives. The full list of independent variables used in both specifications are reported in Appendix A. The sample period is 1996-2023.

(a) Panel A: Panel regression

	Dependent variables	
	Added below	Added above
Independent variables	$I(INTRO_q3 = -1)$	$I(INTRO_q3 = 1)$
ret_1M	-0.12***	0.26***
	(-28.27)	(31.92)
MOM	0.00+	0.02***
	(1.89)	(13.65)
Max strike to price ratio	-0.01***	0.01***
	(-13.36)	(10.24)
Min strike to price ratio	0.09***	0.04***
	(7.09)	(5.09)
ΔImplied volatility (OTM put)	-0.01***	0.00***
	(-7.05)	(3.80)
ΔImplied volatility (OTM call)	0.00	-0.01***
	(0.16)	(-4.64)
Stock variables	Yes	Yes
Option variables	Yes	Yes
Stock fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
Errors clustered by stock	Yes	Yes
Adj. R^2	14.7%	16.0%
Within R^2	6.0%	9.8%

(b) Panel B: XGBoost algorithm

	$I(INTRO_q3 = -1)$	$I(INTRO_q3 = 1)$
Precision	32.7%	19.0%
Recall	85.4%	86.5%

Table 2: Event-Time Cumulative Portfolio Returns for $INTRO_q3 = -1, 1$, and spread

This table shows the event-time cumulative stock portfolio returns (in %) up to 12 months after portfolio formation sorted by $INTRO_q3$. Month=0 is the option introduction month. At the end of each month, stocks are sorted into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price. The '1 - (-1)' column denotes the long-short return spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$. Single-leg returns are reported in excess of the market return. The 'DGTW (1997)' returns are excess returns relative to the 125 characteristic-based benchmark portfolios, five-way sorted by market cap, book-to-market, and momentum, respectively (Daniel et al. (1997)). Market cap is sorted using NYSE breakpoints; book-to-market is sorted relative to the Fama-French 48-industry average; momentum is the past 11-month return, skipping the most recent month. Event-time cumulative returns of a long-short portfolio using only out-of-the-money options with positive trading volume on the day of the introduction are also reported. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

	All options								Options with volume>0 on introduction day			
	$INTRO_q3$				Long-short		DGTW (1997)		Long-short		DGTW (1997)	
Month	-1	t-stat	1	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat
1	-0.37	(-1.85)	0.14	(0.97)	0.51	(1.73)	0.16	(0.82)	1.14	(2.20)	0.47	(1.11)
2	-0.56	(-2.10)	0.29	(1.53)	0.85	(2.19)	0.26	(1.03)	2.00	(3.02)	1.14	(2.27)
3	-1.15	(-3.70)	0.29	(1.25)	1.45	(3.20)	0.46	(1.58)	3.26	(4.19)	1.83	(3.11)
4	-1.31	(-3.69)	0.41	(1.40)	1.72	(3.23)	0.60	(1.71)	4.35	(4.71)	2.31	(3.49)
5	-1.65	(-4.15)	0.51	(1.52)	2.16	(3.59)	0.89	(2.31)	4.95	(4.83)	2.58	(3.58)
6	-1.90	(-4.25)	0.77	(2.19)	2.67	(4.21)	1.15	(2.78)	6.27	(5.52)	3.62	(4.60)
7	-2.35	(-4.61)	0.93	(2.45)	3.29	(4.67)	1.61	(3.64)	6.62	(5.39)	4.03	(4.52)
8	-2.59	(-4.60)	1.06	(2.65)	3.64	(4.82)	1.96	(4.19)	7.17	(5.42)	4.63	(4.97)
9	-2.89	(-4.81)	1.17	(2.70)	4.06	(4.99)	2.24	(4.42)	7.59	(5.26)	4.71	(4.52)
10	-2.93	(-4.65)	1.29	(2.87)	4.23	(5.01)	2.32	(4.38)	7.96	(5.30)	5.08	(4.50)
11	-2.92	(-4.36)	1.48	(3.12)	4.40	(4.89)	2.39	(4.26)	8.57	(5.47)	5.19	(4.54)
12	-3.03	(-4.22)	1.36	(2.67)	4.39	(4.51)	2.38	(3.98)	8.40	(4.99)	4.91	(4.06)

Table 3: Event-Time Option Strategy Returns for $INTRO_q3 = -1, 1$

This table shows the event-time cumulative option returns (in %) up to 12 months after portfolio formation sorted by $INTRO_q3$. At the end of each month, out-of-the-money puts (calls) newly introduced below (above) the existing strike range, are purchased for stocks with $INTRO_q3 = -1$ ($INTRO_q3 = 1$), and sold at the end of the month preceding expiration. An effective spread of 50% is applied to the bid and ask prices. Option returns are reported in excess of the market return. Option portfolios split by a time to maturity of 100 days are also reported. The option returns first are equally weighted at the stock-month level, then value-weighted by the underlying stock's market capitalization at the month level. At the end of each month, options are required to have expiration dates beyond the end of the following month. The option returns are winsorized at the 1% and 99% levels each month. Options that are delisted or have missing prices prior to expiration are assigned a return of -100% in the following month. T-statistics are reported in parentheses. The sample period is 1996-2023.

	New out-of-the-money option introductions				Time to maturity < 100 days				Time to maturity ≥ 100 days			
	$INTRO_q3 = -1$		$INTRO_q3 = 1$		$INTRO_q3 = -1$		$INTRO_q3 = 1$		$INTRO_q3 = -1$		$INTRO_q3 = 1$	
Month	OTM puts	t-stat	OTM calls	t-stat	OTM puts	t-stat	OTM calls	t-stat	OTM puts	t-stat	OTM calls	t-stat
1	11.51	(0.67)	-1.10	(-0.53)	20.13	(1.04)	5.13	(1.03)	-6.86	(-1.99)	-3.00	(-1.80)
2	16.50	(0.95)	7.10	(2.01)	17.94	(0.89)	5.80	(0.77)	4.23	(0.94)	8.25	(2.97)
3	24.58	(1.40)	13.52	(3.34)	17.22	(0.85)	0.07	(0.01)	16.35	(2.95)	17.06	(4.99)
4	33.00	(1.86)	20.91	(3.97)					27.43	(4.32)	26.79	(5.20)
5	44.26	(2.45)	25.75	(4.41)					42.11	(5.24)	33.12	(5.73)
6	58.34	(3.06)	37.98	(4.20)					60.55	(5.58)	47.19	(5.11)
7	67.85	(3.43)	42.55	(4.53)					74.86	(5.95)	54.13	(5.56)
8	71.25	(3.49)	43.92	(4.51)					80.25	(5.50)	56.85	(5.51)
9	71.45	(3.47)	49.27	(4.82)					81.26	(5.43)	63.59	(5.79)
10	73.34	(3.50)	49.79	(4.82)					83.40	(5.45)	64.47	(5.81)
11	75.93	(3.59)	50.72	(4.80)					86.84	(5.54)	65.98	(5.81)
12	80.02	(3.63)	50.70	(4.72)					91.66	(5.39)	66.57	(5.77)

Table 4: Fama-MacBeth Regressions

This table shows Fama and MacBeth (1973) regressions of 1-month and 12-month returns, respectively, on $INTRO_q3$ and stock characteristics. Each month, future 1-month returns (or future 12-month average returns) are regressed on $INTRO_q3$ and stock characteristics in the cross-section. The average coefficients of each cross-sectional regression are then reported, along with their corresponding t-statistics across the time series in parentheses. When 12-month returns are used as the dependent variable, t-statistics are calculated using Newey and West (1987) standard errors with 12 lags. Stock-level variables from stock characteristics include: 1-month return ('ret_1M'), log market cap ('mcap_log'), log book-to-market ratio ('beme_log'), 11-month return skipping the most recent month ('ret_MOM'), 1-month idiosyncratic volatility ('ret_ivol'), profitability ('profitability'), and investment ('investment'). Stock-level variables from option characteristics include: 1-month change in call volatility surface ('ΔCVOL', An et al. (2014)), 1-month change in put volatility surface ('ΔPVOL', An et al. (2014)), 1-month option-to-stock volume ratio ('O/S ratio', Johnson and So (2012)), 1-month average implied volatility spread ('volatility_spread', Cremers and Weinbaum (2010)), 1-month average implied volatility skew ('volatility_skew', Xing et al. (2010)), 1-month percentage change in call-put open interest ratio ('Δcall_put_open_interest', Fodor et al. (2011)), out-of-the-money put-call open interest ratio ('OTM_put_call_OI', Kang et al. (2022)), out-of-the-money put-call volume ratio ('OTM_put_call_volume', Kang et al. (2022)), open interest-weighted average moneyness ('avg_moneyness_OI', Bergsma et al. (2020)), volume-weighted average moneyness ('avg_moneyness_volume', Bergsma et al. (2020)). Details on constructing the option variables are presented in Appendix A. Variables not expressed in percentage changes are standardized by subtracting their cross-sectional mean and dividing by their cross-sectional standard deviation. If variables are unavailable for a given stock, 0 is assigned to preserve the sample size. The average adjusted-R^2's ('Avg. Adj. R^2'), average number of stocks ('Avg. N. stocks'), and the number of cross-sectional regressions ('Months') are reported. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. The sample period is 1996-2023.

	Stocks with options					
variables		ret 1M			ret 12M	
$INTRO_q3$	0.20	0.06	0.04	0.12	0.12	0.10
	(1.23)	(0.73)	(0.55)	(1.66)	(2.68)	(2.59)
ret_1M		-0.23	-0.22		-0.02	-0.02
		(-2.80)	(-2.85)		(-0.56)	(-0.47)
mcap_log		-0.06	-0.06		-0.03	-0.02
		(-1.33)	(-1.23)		(-0.39)	(-0.32)
beme_log		0.07	0.06		0.09	0.09
		(0.86)	(0.82)		(0.59)	(0.59)
ret_MOM		0.15	0.15		-0.08	-0.07
		(1.50)	(1.58)		(-0.89)	(-0.89)
ret_ivol		-0.26	-0.26		-0.04	-0.05
		(-2.48)	(-2.54)		(-0.30)	(-0.40)
profitability		0.08	0.08		0.06	0.06
		(2.25)	(2.23)		(1.47)	(1.54)
investment		-0.06	-0.06		-0.08	-0.08
		(-2.01)	(-1.88)		(-3.91)	(-3.82)
ΔCVOL			0.19			0.01
			(5.41)			(1.90)
ΔPVOL			-0.21			-0.02
			(-6.55)			(-3.34)
O/S ratio			-0.11			-0.03
			(-3.29)			(-1.06)
volatility_spread			0.07			0.08
			(2.01)			(2.45)
volatility_skew			-0.02			-0.02
			(-0.79)			(-1.25)
Δcall_put_open_interest			0.03			0.01
			(1.99)			(1.40)
OTM_put_call_OI			-0.02			-0.01
			(-0.72)			(-1.08)
OTM_put_call_volume			-0.03			0.03
			(-0.68)			(1.80)
avg_moneyness_OI			-0.14			-0.02
			(-1.66)			(-0.46)
avg_moneyness_volume			0.13			0.03
			(1.76)			(0.96)
(intercept)	0.85	1.84	1.75	0.92	1.39	1.30
	(2.41)	(2.11)	(2.03)	(1.80)	(1.01)	(0.97)
Avg. Adj. R^2	0.5%	5.9%	6.6%	0.3%	6.9%	7.6%
Avg. N. stocks	2,134	2,134	2,134	2,005	2,005	2,005
Months	330	330	330	323	323	323

Table 5: Time-series Regressions

This table shows monthly regressions of portfolio returns with 12-month holding horizons on the six-factor returns of Fama and French (2018) plus the short-term reversal factor. At the end of each month, stocks are sorted into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price. The '1 - (-1)' row denotes the long-short portfolio of stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$. The '1 - (-1) volume > 0' row denotes the long-short portfolio only using options with positive trading volume on the day of the introduction. In a given month, all portfolios are formed by holding the 12 portfolios constructed in the current and previous 11 months, and rolled forward each month by adding new portfolios and closing those that were included 12 months earlier (Jegadeesh and Titman (1993)). Returns are reported in excess of the 1-month T-bill rate ('excess'), and alphas are estimated against the six-factor model and the six-factor model plus the short-term reversal. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

INTRO_q3	excess	FF6	FF6 + STREV	Mkt_RF	SMB	HML	RMW	CMA	MOM	STREV
-1	0.43	-0.19	-0.21	0.99	0.12	0.01	-0.01	-0.01	-0.21	0.09
	(1.42)	(-2.44)	(-2.75)	(52.87)	(4.65)	(0.43)	(-0.29)	(-0.17)	(-13.50)	(4.22)
0	0.66	-0.02	-0.02	0.99	-0.05	0.03	0.05	0.02	-0.04	0.01
	(2.60)	(-1.17)	(-1.32)	(265.26)	(-8.84)	(5.01)	(7.33)	(2.76)	(-12.67)	(2.54)
1	0.78	0.04	0.05	1.04	0.03	-0.05	-0.01	-0.08	0.23	-0.05
	(2.94)	(0.58)	(0.73)	(65.90)	(1.45)	(-1.99)	(-0.49)	(-2.27)	(16.91)	(-2.56)
1 - (-1)	0.35	0.23	0.25	0.05	-0.09	-0.07	0.00	-0.08	0.44	-0.13
	(2.05)	(1.95)	(2.24)	(1.86)	(-2.28)	(-1.41)	(-0.08)	(-1.16)	(18.51)	(-4.25)
1 - (-1) volume>0	0.68	0.51	0.55	0.07	-0.26	0.03	0.13	-0.21	0.54	-0.19
	(2.79)	(2.66)	(2.92)	(1.48)	(-4.01)	(0.37)	(1.53)	(-1.91)	(13.69)	(-3.66)

Table 6: Event-Time Returns based on Portfolios sorted by Prediction Model

This table shows the event-time cumulative long-short returns (in %) up to 12 months after portfolio formation sorted by $INTRO_q3$, based on predictions and outcomes of option introductions using the XGBoost prediction model from Section 5.1. True positives consist of stocks where the model predicted an option introduction and an actual option introduction occurred. False positives consist of stocks where the model predicted an option introduction but no actual option introduction occurred. True positives plus false positives consist of stocks where the model predicted an option introduction ex-ante. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

	Model-based option introductions					
	True positive		False positive		True positive + False positive	
Month	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat
1	0.40	(1.20)	0.17	(0.93)	0.23	(1.14)
2	0.78	(1.82)	0.52	(1.97)	0.62	(2.26)
3	1.45	(2.89)	0.78	(2.49)	0.90	(2.75)
4	1.95	(3.41)	1.03	(2.81)	1.16	(3.09)
5	2.54	(3.86)	1.16	(2.81)	1.39	(3.28)
6	3.08	(4.36)	1.17	(2.57)	1.51	(3.28)
7	3.69	(4.60)	1.14	(2.35)	1.62	(3.23)
8	3.87	(4.47)	1.40	(2.72)	1.84	(3.43)
9	4.45	(4.84)	1.41	(2.61)	1.95	(3.48)
10	4.67	(4.92)	1.61	(2.87)	2.17	(3.71)
11	4.63	(4.57)	1.56	(2.70)	2.11	(3.47)
12	4.59	(4.33)	1.53	(2.55)	2.02	(3.18)

Table 7: Earnings Forecast Errors

This table shows the earnings forecast errors for portfolios sorted by $INTRO_q3$. At the end of each month, stocks are sorted into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price. The '1 - (-1)' row denotes the spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$. At the stock-month level, earnings forecast errors are calculated as actual earnings per share in the next yearly earnings announcement minus the median forecasted earnings per share, divided by the previous month-end's stock price. The forecast errors are then value-weighted by market capitalization at the month-level, and the t-statistics are computed across months and reported in parenthesis. Forecast errors with an absolute value greater than 20% are dropped to mitigate the effect of outliers (Richardson et al. (1999), Eames and Glover (2003)). A positive (negative) forecast error is consistent with analysts underreacting to positive (negative) information. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. The sample period is 1996-2023.

	Stock Earnings Forecast Errors	
$INTRO_q3$	All option introductions	OTM options with volume > 0 on introduction day
-1	-0.04 (-2.23)	-0.09 (-2.62)
1	0.01 (2.80)	0.02 (2.48)
1 - (-1)	0.05 (2.63)	0.10 (2.92)

Table 8: Call-to-total Open Interest Ratio of New Options

This table compares the call-to-total open interest ratio between newly introduced options and existing options. At the end of each month, the call-to-total open interest ratio is calculated for the options that are newly introduced outside of the existing strike range, as well as for options at the previous month's maximum and minimum strike prices. For each strike price, the total open interest is defined as the sum of call and put option open interest across all expiration dates. The call-to-total open interest ratio is first averaged within each month and then averaged across months. The differences are reported for each $INTRO_q3 = -1, 1$, and corresponding t-statistics are computed across months. The sample period is 1996-2023.

	Call/(Call+Put) Open interest ratio		
$INTRO_q3$	New options (outside of strike range)	Existing options (previous max or min strike price)	Difference
-1	0.38	0.41	-0.03 (-11.68)
1	0.89	0.85	0.04 (32.40)

Table 9: Event-Time Cumulative Portfolio Returns Double-sorted by Proxies for Informed Trading

This table shows the event-time cumulative stock portfolio returns (in %) up to 12 months after portfolio formation double-sorted by $INTRO_q3$ and proxies for informed trading: Informed Trading Intensity (ITI, Bogousslavsky et al. (2024)), option-to-stock volume ratio (Easley et al. (1998), Johnson and So (2012)), and idiosyncratic volatility (Ang et al. (2006)). Firm-day ITI is averaged within each firm-month. The option-to-stock volume ratio is calculated as the ratio of the monthly sum of daily option volume to the monthly sum of daily stock volume. Idiosyncratic volatility is estimated as the standard deviation of residuals from within-month regressions of daily returns on the Fama and French (1993) 3-factor model. Stocks are first split into high and low groups based on the median value of each proxy. Independently, stocks are sorted by $INTRO_q3$. The event-time cumulative return spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is reported separately for the high- and low- proxy groups. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

	ITI				O/S ratio				IVOL			
	Low		High		Low		High		Low		High	
Month	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat	1 - (-1)	t-stat
1	0.61	(1.83)	0.22	(0.63)	-0.10	(-0.32)	0.75	(2.06)	0.22	(0.76)	1.23	(2.78)
2	0.52	(1.12)	0.54	(1.20)	0.02	(0.06)	1.04	(2.23)	0.55	(1.41)	1.96	(3.11)
3	0.86	(1.74)	0.95	(1.71)	0.61	(1.22)	1.82	(3.33)	0.99	(2.34)	3.14	(3.97)
4	0.93	(1.62)	1.43	(2.21)	0.55	(0.98)	2.39	(3.76)	1.05	(2.09)	4.11	(4.56)
5	1.26	(1.83)	2.32	(3.07)	1.16	(1.82)	2.88	(4.02)	1.33	(2.34)	5.52	(5.46)
6	1.61	(2.19)	3.02	(3.89)	1.57	(2.27)	3.40	(4.40)	1.58	(2.59)	7.08	(6.72)
7	1.90	(2.39)	3.86	(4.71)	1.91	(2.69)	4.27	(4.88)	2.13	(3.14)	7.97	(7.01)
8	2.11	(2.47)	4.14	(4.75)	2.21	(2.94)	4.65	(5.00)	2.45	(3.35)	8.21	(7.04)
9	2.21	(2.40)	4.54	(4.84)	2.55	(3.18)	5.15	(5.24)	2.69	(3.41)	9.27	(7.34)
10	2.31	(2.39)	4.56	(4.79)	3.06	(3.65)	5.13	(5.01)	2.74	(3.31)	9.53	(7.43)
11	2.21	(2.18)	4.78	(4.68)	3.20	(3.56)	5.37	(4.88)	2.84	(3.20)	9.78	(7.10)
12	2.27	(2.09)	4.71	(4.24)	3.28	(3.32)	5.37	(4.55)	2.85	(2.99)	9.69	(6.66)

Table 10: Event-Time Returns for $DELIST_q3 = -1, 1$

This table shows the event-time cumulative stock portfolio returns (in %) up to 12 months after portfolio formation sorted by $DELIST_q3$. At the end of each month, stocks are sorted into three groups: $DELIST_q3 = -1, 0, 1$. $DELIST_q3 = -1$ includes stocks with options delisted below the current month's maximum strike price; $DELIST_q3 = 1$ includes those above the current month's minimum strike price. The 'spread' column denotes the long-short return spread between stocks with $DELIST_q3 = -1$ and $DELIST_q3 = 1$, with the spread aligned to match the direction with $INTRO_q3$. The spread using only stocks with delistings without introductions is also reported. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

	Delistings		Delistings w/o introductions	
Month	spread	t-stat	spread	t-stat
1	-0.08	(-0.44)	-0.08	(-0.39)
2	0.03	(0.09)	0.05	(0.17)
3	0.28	(0.85)	0.27	(0.79)
4	0.38	(0.96)	0.34	(0.82)
5	0.36	(0.80)	0.09	(0.19)
6	0.59	(1.29)	0.13	(0.27)
7	0.66	(1.30)	0.05	(0.09)
8	0.80	(1.44)	0.21	(0.35)
9	0.95	(1.58)	0.41	(0.65)
10	0.95	(1.51)	0.59	(0.89)
11	0.70	(1.02)	0.22	(0.30)
12	0.96	(1.33)	0.42	(0.55)

Table 11: Index Options

This table shows the coefficient estimates of time-series regressions of next month's index returns on current month's $INTRO_q3$ and $INTRO$ for each corresponding index. The following indexes are considered: Dow Jones Industrial Average (1997-10 to 2023-08), NASDAQ 100 (1996-01 to 2023-08), S&P 500 (1996-01 to 2023-08), S&P 100 (1996-01 to 2023-08), S&P Midcap 400 (1996-01 to 2018-06), S&P Smallcap 600 (1996-01 to 2012-11), Russell 2000 (1996-01 to 2023-08), and AMEX (1996-01 to 2018-06). The ticker, coefficient estimate, t-statistic, R^2 (in %), and the number of months are reported. The number of months varies based on the data availability from OptionMetrics for each index. The sample period is 1996-2023.

		$INTRO_q3$			$INTRO$			
Index	Ticker	Estimate	t-stat	R^2	Estimate	t-stat	R^2	N
Dow Jones	DJX	-0.30	(-0.36)	0.04%	0.00	(0.06)	0.00%	311
NASDAQ 100	NDX	0.82	(1.06)	0.34%	0.00	(-0.14)	0.01%	332
S&P 500	SPX	-0.12	(-0.18)	0.01%	-0.01	(-0.78)	0.18%	332
S&P 100	OEX	-0.74	(-1.49)	0.67%	0.00	(-0.06)	0.00%	332
S&P Midcap 400	MID	-0.41	(-0.57)	0.12%	0.02	(1.34)	0.67%	269
S&P Smallcap 600	SML	-0.32	(-0.47)	0.11%	0.05	(1.53)	1.16%	202
Russell 2000	RUT	-0.21	(-0.30)	0.03%	0.01	(0.80)	0.19%	332
AMEX	XMI	0.16	(0.20)	0.01%	0.01	(0.45)	0.08%	259

Table 12: Option Straddle Portfolios

This table shows the event-time cumulative returns for purchasing both newly introduced call and put options for stocks sorted by $INTRO_q3$. At the end of each month, call and put options are both purchased for stocks with $INTRO_q3 = -1$ and $INTRO_q3 = 1$, and sold at the end of the month preceding expiration. Options are required to be listed in pairs—calls and puts—for each stock, month, strike price, and expiration date. An effective spread of 50% is applied to the bid and ask prices. Option returns are reported in excess of the market return. The option returns first are equally weighted at the stock-month level, then value-weighted by the underlying stock's market capitalization at the month level. At the end of each month, options are required to have expiration dates beyond the end of the following month. The option returns are winsorized at the 1% and 99% levels each month. T-statistics are reported in parentheses. The sample period is 1996-2023.

	Straddle (call + put) portfolios			
Month	-1	t-stat	1	t-stat
1	5.75	(0.63)	-2.84	(-3.01)
2	8.05	(0.87)	-0.75	(-0.44)
3	11.43	(1.22)	0.78	(0.40)
4	15.57	(1.65)	3.44	(1.32)
5	20.93	(2.19)	4.10	(1.44)
6	28.05	(2.78)	7.98	(1.74)
7	31.85	(3.07)	8.12	(1.70)
8	33.08	(3.10)	6.92	(1.40)
9	32.91	(3.05)	8.82	(1.72)
10	33.59	(3.07)	8.24	(1.59)
11	34.72	(3.15)	8.08	(1.53)
12	36.31	(3.17)	7.32	(1.35)

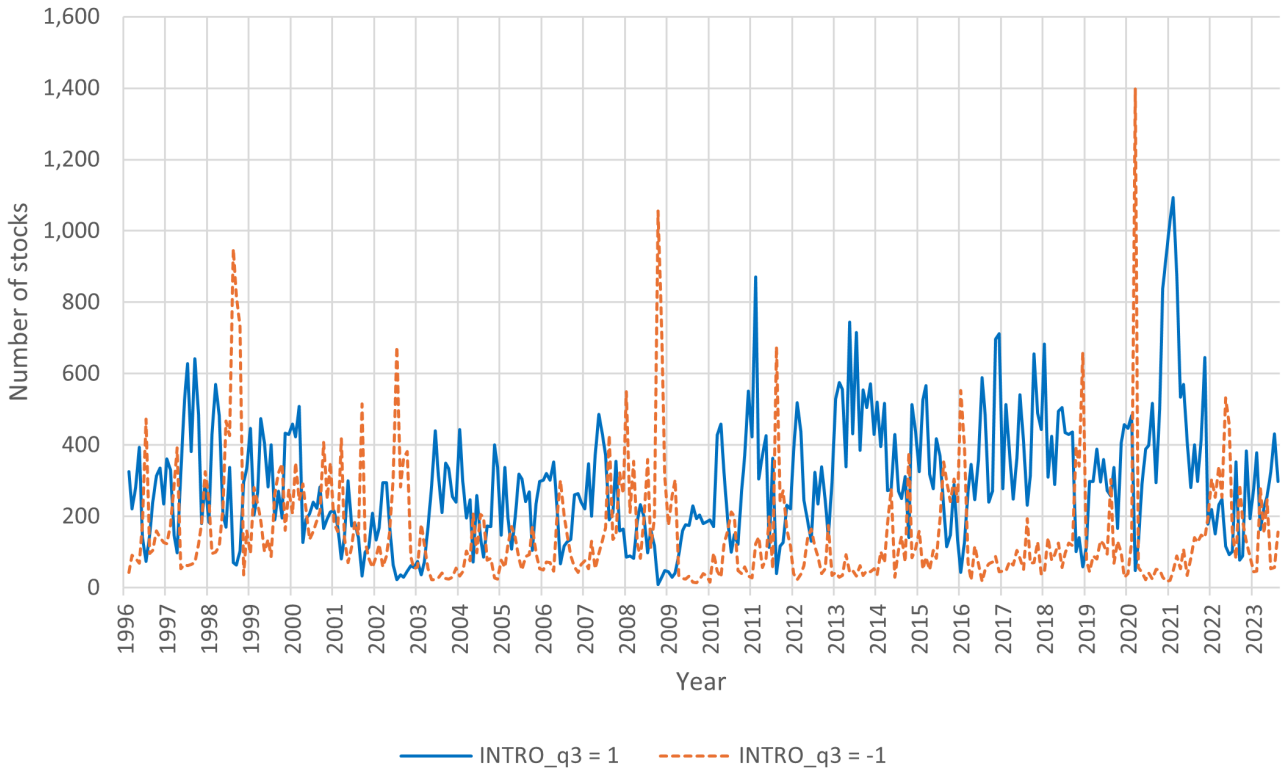


Figure 1: Number of Stocks with Option Introductions

This figure shows the number of stocks with $INTRO_q3 = -1$ and $INTRO_q3 = 1$ in each. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price. The sample period is 1996-2023.

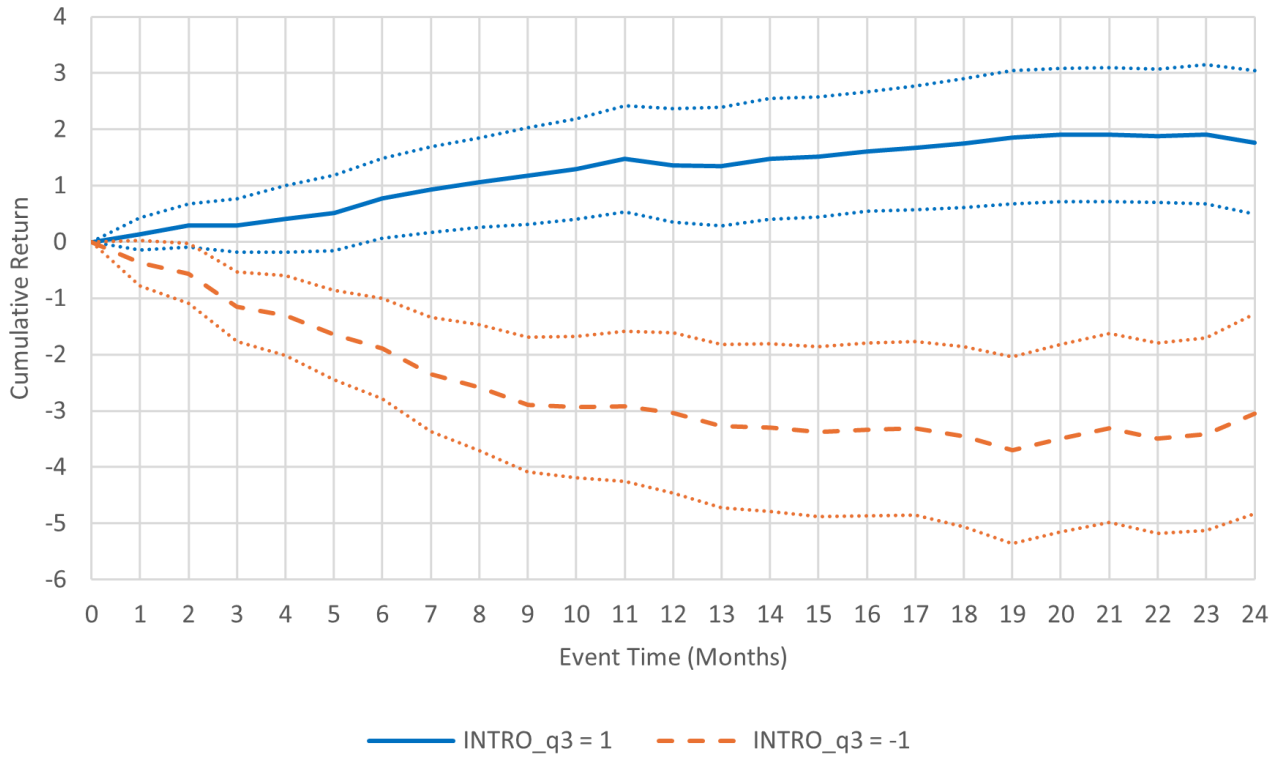


Figure 2: Event-Time Stock Portfolio Returns (Monthly)

This figure shows the event-time cumulative stock portfolio returns (in %) up to 12 months after portfolio formation sorted by $INTRO_q3$. Month=0 is the option introduction month. At the end of each month, stocks are sorted into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price. The solid and dashed lines represent cumulative returns following portfolio formation. The dotted lines represent 95% confidence levels of the cumulative returns. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. The sample period is 1996-2023.

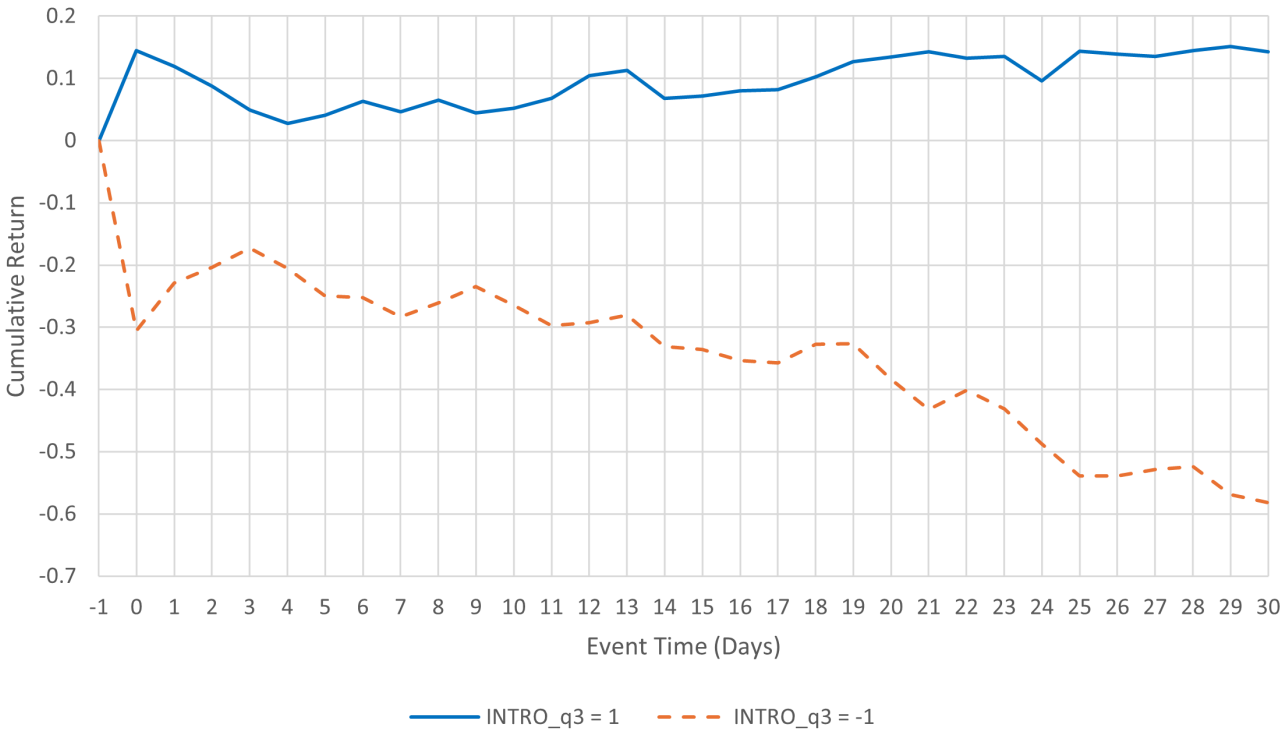


Figure 3: Event-Time Stock Portfolio Returns (Daily)

This figure shows the event-time cumulative stock portfolio returns (in %) up to 30 days after portfolio formation sorted by $INTRO_q3$. Day=0 is the option introduction day. On the introduction day, stocks are sorted into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price. The solid and dashed lines represent cumulative returns following portfolio formation. Portfolios are value-weighted by each stock's market capitalization on the previous trading day. The sample period is 1996-2023.

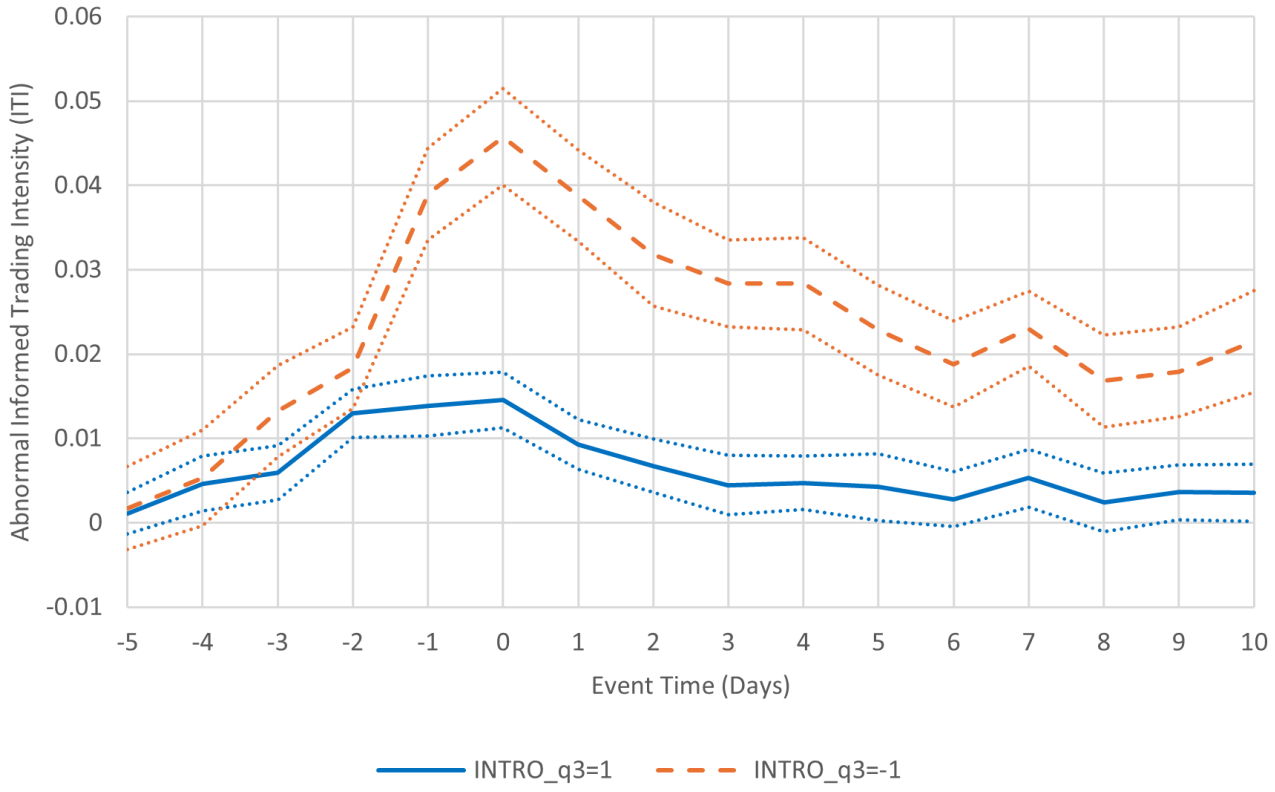


Figure 4: Event-Time Informed Trading Intensity (Daily)

This figure shows the coefficients and their 95% confidence intervals of daily Informed Trading Intensity (ITI, Bogousslavsky et al. (2024)) regressed on event-day dummies from day -5 to day 10 around the day of option introductions for $INTRO_q3 = 1$ and $INTRO_q3 = -1$, respectively. Day=0 is the option introduction day. The regressions include stock and day fixed effects, with standard errors clustered at the stock level, and weighted by the stock's market capitalization at the end of the previous day. On the introduction day, stocks are sorted into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ includes stocks with options introduced above the previous month's maximum strike price; $INTRO_q3 = -1$ includes those below the previous month's minimum strike price.The sample period is 1996-2023.

Internet Appendix

Appendix A Variable Descriptions

A.1 Stock and option variables used for predicting option introductions

The independent variables used for predicting option introductions are listed as the following. All variables are measured at the end of each month, and all changes are computed at the monthly frequency.

- Stock variables: 1-month return ('ret_1M'), 11-month return skipping 1-month ('MOM'), monthly standard deviation of daily returns, price level, turnover, change in turnover, maximum daily return within a month, bid-ask spread, change in bid-ask spread, monthly volume, percent change in monthly volume.
- Option variables: monthly open interest, change in monthly open interest, bid-ask spread, change in bid-ask spread, implied volatility, change in implied volatility, monthly volume, change in monthly volume.
  - Each option variable is divided into six components based on option type (call vs. put) and moneyness (in-the-money, at-the-money, and out-of-the-money). Each component is equal-weighted at the stock-month level.
- Variables related to option strike prices: current number of strike prices, maximum strike price, minimum strike price, maximum strike price-to-stock price ratio, minimum strike price-to-stock price ratio.

A.2 1-month Change in Call and Put Volatility Surface (An et al. (2014))

Following An et al. (2014), (implied) volatility surfaces are from retrieved from OptionMetrics. OptionMetrics computes call and put volatility surfaces separately from options with different strikes and maturities using a smoothing algorithm. OptionMetrics provides data on constant maturity volatility surfaces. I follow An et al. (2014) and retrieve volatility surfaces for call and put options with absolute delta of 0.5 and a constant maturity of 30 days. The ΔCVOL (ΔPVOL) is the 1-month change in call (put) volatility surface for each stock.

An et al. (2014) shows ΔCVOL (ΔPVOL) positively (negatively) predicts monthly stock returns for up to 6 months. Note that the original paper uses equal-weighted returns for the ΔCVOL and ΔPVOL portfolios. I use value-weighted portfolios in this paper.

A.3 1-month Implied Volatility Spread (Cremers and Weinbaum (2010))

For each stock-day, I calculate the implied volatility spread between a given call and put option with the same strike price and same expiration date. Then, I weight the implied volatility spread by the sum of the open interest of the given call and put option to get a stock-day measure of the implied volatility spread. Finally, I average the implied volatility spread of each stock across each month to get a stock-month measure. I consider options with best bids greater than 0, implied volatility between 0% and 200%, and positive open interest.

Note that Cremers and Weinbaum (2010) calculates the weekly implied volatility spreads to predict future weekly returns. Cremers and Weinbaum (2010) reports predictability for up to 4 weeks with the weekly implied volatility spreads. In this paper, I use monthly implied volatility spreads.

A.4 1-month Implied Volatility Skew (Xing et al. (2010))

For each option-day, I calculate its moneyness as the stock price divided by the strike price. Then, for each stock-day, I get the at-the-money call option where its moneyness is closest to 1 given that its moneyness lies between 0.95 and 1.05. Also, I get the out-of-the-money put option where its moneyness is closest to 0.95 given that its moneyness is less than equal to 0.95. Finally, for each day, I calculate the implied volatility skew measure as the out-of-the-money put option's implied volatility minus the at-the-money call option's implied volatility. The monthly implied volatility skew measure takes the average of the daily implied volatility skew for each stock for a given month.

Note that Xing et al. (2010) calculates the weekly implied volatility skew (average of daily implied volatility skew) to predict weekly returns. Xing et al. (2010) reports predictability for up to 24 weeks with the weekly implied volatility skew measure. In this paper, I use monthly implied volatility skew.

A.5 1-month Option to Stock Volume Ratio (Johnson and So (2012))

At the end of each month, I replicate the option-to-stock volume ratio by taking ratio of total option volume across maturities from 10 to 60 days to total stock volume over the past month for simplicity and monthly

comparability. Johnson and So (2012) uses a weekly measure using firms with at least 25 call and put contracts traded, and aggregating option volume across maturities expiring in the 30 days after the option is traded.

Note that Johnson and So (2012) shows return predictability at the weekly level. But the replicated monthly measure has significant predictability in the Fama-MacBeth regressions in Table 4.

A.6 Average Moneyness weighted by Volume and Open Interest (Bergsma et al. (2020))

Following Bergsma et al. (2020), I replicate volume- and open interest-weighted average moneyness of a stock as the following:

$$avg_moneyness_volume = \sum_i \frac{K_i}{S} \times \frac{MP_i \times Vol_i}{TotalVol}$$

$$TotalVol = \sum_j MP_j \times Vol_j$$

$$avg_moneyness_OI = \sum_i \frac{K_i}{S} \times \frac{MP_i \times OI_i}{TotalOI}$$

$$TotalOI = \sum_j MP_j \times OI_j$$

where K_i is the strike price of option i, S is the stock price, MP_i is the midpoint option price, Vol_i is the volume of option i, OI_i is the open interest of option i, $TotalVol$ and $TotalOI$ is the total volume and open interest, respectively.

Note that Bergsma et al. (2020) documents stock return predictability at the daily level using equal-weighted returns.

A.7 1-month Percentage Change in Call and Put Option Open Interest (Fodor et al. (2011))

Following Fodor et al. (2011), at the end of each month, I replicate the percentage changes of open interest for call and put options using all options with days to expiration between 30 and 365 days and with open interest of at least 50 (call and put options combined). I use the monthly changes in call-put open interest ratio as the main measure in my tests.

Note that Fodor et al. (2011) constructs a weekly measure and uses options with days to expiration between 30 and 365 at the initial measurement day, and drops options with less than 50 open interest for

either call or put options.

A.8 1-month Out-of-the-money Put-Call Volume and Open Interest Ratio (Kang et al. (2022))

Following Kang et al. (2022), I replicate the out-of-the-money put-call ratio with volume and open interest as the following:

$$OTM_put_call_volume_{i,t} = \frac{\text{Sum of OTM trading volume of puts}}{\text{Sum of OTM trading volume of all options}}$$

$$OTM_put_call_OI_{i,t} = \frac{\text{Sum of } \Delta_+\text{OTM open interest of puts}}{\text{Sum of } \Delta_+\text{OTM open interest of all options}}$$

The $OTM_put_call_volume$ is calculated using trading volume of puts and calls which is straightforward. The $OTM_put_call_OI$ requires more details on the construction. Δ_+OTM open interest is the sum of positive changes in open interest for each option, when the option was out-of-the-money during the month. Then, it is aggregated across all options to obtain a stock-level measure.

Note that Kang et al. (2022) constructs a monthly measure as above and documents stock return predictability at the month-level.

A.9 Stock borrowing fees (Muravyev et al. (2022))

Muravyev et al. (2022) show that stock borrowing fees explain a significant portion of stock returns predicted by options market information. Muravyev et al. (2022) introduce how to accurately estimate stock borrowing fees from implied volatility surfaces of OptionMetrics. Accordingly, stock borrowing fees are estimated with 30-day implied volatility surfaces using delta of 0.50 for calls and -0.50 for puts. This measures the expected borrowing fee over the next 30 days. Formally, the borrowing fees for each stock in each month are calculated as follows:

$$\frac{-(\sigma_C - \sigma_P)}{\sqrt{2\pi(T-t)}}$$

where σ_C (σ_P) is the implied volatility surface of the call (put) options, $T - t$ is the time to maturity that determines the horizon of the expected borrowing fee.

A.10 5-way sorts: $INTRO_q5$

To capture the magnitude of demand, I further sort stocks by $INTRO_q5$ into five groups. If exchanges are responding to stronger demand from informed investors, they are likely to introduce more options across different maturities and across different strike prices. Then, the *INTRO* variable, which takes into account the number of new options introduced, might capture the level of demand. Based on this hypothesis, I split $INTRO_q3 = 1$ into two groups based on the median value of INTRO: $INTRO_q5$=2 (above median number of options introduced) and $INTRO_q5$=1 (below median number of options introduced). Likewise, I split $INTRO_q3 = -1$ into two groups based on the median value of INTRO: $INTRO_q5$=-1 (below median number of options introduced), $INTRO_q5$=-2 (above median number of options introduced). $INTRO_q5$ can be expressed as the following:

$$
\begin{aligned}
&= 2 \quad && (INTRO_t > 0, \text{ above median}) \\
&= 1 \quad && (INTRO_t > 0, \text{ below median}) \\
INTRO_q5_t &= 0 && (INTRO_t = 0) \\
&= -1 \quad && (INTRO_t < 0, \text{ below median}) \\
&= -2 \quad && (INTRO_t < 0, \text{ above median})
\end{aligned}
\tag{5}
$$

Appendix B XGBoost: a supervised learning algorithm

In this section, I describe the rationale of choosing the XGBoost algorithm for predicting option introductions, along with chosen parameter settings.

The goal of using a machine learning model is to capture the potential nonlinearities of and interactions between independent variables that is difficult to be captured in linear regressions. Also, the goal is to predict option introductions with high accuracy, not to select the most interpretable model or to examine the actual nonlinearities and interactions across variables. XGBoost seems to be a reasonable selection for a supervised learning algorithm, where the option introductions are given. See Hastie et al. (2009) for descriptions on tree boosting models.

XGBoost (Extreme Gradient Boosting) is a state-of-the-art supervised learning algorithm that implements gradient-boosted decision trees through sequential tree construction. In simple terms, the data is split into multiple groups based on the independent variables. A model makes a prediction for each group, and the prediction errors are used to adjust the next round of groups and corresponding predictions. In other words, the data is newly split into multiple groups based on the previous step's prediction, and new predictions are made in each group. This process is sequentially repeated at each step. The final prediction is the most recent one and can be understood as the result of combining the predictions in each step. The 'extreme' part of gradient boosting comes from the more sophisticated models used for each group, a different data-splitting method, and the speed of the algorithm. See Chen and Guestrin (2016) for a full description of the XGBoost algorithm.

I train XGBoost algorithms that predict stocks with $INTRO_q3 = 1$ and predicts stocks with $INTRO_q3 = -1$, respectively, in the next month, using stock and option variables in the current month.

R package 'xgboost' is used to train the XGBoost algorithm[2]. The objective is set to 'binary:logistic' which outputs a probability score of 0 and 1 for each observation and minimizes losses from a logistic regression. The performance evaluation metric is set as 'AUCPR (Area Under the Precision-Recall Curve)' that balances between precision and recall rates. Since the unconditional probability of a stock being in $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is around 12% and 6%, respectively, the variables are scaled by these weights, which heavily penalizes mistakes on predicting 1's. The goal is to prevent the model to achieve a 88% and 94% accuracy, respectively, by just predicting no option introductions. The learning rate is set to

[2]https://cran.r-project.org/web/packages/xgboost/index.html

0.05, which decreases the weight the tree after each tree is built. The maximum depth of individual decision trees are limited to 4.

The sample is split into five different periods. First, I use the first 80% of the sample to train the XGBoost algorithm to get out-of-sample predictions for the last 20% of the sample. Next, I use the first 60% and last 20% of the sample as the training sample to get out-of-sample predictions for the 4th period of the sample. Repeating this process, I get out-of-sample predictions for each of the five periods based on the training sample that excludes the target period.

The training sample is trained with a maximum number of iterations set as 300. I set the early stopping rounds as 30, which means that if AUCPR does not improve for 30 consecutive rounds, the model stops training. In each training session, 80% of randomly selected observations are used, and 80% of randomly selected variables are used. This prevents the model to train heavily on a short period or on a small set of variables.

As mentioned earlier, the model outputs a number between 0 and 1 for each stock-month. I classify the model's prediction as 1 if the output is larger than 0.5, and 0 otherwise. The true positive occurs when the model predicts 1 and the stock belongs to either $INTRO_q3 = -1$ or $INTRO_q3 = 1$, depending on the direction of the model.

Appendix C Supplementary Figures and Tables

Firm A's stock price in Figure C.1 went from \$46 in February 2022 to \$56.2 in March 2022. The strike prices \$75 and \$80 were introduced above the prevailing maximum strike price of \$70. In this case, INTRO=2, since two strike prices were introduced above the prevailing maximum strike price. Also, $INTRO_q3 = 1$, since *INTRO* is positive. On the other hand, Firm B's stock price went from \$28.1 in February 2022 to \$28.8 in March 2022. The strike price \$12.5 was introduced below the prevailing minimum strike price of \$15. In this case, INTRO=-1, since one strike price was introduced below the prevailing minimum strike price. Also, $INTRO_q3 = -1$ because *INTRO* is negative.

	Firm A		Firm B	
	2022-02-28	2022-03-31	2022-02-28	2022-03-31
Stock Price	46	56.2	28.1	28.8

INTRO_q3 = 1		INTRO_q3 = -1	
Strike price	Strike price	Strike price	Strike price
	80	...	...
	75	22.5	22.5
70	70	20	20
65	65	17.5	17.5
60	60	15	15
...	...		12.5

Figure C.1: Examples of Strike Price Introductions

This figure shows two firms with different values of $INTRO_q3$. For Firm A, strike prices were introduced above the prevailing maximum strike price. For Firm B, strike price were introduced below the prevailing minimum strike price. Both examples are from the end of February 2022 to the end of March 2022. Each firm's stock prices and $INTRO_q3$ variable are shown in the figure.

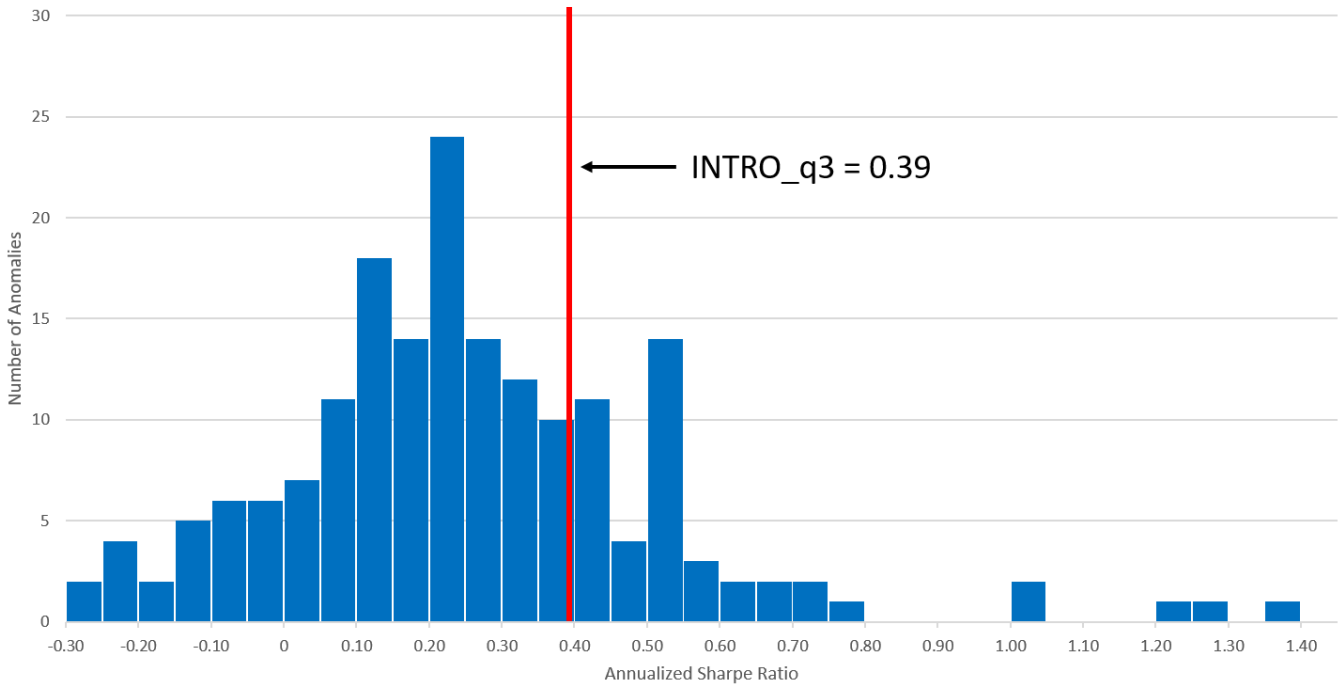


Figure C.2: Annualized Sharpe Ratio of Stock Anomalies

This figure shows the annualized Sharpe ratio of 180 stock anomalies from Chen and Zimmermann (2022) based on value-weighted long-short decile portfolios. The red vertical line denotes the Sharpe ratio of the portfolio constructed from $INTRO_q3$. The sample period is 1996-2023.

Table C.1: Summary Statistics

This table shows summary statistics of stocks with listed options, along with average characteristics of the stocks with $INTRO_q3 = -1$ and $INTRO_q3 = 1$, respectively. The sample period is 1996-2023.

	$INTRO_q3$			
Variables	-1	0	1	Total
Avg. number of stocks	156	1,894	293	2,343
Avg. market cap ($MM)	4,665	4,050	9,991	
Avg. 1-month return (%)	-7.5%	1.1%	9.2%	
Avg. momentum return (%)	6.0%	27.5%	50.4%	
Avg. log book-to-market	-1.09	-0.89	-1.00	

Table C.2: Daily Event-Time Returns for $INTRO_q3 = -1, 1$

This table shows event-time returns (in %) K (=-1,0,1,2,3) days before and after option introductions for $INTRO_q3 = -1, 1$. The event at Day=0 is the day when the options are introduced. At the end of each month, stocks are divided into three groups: $INTRO_q3 = -1, 0, 1$. $INTRO_q3 = 1$ represents a portfolio of stocks with new options introduced above the prevailing maximum strike price (at the end of the previous month) over the past month, $INTRO_q3 = -1$ represents a portfolio of stocks with new options introduced below the prevailing minimum strike price (at the end of the previous month) over the past month, and $INTRO_q3 = 0$ represents the rest of the stocks. The returns are reported in excess of the market return for $INTRO_q3 = 1$ and $INTRO_q3 = -1$. Stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ are value-weighted by their market cap at the end of each event day. Then, the portfolio returns averaged across each day are reported. Stocks are required to have at least one listed option before portfolio formation. Stocks with prices less than $1 are excluded. The sample period is 1996-2023.

	$INTRO_q3$				Long-short	
Day	-1	t-stat	1	t-stat	1 - (-1)	t-stat
0	-0.31	(-5.60)	0.15	(5.78)	0.45	(7.24)
1	0.08	(2.30)	-0.03	(-1.22)	-0.10	(-2.45)
2	0.03	(0.72)	-0.03	(-1.51)	-0.06	(-1.34)
3	0.03	(0.99)	-0.04	(-1.71)	-0.07	(-1.75)
4	-0.03	(-1.11)	-0.02	(-1.01)	0.01	(0.28)
5	-0.04	(-1.20)	0.01	(0.62)	0.06	(1.28)
6	0.00	(-0.06)	0.02	(0.80)	0.02	(0.53)
7	-0.03	(-1.02)	-0.02	(-0.68)	0.01	(0.35)
8	0.02	(0.81)	0.02	(0.74)	0.00	(-0.10)
9	0.03	(0.85)	-0.02	(-0.92)	-0.05	(-1.15)
10	-0.03	(-0.93)	0.01	(0.27)	0.04	(0.87)

Table C.3: Leverage Ratio and Bid-Ask Spread for New Options

This table shows the bid-ask spread and embedded leverage of new out-of-the-money call options for $INTRO_q3 = 1$ and new out-of-the-money put options for $INTRO_q3 = -1$. Newly introduced out-of-the-money options are required to have positive trading volume on the introduction day to be included in the sample. The old options represent the previous month-end's out-of-the-money call options with the maximum strike price or out-of-the-money put options with the minimum strike price. The newly introduced and existing options are matched by their underlying stock, month, strike price, and expiration date. The bid-ask spread is calculated as the best offer minus the best bid divided by the bid-ask midpoint. The embedded leverage of each option is calculated as the delta times the underlying stock price divided by the option price given by the bid-ask midpoint Frazzini and Pedersen (2022). The average bid-ask spread and leverage are reported. The sample period is 1996-2023.

	OTM put	OTM call
Variables	(Volume > 0 on introduction day)	
Bid-ask spread (%, new)	49.1%	64.7%
Bid-ask spread (%, old)	25.6%	30.6%
New-minus-old (%P)	23.4%	34.2%
Leverage (new)	8.72	4.94
Leverage (old)	6.87	4.17
New/old ratio	1.27	1.19
Option-level obs.	62,650	22,811

Table C.4: Correlation of $INTRO_q3$ with Existing Variables

This table shows the correlation between $INTRO_q3$ with existing variables in the literature. Stock-level variables from stock characteristics include: 1-month return ('ret_1M'), 1-month idiosyncratic volatility ('ret_ivol'), 11-month return skipping the most recent month ('ret_MOM'), log market cap ('mcap_log'), log book-to-market ratio ('beme_log'), profitability ('profitability'), and investment ('investment'). Stock-level variables from option characteristics include: 1-month change in call volatility surface ('ΔCVOL', An et al. (2014)), 1-month change in put volatility surface ('ΔPVOL', An et al. (2014)), 1-month option-to-stock volume ratio ('O/S ratio', Johnson and So (2012)), 1-month average implied volatility spread ('volatility_spread', Cremers and Weinbaum (2010)), 1-month average implied volatility skew ('volatility_skew', Xing et al. (2010)), 1-month percentage change in call-put open interest ratio ('Δcall_put_open_interest', Fodor et al. (2011)), out-of-the-money put-call open interest ratio ('OTM_put_call_OI', Kang et al. (2022)), out-of-the-money put-call volume ratio ('OTM_put_call_volume', Kang et al. (2022)), open interest-weighted average moneyness ('avg_moneyness_OI', Bergsma et al. (2020)), volume-weighted average moneyness ('avg_moneyness_volume', Bergsma et al. (2020)). Details on constructing the option variables are presented in Appendix A. Stocks are required to have at least one listed option before portfolio formation. Stocks with prices less than $1 are excluded. The sample period is 1996-2023.

variables	$INTRO_q3$
ret_1M	0.25
ret_ivol	-0.08
ret_MOM	0.16
mcap_log	0.10
beme_log	0.01
profitability	0.00
investment	0.00
volatility_spread	-0.06
volatility_skew	-0.01
ΔCVOL	-0.04
ΔPVOL	-0.04
avg_moneyness_volume	-0.09
avg_moneyness_OI	-0.14
O/S ratio	0.03
Δcall_put_open_interest	0.00
OTM_put_call_volume	0.10
OTM_put_call_OI	0.07

Table C.5: Robustness Tests

This table shows the event-time cumulative return spread (in %) between $INTRO_q3 = 1$ and $INTRO_q3 = -1$, 12 months after portfolio formation, under different specifications. For option maturity, stocks are sorted by $INTRO_q3$ using only options that have time-to-maturity below or above 100 days. For borrowing costs, stocks are sorted by $INTRO_q3$ using only stocks identified as easy-to-borrow and hard-to-borrow. For the subperiod analysis, the sample is split into two periods at September 2009. For the event-time analysis specifications, the t-statistic is calculated using Newey and West (1987) standard errors with 12 lags, and stocks are sorted by $INTRO_q3$ using only those that had no options introduced in the prior 12 months. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

Variables	Month 12 1 - (-1)	t-stat
Option maturity (all options)		
Time to maturity < 100	4.11	(3.95)
Time to maturity ≥ 100	4.92	(4.65)
Borrowing costs		
Easy-to-borrow stocks (fee ≤ 1%)	4.39	(4.15)
Hard-to-borrow stocks (fee > 1%)	4.23	(3.82)
Subperiod analysis		
1996-2009	4.36	(2.90)
2009-2023	4.46	(3.63)
Event-time analysis specifications		
NW 12 lags	4.39	(2.32)
No introductions in past 12 months	4.23	(2.10)

Table C.6: Event-time Cumulative Portfolio Returns Double-sorted by Number of Option Introductions

This table shows the event-time cumulative stock portfolio returns (in %) up to 12 months after portfolio formation double-sorted by $INTRO_q3$ and the number of options introduced. The stocks in $INTRO_q3 = -1$ and $INTRO_q3 = 1$ are split into two groups each, respectively, by the median number of options introduced. The event-time cumulative return spread between stocks with $INTRO_q3 = 1$ and $INTRO_q3 = -1$ is reported separately for stocks with a high and low number of newly introduced options. Portfolios are value-weighted by each stock's market capitalization at the end of the formation period. Stocks must have at least one listed option prior to portfolio formation. Stocks with prices below $1 are excluded. T-statistics are reported in parentheses. The sample period is 1996-2023.

	Number of options introduced			
	Low		High	
Month	1 - (-1)	t-stat	1 - (-1)	t-stat
1	0.39	(1.32)	0.51	(1.44)
2	0.57	(1.50)	0.99	(2.02)
3	0.89	(2.07)	1.85	(3.08)
4	0.70	(1.35)	2.66	(3.99)
5	1.02	(1.77)	3.39	(4.47)
6	1.25	(1.98)	4.34	(5.27)
7	1.45	(2.07)	5.11	(5.63)
8	1.79	(2.34)	5.50	(5.63)
9	2.08	(2.55)	6.04	(5.77)
10	2.25	(2.66)	6.22	(5.80)
11	2.51	(2.79)	6.20	(5.37)
12	2.48	(2.54)	6.50	(5.36)