Trading Options with Uncertainty Risk around Earnings Announcements

Eric Ahn

Advisor: Dr. Daniel Rigobon

Submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Engineering
Department of Operations Research and Financial Engineering
Princeton University

May 2024

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Eric Ahn

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Eric Ahn

Abstract

Quarterly earnings announcements release a substantial amount of financial information about a firm all at once, often inducing significant, discontinuous jumps in stock prices. Traditional option pricing models, such as the Black-Scholes model and stochastic volatility models, fail to model these jumps. In this thesis, we aim to estimate the earnings volatility as priced into existing options contracts. To do so, we use a modified Black-Scholes model and a modified Heston model that incorporates jumps at earnings. Since both Black-Scholes and Heston have differing underlying models for the stock price, we obtain two distinct estimates for earnings volatility. These estimates are observed to be similar, although the estimate derived from the Heston model is marginally higher. We then attempt to explain the estimates of the earnings volatility using its lags, market volatility, and quarter information, among other features. We observe that the previous two earnings volatilities are most predictive of future earnings volatility. We then build more sophisticated prediction-focused models and observe that a LASSO model exhibits the best out-of-sample error. These predictive models allow us to construct four distinct trading strategies based on the predicted earnings volatility, which are evaluated against several benchmarks and prove to be profitable in a back-test. This work shows that simple heuristic-based trading strategies can exploit discrepancies in the valuation of volatility in options contracts around earnings announcements.

Acknowledgements

I would like to first thank my advisor, Professor Daniel Rigobon, for his mentorship and advice throughout this entire thesis process. His enthusiasm and guidance kept me focused and grounded throughout the year. I would also like to thank Rajita Chandak for all her support. Without her feedback, this thesis would not be at the state it is today.

To my family, thank you so much for all the love and support you have given me over the past four years. To Mom and Dad, thank you for believing in me and pushing me to accomplish things I was not sure would be possible. To Amy, thank you for being the best older sister I could have asked for.

Thank you to my ORFE friends, Megan King, Vivek Kolli, Katie Kolodner, and Burke Pagano. No matter what ORFE classes I took, I knew you guys would be the one constant through them all. It is with your support that I survived the barrage of ORFE problem sets.

To my roommates, Bill Ao, Peter Ng, and Harvey Wang, thank you for all the memories we made this year and for making me a better person.

And to all my friends here at Princeton, thank you so much. You have made my college experience a much richer and greater experience than I could have ever imagined.

To my friends and family

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Motivation

Every quarter, firms are required by the Securities and Exchange Commission to report their "earnings". Earnings refer to the quarterly financial reports issued by publicly traded companies, which provide insights into the companies' profitability. These reports generally include three fundamental financial statements: the income statement, the balance sheet, and the cash flow statement, each providing information about the aspects of the company's financial health. Additionally, many firms frequently provide forward estimates, or "company guidance", which are the companies' public forecasts of their performance in the current quarter or upcoming year. As companies tend to release a substantial amount of information at once through these announcements, this often leads to volatility in the stock prices, sometimes even leading to "gapping" — jumps in the stock price — as investors rapidly absorb previously undisclosed information about the company.

In financial markets, options are among the most frequently traded derivatives. These financial derivatives give buyers, referred to as "the long party", the right, but not the obligation, to buy or sell an underlying asset at a predetermined strike price

K and by a specific date T. One of the longest-standing problems in the field of financial mathematics has been the problem of accurately pricing these derivatives.

The first of these attempts was pioneered by Fischer Black and Myron Scholes with the introduction of their coveted Black-Scholes model [4]. The celebrated Black-Scholes model for pricing European options assumes that the price dynamics of the underlying asset follow a geometric Brownian motion with constant volatility σ, signifying that the volatility of the stock's return is constant. This simplification allows us to solve for a closed-form solution to the option pricing problem. Nevertheless, a consensus among experts suggests that this constant volatility assumption does not hold empirically when observing market prices. More precisely, empirical evidence has shown that the volatility of an asset is itself randomly distributed [10], often also following another geometric Brownian model that is correlated to the Brownian motion that drives the stock price. These models are commonly characterized as stochastic volatility models.

Bridging these two concepts, we can see the interconnectedness of earnings and options pricing through the lens of volatility. Typically, the volatility of an asset increases around the time of earnings announcements, as after earnings, the asset price tends to jump as investors rapidly incorporate new fundamental information about a company. Therefore, we need an options model that can account for these jumps in stock price and explain the increase in volatility around earnings, providing a more accurate description of the price dynamics.

The concept of implied volatility will be extensively discussed in this thesis. Unlike realized volatility, which is calculated from the historical price fluctuations of the underlying asset, implied volatility is derived from the market price of the option itself. When trading options, all elements of an option — such as its market price, the price of the underlying asset, strike price, maturity, and the risk-free rate — are typically known, except for the volatility. By "backing out" the volatility that explains the current market price of the option, we can define this value as the implied volatility. Implied volatility is a key concept, particularly in the realm of options trading, as it represents the market forecast of future realized volatility of the asset over the life of the option [9]. Therefore, trading options, to a certain extent, can be perceived as trading volatility, as each party has different predictions on what the volatility of the underlying contract is, presupposing that all traders adhere to the same theoretical option pricing model.

In this thesis, we aim to estimate the earnings volatility that is priced into existing options contracts. We do so under the assumption of a modified constant volatility model and a modified stochastic volatility model that incorporate jumps at earnings. As the two mdoels posit different underlying dynamics for the stock price, we obtain two distinct earnings volatility estimates, which we then compare. Following this, we explore potential relationships between the priced-in earnings volatility and different fundamental features such as its past earnings volatility, dispersion of analyst forecasts, and the size of a company to gain intuition on how such features could potentially influence earnings volatility. Finally, using a predictive model, we aim to forecast earnings volatility and investigate trading strategies that could leverage this information.

1.2 Literature Review

1.2.1 Option Pricing

We examine various existing options and stock price models. Since the influential work of Black and Scholes [4], a diverse array of models have emerged that aim to more accurately model the dynamics of stock prices. Empirically, it has been observed that stock returns frequently exhibit fat tails and are skewed to the left [13], a phenomenon that is inconsistent with the assumptions of the Black-Scholes model. To rectify these discrepancies, two approaches have been proposed: stochastic volatility models and jump processes.

Within stochastic volatility models, the most popular approach is the Heston model [10], which extends the Black-Scholes model by having a stochastic variance process that follows a Cox-Ingersoll-Ross (CIR) process. A CIR process, first used to describe the evolution of interest rates, is characterized by mean reversion, meaning that it models the process to move towards a long-term average over time. Heston goes on further to provide a closed-form solution for the price of a call option under this model. Further in-depth explanation of the Heston model is detailed in Chapter 2.

Under the category of models with jump processes, the most prominent model is the Merton jump-diffusion model [14]. This model presupposes that the stock price exhibits small, continuous fluctuations in addition to large, discontinuous jumps:

$$\frac{dS_t}{S_t} = (r - \lambda k)dt + \sigma dW_t + kdq_t$$

Here, S_t is the price of the stock, r is the risk-free rate, σ is the volatility of the asset, W_t is the Brownian motion, k is the magnitude of the jumps, q_t is the Poisson process, and λ is the rate. Merton assumes that k is also randomly distributed. The diffusion component is represented by the familiar geometric Brownian motion that is observed in the Black-Scholes model, denoting continuous price movements. The model then also introduces a jump component, characterized by log-normal jumps of size k, driven by a compound Poisson process q_t with rate λ. These jumps signify sudden discontinuous changes in the underlying asset price due to the arrival of new information. Therefore, stock price dynamics can now follow non-continuous paths. We see that if $k = 0$, then the Merton model simply reduces back to the familiar

Black-Scholes model. If the distribution of k obeys:

$$\ln(1+k) \sim N(\gamma, \delta^2)$$

where γ and δ^2 are the average log jump size and the variance of the jump sizes, respectively, then Merton shows a closed-form solution to the option pricing problem under this model.

Kou further expands upon this in [12] where he offers a new model that also explains the skew and tail in the returns distribution but also provides analytical solutions to option pricing problem for path-dependent options, options that depend not only on the final price of the underlying asset but also the path that it took to reach that final price.

In this model, the asset price once again follows a geometric Brownian motion, similar to other models, but is further augmented by a compound Poisson process where the jump sizes follow a double exponential distribution. The model is expressed as:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t + d\left(\sum_{i=1}^{q_t}(V_i - 1)\right)$$

where q_t is a Poisson process with rate λ, similar to the Merton model, and $\{V_i\}$ denotes a sequence of independent, identically distributed random variables such that $\log(V)$ follows an asymmetric double exponential distribution. Once again, Kou is able to derive an analytical solution to the option pricing problem under this model. Both Merton's and Kou's models feature randomly distributed jump times and magnitudes. Kou interprets the jump component of the models as the response to outside news that comes out at random times. Due to the double exponential distribution of the jump, which is characterized by high peaks and heavy tails, both an overreaction and underreaction to news can be modeled. However, this stands in contrast to earnings announcements where the timing of potential jumps is predetermined.

1.2.2 Pricing Earnings Announcements

Next, we look at different works in the field of earnings announcements. The pioneering study by Ball and Brown [2] is said to have reshaped how investors thought about capital markets research and interpret accounting earnings.

Ball and Brown seek to explain the relationship between earnings and stock price residuals. Earnings residuals are defined as the difference between the anticipated and actual released earnings numbers, while stock price residuals are the difference in stock price preceding and following the earnings announcement. Fundamentally, their research aims to explain stock price movement with respect to the discrepancy in earnings expectations. To do this, Ball and Brown estimate a simple linear model with firm-specific coefficients. Their results confirm that accounting figures and earnings are indeed factors that investors consider when trading voluntarily.

Similarly, Ball and Brown's methodology select the most representative and informative accounting numbers: earnings per share (EPS) and net income. This thesis adopts a similar approach and focuses on the projections of EPS and Sales as these two metrics have emerged as the most important fundamental accounting values, evidenced by the prevalence of multiples such as enterprise value-to-sales (EV/Sales) and price-to-earnings (P/E) by both sell-side and buy-side analysts for fundamental valuation of firms. Like Ball and Brown, we also explore both firm-specific models and a pooled model, enabling us to compare how coefficients vary across firms.

Donelson in his paper [7] discusses the topic of earnings uncertainty and its capacity to predict earnings returns. The study introduces a firm-specific measure of earnings uncertainty by matching firm i at time t with firms of comparable characteristics in the preceding period $t-1$. The variance of the earnings realizations of matched firms during period t represents the earnings uncertainty. Donelson posits that the higher the variance of earnings found in the past of matching firms, the broader dispersion of earnings results that could be realized by firm i. The study then empirically evaluates these measures and determines how predictive they are of future earnings. We take a similar approach to Donelson, where we employ a measure of earnings volatility and assess its relationship to other fundamental values. Emulating the features considered by Donelson such as analyst dispersion and firm market cap, this thesis includes similar features as well.

Patell and Wolfson find in [19] that the majority of price movement occurs minutes after the release of the earnings announcement, which then makes it reasonable to model this information as a discontinuous jump in the stock price. Further exploration by Patell and Wolfson in [18] shows that option prices contain information about earnings announcements. To do so, they describe the time series behavior of implied volatility around earnings.

They assume the following model: The instantaneous variance remains at baseline level γ, except during the announcement period which begins at time t_0. Throughout the disclosure interval of duration τ, the variance now increases to a level $\gamma + \delta$. They then find that the average variance up until expiration T is equal to the area under the instantaneous variance curve from that point to expiration divided by the length of time to expiration $T - t$. Therefore, at a time t which precedes the announcement date, the average variance can be given by:

$$\sigma^2(t) = \frac{1}{T-t}(\gamma(T-t) + \tau\delta) = \gamma + \frac{\tau\delta}{T-t} \tag{1.1}$$

They then use this model to find the magnitude of the increase in earnings volatility around earnings from empirical data.

Dubinsky and Johannes in [8] derive a very similar model and estimator to that of the one described by Patell and Wolfson above. They describe methods for estimating earnings volatility and uncertainty from option prices using a stock price model that incorporates discontinuous jumps similar to the one found in the effect of macroeconomic news by Beber and Brandt in [3]. In their research, Dubinsky and Johannes analyze how to quantify earnings uncertainty under these models and demonstrate that a high earnings volatility is highly correlated with subsequent volatile earnings returns. This thesis draws heavily upon the methods proposed by Dubinsky and Johannes for estimating earnings volatility and uncertainty. A more comprehensive exposition of their methodological approaches is provided in Chapters 2 and 3.

Most of the literature analyzing options pricing models deals with pricing index options. Bakshi and Cao describe in [5] that the return distributions of individual equity firms were more volatile, less skewed, and had fatter tails than the market index. Bakshi and Cao then show in [1] that return jumps and volatility jumps are critical for pricing options, with return jumps being empirically more relevant. In alignment with these developments, this paper focuses predominantly on the analysis of return jumps, and not volatility jumps.

Chapter 2

Option Pricing Models

2.1 Stock Models

In this section, we formulate the options pricing problem. To do so, we first introduce the Black-Scholes and Heston models and subsequently extend these models with discontinuous jumps at earnings announcements. Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a probability space, $\mathcal{F}$ be the normal filtration, and $\mathbb{Q}$ be the risk-neutral measure. The risk-neutral measure $\mathbb{Q}$ is defined such that the current share price is exactly equal to the discounted expectation of the share price at time T under this measure, mathematically represented as:

$$S_t = e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[S_T]$$

From now on, we define S_t to be the price of the asset at time t, r to be the risk-free rate, σ to be the volatility of the stock, and W_t be the $\mathbb{Q}$-Brownian motion.

2.1.1 Black-Scholes Model

First, we model the stock price under the classic Black-Scholes model given in [4]. Under the Black-Scholes model, the underlying stock dynamics are modeled by the following stochastic differential equation:

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t \tag{2.1}$$

The Black-Scholes model above assumes that both the risk-free rate r and the volatility σ are constant and non-negative. The simplifying assumption of volatility being constant is a necessary assumption that allows for a closed-form derivation of the price of a European option which we describe in Section 2.2.

2.1.2 Heston Model

In contrast to the Black-Scholes model where the volatility is treated as a constant, we now allow for the evolution of the volatility of an asset. Primarily, the Heston Model [10] assumes that the volatility of the asset is neither constant nor deterministic, but instead follows a stochastic process, specifically the the Cox-Ingersoll-Ross process. The basic Heston model is described by the following set of stochastic differential equations:

$$\begin{aligned} dS_t &= rS_t\,dt + \sqrt{\nu_t}S_t\,dW_t^{(0)} \\ d\nu_t &= \kappa(\theta - \nu_t)\,dt + \xi\sqrt{\nu_t}\,dW_t^{(1)} \end{aligned} \tag{2.2}$$

The parameters κ, θ, ξ are additional parameters found in this model in addition to the ones found also in the Black-Scholes model. The drift factor $\kappa(\theta - \nu_t)$ in the volatility equation ensures mean reversion of the variance back towards the long-term value θ with a rate of reversion controlled by the strictly positive parameter κ. The parameter ξ reflects the volatility of the volatility, representing the rate at which the volatility itself varies.

To account for the leverage effect, the well-established negative relationship between returns and subsequent volatility, it is common to correlate the Wiener processes $W_t^{(0)}, W_t^{(1)}$ such that:

$$dW_t^{(0)} \cdot dW_t^{(1)} = \rho\, dt$$

When the returns on equity are negative, the reactions by investors are more volatile, thereby increasing volatility, and vice versa. The parameter ρ represents the correlation between the underlying asset price and its volatility. Furthermore, we also assume the Feller condition holds where:

$$2\kappa\theta > \eta^2$$

This is a sufficient condition for the volatility process ν_t to remain strictly positive. Once again, we again assume that the risk-free rate is non-negative. These assumptions are ones that Heston also adopts in his paper [10].

2.1.3 Models with Discrete Earnings Jumps

The previously discussed models only account for continuous stock price dynamics and do not account for the rapid changes typically associated with earnings announcements. We model these rapid changes as discontinuous price jumps at earnings. Unlike other models with discontinuous jumps such as the Merton jump-diffusion model which assumes a random distribution of jumps driven by a Poisson process, the timings of earnings are pre-disclosed, and so the occurrence of jumps in this context are predictably non-random. Thus, this paper will adopt an alternative model that reflects this difference, utilizing the same approach taken by Dubinsky and Johannes in [8].

Let N_t^d count the number of predictable earnings announcements prior to time t:

$$N_t^d = \sum_j \mathbb{1}_{[\tau_j \le t]}$$

where the τ_j's denote the sequence of predictable times that represent when the earnings announcements occur. The magnitude of the jump is given by $Z_j = \log(S_{\tau_j}/S_{\tau_j^-})$, the log-return of the stock price right after the announcement relative to right before it. Following Dubinsky, we assume that the sizes of these jumps Z_j are independent and follow a normal distribution $Z_j \sim N\left(-\frac{1}{2}\sigma_j^2, \sigma_j^2\right)$ where σ_j is the anticipated volatility surrounding the earnings announcement. We choose the mean of Z_j such that $\mathbb{E}[S_{\tau_j}/S_{\tau_j^-}] = 1$. This ensures that on average after a jump occurs, there is no systematic increase or decrease in the stock price and reflects the idea that while jumps introduce additional volatility into the stock price process, they do not necessarily have a predictable direction. We then extend the dynamics of the original Black-Scholes model from Equation 2.1 as follows:

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t + d\left(\sum_{j=1}^{N_t^d} S_{\tau_j^-}[e^{Z_j} - 1]\right) \tag{2.3}$$

where now we have discontinuous jumps at the earnings announcements. Similarly, we do the same with the Heston model, which gives us the dynamics:

$$\begin{aligned} dS_t &= rS_t\,dt + \sqrt{\nu_t}S_t\,dW_t^{(0)} + d\left(\sum_{j=1}^{N_t^d} S_{\tau_j^-}[e^{Z_j} - 1]\right) \\ d\nu_t &= \kappa(\theta - \nu_t)\,dt + \xi\sqrt{\nu_t}\,dW_t^{(1)} \end{aligned} \tag{2.4}$$

The above assumptions for each model still hold for these models as well. The anticipated volatility of the jump size Z_j captures the earnings volatility and the fundamental uncertainty priced in by investors.

In the Heston model with jumps, we assume that there is a jump at earnings which leads to a discontinuous stock path. We could also extend the Heston model by incorporating discontinuous jumps in volatility as well at this time. However, due to complexity and lack of estimability from ATM options, and Bakshi and Cao showing in [1] that the effect of price jumps was more empirically relevant than volatility jumps, we choose to simply focus on the price jumps in this paper.

2.2 Option Pricing

Let $\phi(s)$ denote the payoff function of a derivative that only depends on the stock price at maturity i.e. not a path-dependent derivative. We will first concern ourselves with European calls with the payoff $\phi(S_T) = (S_T - K)^+$ where K is the strike price of the option.

Using the fundamental theorem of asset pricing, the price of the European call option under the Black-Scholes model is:

$$v^{BS}(t,s) = e^{-r(T-t)}\mathbb{E}\{\phi(S_T)|S_t = s\}$$

and under the Heston model will be:

$$v^{H}(t,s,v) = e^{-r(T-t)}\mathbb{E}\{\phi(S_T)|S_t = s, \nu_t = v\}$$

Under the Black-Scholes model without jumps from Equation 2.1, we can explicitly solve for $v^{BS}(t,s)$ as shown in [4], which gives us:

$$v^{BS}(t,s) = sN(d_1) - Ke^{-r(T-t)}N(d_2) \tag{2.5}$$

where

$$d_1 = \frac{\ln(s/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}$$

and $N(\cdot)$ is the CDF of the standard normal distribution.

In addition to deriving a closed-form solution to the option-pricing problem as we did with Equation 2.5, options can also be priced numerically via Monte-Carlo simulation, as described in [6]. This process involves simulating a possible path for the stock price over the life of the option, from time t to time T, and then calculating the option's payoff for that path, denoted as $\phi(S_T)_1$. This procedure is then repeated until we record n possible payoffs. The average $\phi(\bar{S}_T) = \frac{1}{n}\sum_{i=1}^{n}\phi(S_T)_i \to \mathbb{E}[\phi(S_T)]$ as $n \to \infty$ by the Law of Large Numbers. The resulting average can then be discounted to estimate the option price. This method is particularly useful for models where it is difficult to find closed-form solutions, such as with the Heston model incorporating jumps as per Equation 2.4.

To simulate stock price paths under the Heston model with jumps, we employ the Euler Scheme as detailed in [15]. We break up the life of the option from time t to time T into M equally spaced time steps where $\Delta t = (T - t)/M$. The stock and volatility paths are simulated using the following equations:

$$
\begin{aligned}
v_{n+1} &= \left(v_n + \kappa(\theta - v_n)\Delta t + \xi Z_v \sqrt{v_n \Delta t}\right)^{+} \\
\ln(S_{n+1}) &= \ln(S_n) + \left(r - \frac{1}{2}v_n\right)\Delta t + Z_s\sqrt{v_n \Delta t} + \sum_{j=1}^{N_t^d} Z_j
\end{aligned}
\tag{2.6}
$$

Here, Z_s, Z_v are the standard normal random variables, with Z_s having a correlation ρ to Z_v. This pair of correlated standard normal random variables can be generated by $Z_v = Z_1$ and $Z_s = \rho Z_v + \sqrt{1-\rho^2}Z_2$ where Z_1 and Z_2 are two independent draws from the standard normal distribution. Z_j, which represents the earnings jump, is generated from a normal distribution with a mean $-\frac{1}{2}\sigma_j^2$ and a variance σ_j^2 as described above. Exponentiating this log-stock price yields the simulated stock price process.

The above formulas and payoffs are valid only for European options. However, most of the equity options traded in the US market are American options. American options differ from European options in that American options can be exercised at any time before and including the day of expiration, while European options can only be exercised on the day of expiration. We show that in the case of no- or low-dividend stocks, the price of the American call option and the European call option should be the same.

To do so, we start by showing that it is never optimal to exercise an American option early. We observe that the payoff function $\phi(s) = (s-K)^+$ is a convex function. We also know that under the risk-neutral measure $\mathbb{Q}$, from the fundamental theorem of asset pricing, we have that $\mathbb{E}_{\mathbb{Q}}[S_T|S_0 = s] = e^{rT}s$. Therefore, by Jensen's inequality:

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[\phi(S_T)] &= \mathbb{E}_{\mathbb{Q}}[(S_T - K)^+|S_0 = s] \\ &\geq \phi(\mathbb{E}_{\mathbb{Q}}[S_T|S_0 = s]) \\ &= (e^{rT}s - K)^+\end{aligned}$$

Therefore, we see that:

$$v(t,s) = e^{-rT}\mathbb{E}\{\phi(S_T)|S_0 = s\}] \geq (s - Ke^{-rT})^+ > (s - K)^+$$

From here, we can conclude with any maturity T, the stopping value of $(s - K)^+$ will always be strictly less than the corresponding value of the European call option. Therefore, it is shown that it is never optimal to exercise early, and therefore, the price of the American call option and the European call option should be the same. For this reason, even though our estimators of earnings volatility in Chapter 3 assume the prices of European call options, even if we use prices of American call options of no- or low-dividend stocks, our estimations should remain relatively robust.

Throughout this thesis, the concept of option moneyness will be referenced, describing the relationship between the strike price of an option and the current spot price of the underlying asset.

- In-The-Money (ITM): A call option is in-the-money when the strike price is below the spot price of the underlying asset, indicating profitability if exercised immediately.
- At-The-Money (ATM): An option is at-the-money when its strike price is equal to or close to the spot price of the underlying asset.
- Out-of-the-Money (OTM): A call option is out-of-the-money when the strike price is above the spot price of the underlying asset, rendering it valueless if exercised at that time.

Chapter 3

Overview of Models

3.1 Overview of Earnings Estimator

3.1.1 Black-Scholes Estimator

In this section, we develop an earnings estimator analogous to the one presented in Dubinsky and Johannes' paper [8]. Let us first consider the Black-Scholes model with discrete earnings jumps from Equation 2.3. In this model, σ is the diffusive volatility, indicative of the level of volatility without jumps, and is assumed to be a constant. By applying Ito's formula for semi-martingales with jumps to the function $f(t, S_t) = \ln(S_t)$ as seen in [20], and integrating from 0 to T, we obtain:

$$\ln(S_T) - \ln(S_0) = \left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma W_T + \int_0^T \ln(S_\tau) - \ln(S_{\tau^-}) dN_T^d$$

We know that:

$$\ln(S_\tau) - \ln(S_{\tau^-}) = \ln\left(\frac{S_\tau}{S_{\tau^-}}\right) = Z_j$$

where the second equality comes from the definition of Z_j. Plugging in and rewriting the integral as a summation as seen in [20], we get:

$$\ln(S_T) - \ln(S_0) = \left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma W_T + \sum_{j=1}^{N_T^d} Z_j$$

which we can simplify as:

$$S_T = S_0 \exp\left[\left(r - \frac{1}{2}\sigma^2\right) T + \sigma W_T + \sum_{j=1}^{N_T^d} Z_j\right] \tag{3.1}$$

We can further simplify this equation as follows. Since we assumed $Z_j \sim N\left(-\frac{1}{2}\sigma_j^2, \sigma_j^2\right)$, we know that the sum of Z_j is also normally distributed:

$$\sum_{j=1}^{N_T^d} Z_j \sim N\left(-\frac{1}{2}\sum_{j=1}^{N_T^d} \sigma_j^2, \sum_{j=1}^{N_T^d} \sigma_j^2\right)$$

and therefore, we have that:

$$\sigma W_T + \sum_{j=1}^{N_T^d} Z_j \sim -\frac{1}{2}\sum_{j=1}^{N_T^d} \sigma_j^2 + W_T\sqrt{\left(\sigma^2 + T^{-1}\sum_{j=1}^{N_T^d} \sigma_j^2\right)}$$

Substituting this into Equation 3.1 above and rewriting, we get:

$$S_T = S_0 \exp\left[\left(r - \frac{1}{2}\left(\sigma^2 + T^{-1}\sum_{j=1}^{N_T^d} \sigma_j^2\right)\right) T + W_T\sqrt{\left(\sigma^2 + T^{-1}\sum_{j=1}^{N_T^d} \sigma_j^2\right)}\right]$$

We can see that this results in the Black-Scholes model with a modified volatility input. Instead of the constant volatility σ as we had in Equation 2.1, we now have:

$$\sigma_{t,T}^2 = \sigma^2 + T^{-1}\sum_{j=1}^{N_T^d} \sigma_j^2 \tag{3.2}$$

where t is the current time t, T is the days until expiration of the contract, and N_T^d is the number of earnings announcements between t and T. This then implies that the price of a European call option is given by the Black-Scholes formula given in Equation 2.5 with the above-modified volatility substitution.

This analysis allows us to draw the following important conclusions. In the absence of any earnings announcements, i.e., $N_T^d = 0$, we have that the above formula, Equation 3.2, simplifies back to the diffusive volatility σ^2. Consequently, the price of the option reverts to the Black-Scholes price formula without any jumps from Equation 2.5. This aligns with our intuition as without any earnings jumps ahead of us, we should simply price the option normally under the Black-Scholes model. We also observe that the volatility around earnings should climb as the announcement date approaches, peaking right before the earnings release. Subsequently, after the announcement, the volatility should then decrease back to σ^2.

The volatility model in Equation 3.2 yields results that are similar to the volatility model assumed by Patell and Watson in Equation 1.1 [18]. However, as described by Dubinsky, despite the similar implications regarding volatility around earnings announcements, there are subtle theoretical differences. Patell and Watson's approach models stock price paths as continuous with a state of elevated volatility around earnings. However, in the model above, there is a discontinuous jump in the stock price path, and it is from this discontinuity where the volatility of earnings is derived from. A notable advantage of this framework is its flexibility, allowing for a more natural integration of stochastic volatility as seen in Equation 2.4. This adaptability enhances our capacity to capture the dynamic nature of volatility.

Building on this framework, we can naturally derive two distinct estimators of σ_j^2, the earnings volatility of the jth earnings event, using the methodology outlined by Dubinsky in his paper [8]. Consider a scenario where we have two options with T_1 and T_2 days until maturity from the time t such that $T_2 > T_1$. This means that the second option has a longer time until maturity compared to the first option. Say that we have a single earnings announcement, the jth earnings event, scheduled to occur before the maturity of either option, such that $N^d_{T_1} = 1$ and $N^d_{T_2} = 1$, as seen in Figure 3.1 below:

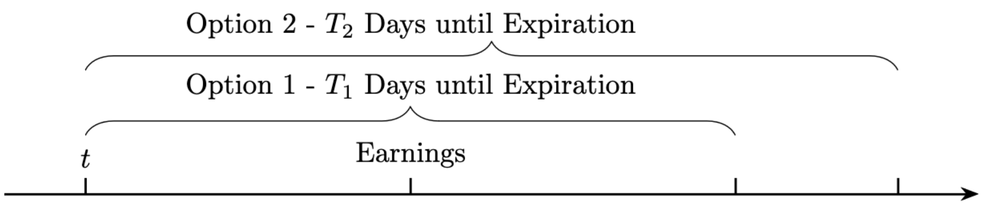

Figure 3.1: Diagram of Option Maturities for Term-Structure Estimator

Under these assumptions, using the volatility model from Equation 3.2, we have the following two equations:

$$\sigma^2_{t,T_1} = \sigma^2 + T_1^{-1}\sigma_j^2$$
$$\sigma^2_{t,T_2} = \sigma^2 + T_2^{-1}\sigma_j^2$$

Solving this system of equations for σ_j^2, we get:

$$\sigma_j^2 = \frac{\sigma^2_{t,T_1} - \sigma^2_{t,T_2}}{T_1^{-1} - T_2^{-1}} \tag{3.3}$$

This is our first estimator, which Dubinsky aptly labels as the term-structure estimator.

For the second estimator, consider the scenario where we have an option with T days until maturity, with an earnings announcement scheduled to occur after the close of trading on time t. At this point in time, the number of earnings announcements before the option's maturity is $N^d_T = 1$. However, after the earnings announcement has taken place, on time $t+1$, this count changes to $N^d_T = 0$. This can be seen in Figure 3.2:

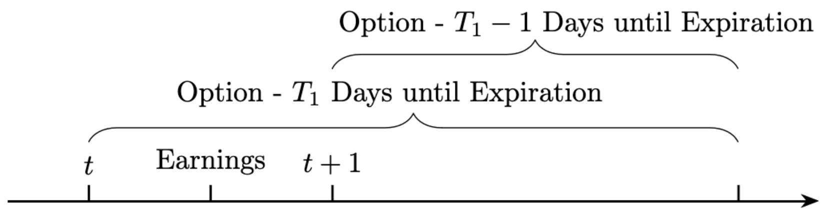

Figure 3.2: Diagram of Option Maturities for Time-Structure Estimator

As described above, the volatility of an option in the absence of any future earnings announcements defaults back to σ^2. Once again using the volatility model from Equation 3.2, we can construct the following two equations:

$$\sigma_{t,T}^2 = \sigma^2 + T^{-1}\sigma_j^2$$
$$\sigma_{t+1,T-1}^2 = \sigma^2$$

Solving this system of equations for σ_j^2, we get:

$$\sigma_j^2 = T(\sigma_{t,T}^2 - \sigma_{t+1,T-1}^2)$$

Dubinsky aptly names this estimator as the time-series estimator.

Empirically, to solve for σ_j^2, we rely on the implied volatilities derived from the market prices of the options we chose. This approach enables us to estimate the earnings volatility that is currently implied by the market, providing a direct method to gauge investor expectations regarding the upcoming earnings announcement. In our research, we encountered significant empirical difficulty in utilizing the second time-series estimator for earnings estimation, with the majority of our attempts yielding non-meaningful values. This aligns with what Dubinsky finds as well. One possible explanation for this difficulty is the time-series estimator's dependence on the implied volatility of the option after earnings, which may introduce a degree of fragility to the estimation process. Also, as Dubinsky mentions, the time-series estimator relies solely on data from a single option, whereas the term-structure estimator leverages data from two distinct options. This contributes to the robustness of the term-structure estimator, making it a more reliable tool for assessing earnings volatility. Consequently, for the estimation of σ_j^2 moving forward, we use the term-structure estimator over the time-series estimator. Throughout this paper, the terms "term-structure estimator" and "Black-Scholes Estimator" are used interchangeably, as a reference to the derivation of the term-structure estimator from the original Black-Scholes model with jumps.

The accuracy of the estimators above is predicated on the assumption that the diffusive volatility of an asset is constant. Consequently, if this assumption proves invalid — as most empirical observations of volatility have indicated — our estimators may fail to offer significant insights. However, Dubinsky in the same paper [8] demonstrates that not only is the the implied variance derived under the Black-Scholes model a good proxy for the risk-neutral variance, but that the term-structure estimator retains its robustness even when applied to stochastic, mean-reverting volatility processes. As a result, in our empirical testing of these estimators, we anticipate that the term-structure estimator will not exhibit substantial biases.

3.1.2 Stochastic Volatility Estimator

So far, the results of the estimator found in this paper are under the assumption that the diffusive volatility σ is constant. We now consider the stochastic volatility model from Equation 2.4. Once again, we employ a method that is employed by Dubinsky in his paper [8] and estimate the parameters in the Heston model using the entire time series of at-the-money (ATM) call options.

Let $C(S_t, \nu_t, \Theta, \sigma_{\tau_n}, \tau_n, K_n)$ represent the theoretical price of the call option generated from the Heston model with jumps at earnings with a strike K_n and τ_n days until maturity. The set of parameters for the stochastic volatility process is denoted by $\Theta = (\kappa, \theta, \xi, \rho)$. Let σ_{τ_n} represent the set of σ_j's that occur from current time t and until the option's expiration, defined as $\sigma_{\tau_n} = \{\sigma_j : t < j < t+\tau_n\}$. Let $C^{act}(t, \tau_n, K_n)$ be the market price of the option at time t with a strike K_n, and τ_n days until maturity. To estimate the parameters, we maximize the following log-likelihood objective function:

$$\log[\mathcal{L}(\Theta, \sigma_{\tau_n}, \nu_t)] = \\ \frac{-TN}{2}\log(\sigma_\varepsilon^2) - \frac{1}{2}\sum_{t=1}^{T}\sum_{n=1}^{N}\left[\frac{C^{act}(t, \tau_n, K_n) - C(S_t, \nu_t, \Theta, \sigma_{\tau_n}, \tau_n, K_n)}{\sigma_\varepsilon S_t}\right]^2 \tag{3.4}$$

The term σ_ε^2 is introduced to denote the variance of the pricing errors, with the assumption that it remains constant across all options. Thus we assume that the actual market price $C^{act}(t, \tau_n, K_n)$ is normally distributed around the theoretical price $C(S_t, \nu_t, \Theta, \sigma_{\tau_n}, \tau_n, K_n)$ with a variance σ_ε^2.

The objective function is designed with a twofold purpose. The first term of the objective function penalizes higher values of σ_ε^2 which pushes our optimization towards solutions that minimize the variance of pricing errors. The second term aims to minimize the differences between market prices and those predicted by the theoretical model, with these differences normalized by the stock price S_t over time T and across N different options. As Dubinsky highlights in [8], this error normalization ensures the stationarity of the data.

For our analysis, we implement the Heston model with jumps through a Monte Carlo simulation as described in Equations 2.6, which was then used to estimate the theoretical option price $C(S_t, \nu_t, \Theta, \sigma_{\tau_n}, \tau_n, K_n)$ utilized in our optimization process. In contrast to the previously discussed Black-Scholes earnings estimator, this method does not provide a time series of earnings volatilities across different quarters and instead only offers a singular aggregate estimate of σ_j that fits across all the earnings announcements. We plan to compare these estimates against the average earnings estimates we obtain from the simpler Black-Scholes estimator.

3.2 Decomposition of Earnings

Having developed a methodology for estimating earnings volatility and uncertainty, our next objective is to identify the underlying drivers behind these values. We aim to decompose the earnings volatility into various features and gain insight into how fundamental values may influence these values. This approach seeks to build intuition and give us an interpretable breakdown of what drives earnings volatility, laying a solid foundation for accurately forecasting this volatility in the upcoming section.

For this thesis, we select a set of features to investigate their impact on earnings volatility: the analyst variance (AN), the logarithm of market capitalization ($\log Mkt$), the firm age (Years), the specific quarter of the earnings announcement (Q), the market implied volatility (VIX), and previous earnings variance (σ^2_{j-k}). The features are denoted in the model by the notations provided in parentheses. Factors such as firm age and market cap already serve as traditional indicators for fundamental uncertainty within firms, owing to their clear relationship with the firm's operational stability. Also, their ease of observability further justifies their inclusion as factors in our analysis.

The perspectives of buy-side analysts, comprised of institutional investors and asset managers, often carry more weight than those of the sell-side brokers and analysts, given the buy-side's direct influence on investment decisions and market movements. However, the proprietary nature of buy-side information and the discretion practiced by the investors render access to their insights relatively challenging. As a workaround, we use sell-side analysts' reports on securities, which offer forecasts on various financial metrics such as EPS and revenue, as a proxy for buy-side opinions. A larger dispersion in these forecasts among analysts may indicate heightened uncertainty and volatility surrounding earnings.

Additionally, we utilize the VIX as a proxy for systematic variance found within the market. The VIX measures the 1-month implied volatility derived from S&P500 options and reflects the expected market volatility in the coming month. By incorporating the VIX, we aim to encompass broader market uncertainties that could influence earnings volatility.

Finally, by considering earnings volatility as part of a time series, we capture the idiosyncratic uncertainties found within a firm. Further explanation of how we systematically defined and derived these values is elaborated in Chapter 4.

In the subsequent sections, N denotes the number of data points, while D denotes the number of factors, excluding the terms related to lagged earnings volatility. In all of the proposed models, the response variable Y will be σ_j^2, the variance surrounding the upcoming earnings calculated using the Black-Scholes estimator. We set aside the estimate derived from the stochastic volatility model due to it only providing a singular average estimate of earnings volatility. This is in contrast with the Black-Scholes estimator which outputs a distinct value for each quarter. The input variable x, comprised of a vector of the identified features above, is expressed as $x = (x_1, \ldots, x_D)$.

3.2.1 Multiple Linear Model

To decompose earnings, we start with a multiple linear model. The rationale behind this approach lies in the interpretive clarity and more straightforward framework linear models provide in understanding the relationships between volatility and the influencing features through their coefficients.

The linear model is structured as follows:

$$\hat{Y}_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \beta_{p+1}^T x$$

where $\beta_0, \ldots, \beta_p \in \mathbb{R}$ and $\beta_{p+1} \in \mathbb{R}^D$. The model presumes that the predicted variable Y, in this context σ_j^2, can be represented as a linear combination of up to p lagged values of Y, in conjunction with the feature data x. The optimal number of lagged values, p, is determined by cross-validation. The model will be fit using the method of least squares, formulated as:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2$$

3.3 Predicting Earnings

We now transition to more complex, predictive models, including the LASSO method and random forest model. While these models may not offer the same level of interpretability as linear models, they hold the potential to yield more accurate predictions of earnings volatility. Achieving higher accuracy in our predictions enables us to develop trading strategies where the precision of forecasts is imperative, even at the cost of sacrificing some interpretability.

In contrast to the previous section, where our focus was on building intuition on the features influencing earnings volatility without concern for overfitting, we now divide our dataset into a training and back-testing set. The latter set serves to evaluate the efficacy of our trading strategies in Chapter 6. Given the chronological nature of our data, when splitting our data into the above sets, it is important we maintain the sequential order of the dataset and prevent the inclusion of future information when predicting, avoiding lookahead bias during the training process. Rather than employing the common approach of randomly selecting data points for each set, we organize the data into a 60-40 split, allocating the initial 60% of the data to training and reserving 40% for back-testing.

To assess the models and evaluate performance, we employ an out-of-sample testing approach. To estimate out-of-sample mean squared error (MSE) and R^2, we divide the training set into n sequential partitions. In each iteration, we incrementally expand the training set by one partition and test on the next, continuing this process until all partitions are utilized. Averaging the scores obtained from each iteration gives us our out-of-sample error scores for both MSE and R^2. These methods ensure the integrity of the temporal order of the data set. We choose our parameters for cross-validation during this step as well.

3.3.1 LASSO Method

The LASSO (Least Absolute Shrinkage and Selection Operator) method is a shrinkage method for the linear model described above. Instead of solely minimizing the squared errors as in the linear model before, a L_1 lasso penalty is also imposed. Thus, the optimization problem is formulated as:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2 + \lambda \sum |\beta_j|$$

Adjusting the regularization parameter λ can lead to the reduction of certain coefficients of the predictors to zero, thereby doing a level of continuous subset selection of the features. While the penalty term adds some level of bias to our coefficient estimates, it can also reduce the variance, potentially improving the model's generalizability and preventing overfitting. We select our regularization parameter λ using cross-validation.

3.3.2 Random Forests

Next, we develop the random forest model. The random forest model is an ensemble learning method that amalgamates a multitude of decision trees to perform regression or classification tasks.

A decision tree is a hierarchically organized structure that segments the data space into different regions based on the predictor values. At each node, a decision is made, leading to the final value that is reached after every decision.

For our analysis, we utilize the traditional Classification and Regression Trees (CART) methodology. The model can be expressed as follows:

$$\hat{Y} = \sum_{i=1}^{M} \beta_m \mathbb{1}(x \in R_m)$$

During the learning process, the model identifies the optimal splitting variables and points, partitions the data into regions, determines whether a node should terminate or undergo further subdivision, and prunes the branches as necessary.

The random forest model, by leveraging the strength of multiple decision trees, aims to enhance prediction accuracy, overcoming the tendency to overfit common with individual decision trees. Each time a split is considered on a tree, a random sample of the features is chosen as candidates for consideration from the full set of D factors, a process that decorrelates the trees with each other. After constructing B trees, the predictor is formulated as:

$$\hat{Y} = \frac{1}{B} \sum_{b=1}^{B} \hat{Y}(x; \Theta_b)$$

where Θ_b characterizes the bth tree in terms of the split, thresholds, and terminal nodes. The hyperparameters for the random tree forest such as the maximum depth, number of trees, and minimum number of leaves will be chosen using a randomized grid search cross-validation.

Chapter 4

Data Acquisition

4.1 Sourcing

The data for this section can be split into two categories: fundamental data and options data. We obtain closing options and stock prices from Ivy DB OptionMetrics [17], a frequently used data source for equity options prices in options research literature. We also obtain our fundamental values from the data source IBES [11]. The analysis utilized data covering the past decade, from February 28, 2013, to February 28, 2023.

4.2 Filtering Firms

To determine which firms to include in our analysis, we first identify the top firms based on the highest average dollar notional volume traded over the previous 90 days, representing the most liquid firms. We select the most liquid stocks to ensure adequate price discovery of option and stock prices. From here, we remove firms with a median dividend rate exceeding 1%, firms whose stock price traded below $5, firms involved in significant mergers and acquisitions (M&A), and any American Depository Receipts (ADR).

Our choice of low-dividend stocks was mostly computational, as the estimators described in the previous section assumed no dividend payments from the stocks, and a high dividend rate could potentially skew the estimations. While we could have amended our models to incorporate dividends, such adjustments would have introduced complexity that could affect option pricing, particularly regarding the early-exercise premiums in American options. As shown in Chapter 2, without dividends, the value of American and European call options should be the same. In addition, as mentioned in [8], unlike index options which model dividend payments continuously, dividends on individual equities are relatively lumpy and could further affect our model. We exclude stocks that traded below $5 due to the restricted range of available option strikes. This could lead us to analyze options that are either very in-the-money or out-of-the-money, potentially leading to implied volatility levels that are heavily skewed. Removing firms involved in major M&As and ADRs simplified our calculations as the volatility of firms in this category tends to be rather unpredictable.

From here, we choose the top 40 firms that fit these criteria. Empirically, upon analyzing the tickers DELL and AMC, we observe that the estimators are highly unstable. In DELL's case, this instability could be attributed to its recent transition back to public trading after a period of being private, a change that occurred within the last five years. Consequently, there was a lack of earnings data for us to make any meaningful conclusions. For AMC, the stock exhibited erratic trading behavior during the 2020 surge in trading, driven predominantly by retail investors coordinating through online platforms. This unusual market activity has made valuation and prediction models for AMC less reliable. To maintain the robustness of our model, we decide to exclude the two tickers from our firms, reducing the number of firms under consideration to 38. We list the selected firms in order in Table 4.1 and Table 4.2 at the end of the chapter. We observe that most of the selected companies are tech or semiconductor companies, which aligns with our selection criteria of mostly high-growth, low-dividend stocks.

4.3 Data Cleaning and Processing

4.3.1 Processing Earnings Estimator

For the earnings data, we first collect the dates of every earnings announcement for each firm and then label which quarter earnings they were.

We then collect the following data from the daily closing call options: Ticker Symbol, Best Bid Option Price, Best Offer Option Price, Strike Price, Date Traded, Expiration Date, and Closing Stock Price. The price of the option was determined to be the midpoint between the best bid and offer price provided by OptionMetrics. We calculate the implied volatilities of the options using this price. We decide to focus only on at-the-money (ATM) call options due to their high liquidity and to mitigate the effects of volatility skew that is found in deeply in-the-money or deeply out-of-the-money options. To filter for ATM options, we find every option traded on a given date and a given expiration date. We then select the option that had the strike closest to the closing stock price on that day.

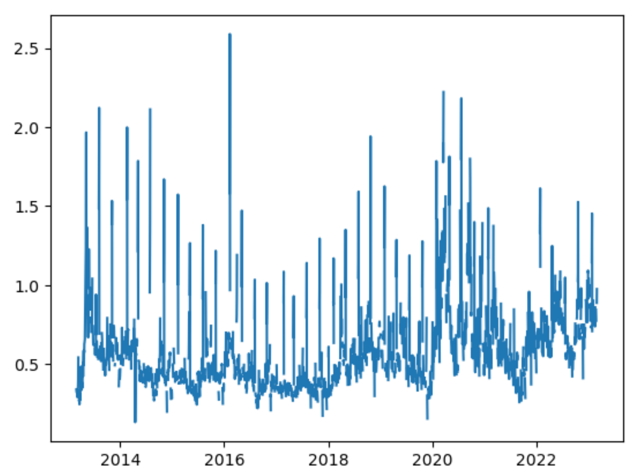

Figure 4.1: Implied Volatility of the Nearest Expiration ATM option for TSLA

In Figure 4.1, we plot the implied volatility of the nearest expiration ATM option for TSLA. As explained by our earnings estimator from Equation 3.2 in Chapter 3, we observe a gradual increase in the volatility leading up to the earnings announcement, hitting a peak right just before earnings are disclosed. This is then followed by a rapid decline in volatility after earnings.

We analyze the options of all 38 firms traded two days and seven days prior to the earnings announcement. Applying the estimator a few days before the actual earnings date enables us to obtain more stable estimator values, which are not affected by the increase in trading flow that usually precedes an announcement. This approach allows our earnings estimator to "settle" to a more definitive value. We evaluate the earnings estimator values at two distinct intervals for the following reason: The value determined 2 days before the earnings is a more immediate estimator value, which we believe reflects a more accurate representation of the market sentiment on that earnings volatility. The value determined 7 days before the earnings serves as a baseline, allowing us to track the evolution of market expectations and refine our trading strategies by comparing the two measurements. More detail is described in Chapter 6 when discussing the strategies. When computing the Black-Scholes estimator, we sort the options by maturity and select the first two options closest to maturity. Although the choice of these options was made arbitrarily, we experimented with other combinations of options and verified that they resulted in similar values of earnings volatility, confirming the robustness of our estimator.

With regards to the estimation under the stochastic volatility model, due to the computational complexity of the maximum likelihood estimation (MLE) described in Equation 3.4, we compute earnings estimates only for the following few firms: AAPL, META, TSLA, AMZN, NFLX, MSFT, AMGN, and INTC. The first five firms were chosen due to the fact that they were the most liquid stocks, and the last three were chosen so that we could compare our results to that of the firms chosen by Dubinsky in his paper [8].

When estimating the theoretical price of the option, we utilize $M = 252$ time steps to correspond with each trading day within a year and simulate $N = 1000$ stock price paths. The number of stock price paths is kept relatively low as to expedite the convergence of the optimization problem. To solve the optimization problem, we utilize Differential Evolution, a type of genetic algorithm specifically designed for optimizing real-valued multi-dimensional functions. The algorithm iteratively improves a solution by maintaining a population of candidate solutions, creating new solutions by combining existing ones, and keeping the solution that has the best "fitness" as determined by the optimization problem. Differential Evolution operates without the need for gradient information, making it particularly helpful in our case where solving for the gradient may be difficult to compute.

4.3.2 Finding Features

All features are collected seven days before an earnings announcement, ensuring that when we predict volatility for trading, we avoid the inclusion of any future data.

To calculate the log market cap, we first multiply the stock price on that day by the total number of shares outstanding to find the total market cap. We then take the logarithm of this value to compute this feature.

For the age of the firm, we calculate the number of years from the firm's Initial Public Offering (IPO) date to the current day. This approach provides a standardized measure of firm age, which is useful given the absence of clear founding dates for many of the firms under consideration.

To represent earnings quarters, we introduce three dummy variables to represent the second, third, and fourth quarters, respectively. Each dummy variable is assigned a value of 1 for observations falling within the corresponding quarter, isolating the impact of each quarter relative to the first quarter.

Additionally, we collect the closing VIX prices. To translate these values from a volatility measure into a variance measure, the VIX values are divided by 100 and squared, as a measure for market forecast variance.

For the dispersion of analyst forecasts, we first compile a list of forecasts made by analysts between consecutive earnings releases, focusing on the most recent forecast by each analyst within this period. The variance among these forecasts among the analysts serves as a measure of analyst dispersion. For example, for TSLA, with an earnings release scheduled for February 7, 2018, we examine the variance among the forecasts made by analysts from the previous earnings announcement on November 1, 2017, until January 31, 2018. The variance among the forecasts then serves as the measure of analyst dispersion for the February 7 earnings. To normalize the scale of these forecasts, we utilize the coefficient of variation:

$$\text{Coefficient of Variation} = \frac{\sqrt{\text{Variance of Forecasts}}}{\text{Mean of Forecasts}}$$

This process is done for both EPS and revenue forecasts for the upcoming quarter and year, yielding four distinct measures of analyst dispersion for each earnings announcement.

Firm	**Stock Ticker**
Tesla, Inc.	TSLA
Meta Platforms, Inc.	META
Apple Inc.	AAPL
Amazon.com, Inc.	AMZN
Netflix, Inc.	NFLX
Advanced Micro Devices, Inc.	AMD
Zoom Video Communications, Inc.	ZM
Moderna, Inc.	MRNA
PayPal Holdings, Inc.	PYPL
Alphabet Inc.	GOOGL
Microsoft Corporation	MSFT
NVIDIA Corporation	NVDA
Enphase Energy, Inc.	ENPH
DocuSign, Inc.	DOCU
Coinbase Global, Inc.	COIN
The Boeing Company	BA
Booking Holdings Inc.	BKNG
Adobe Inc.	ADBE
SolarEdge Technologies, Inc.	SEDG
Oracle Corporation	ORCL

Table 4.1: List of Firms with Corresponding Stock Tickers (1-20).

Firm	**Stock Ticker**
Super Micro Computer, Inc.	SMCI
Palantir Technologies Inc.	PLTR
Walmart Inc.	WMT
Snowflake Inc.	SNOW
Etsy, Inc.	ETSY
Upstart Holdings, Inc.	UPST
Palo Alto Networks, Inc.	PANW
C3.ai, Inc.	AI
Salesforce, Inc.	CRM
Qualcomm Inc.	QCOM
Applied Materials, Inc.	AMAT
Amgen Inc.	AMGN
Cisco Systems, Inc.	CSCO
The Home Depot, Inc.	HD
International Business Machines Corporation	IBM
Intel Corporation	INTC
Texas Instruments Incorporated	TXN
Micron Technology, Inc.	MU

Table 4.2: List of Firms with Corresponding Stock Tickers(21-38).

Chapter 5

Empirical Results on Earnings

5.1 Descriptive Statistics

5.1.1 Black-Scholes Earnings Estimator

For each firm, the earnings estimations are computed using Equation 3.3 and following the methodology described in the previous chapter. For readability, when discussing the values of the earnings estimator, we present these in volatility units, formatted as $100 \times \sigma_j$. In the interest of conciseness, for firm-specific analysis, we focus on the first ten stocks, in addition to a pooled assessment of all 38 tickers. The compiled statistics summarizing the earnings volatility are detailed in Table 5.1.

During the empirical estimation, we encountered instances where $\sigma_{t,T_1} < \sigma_{t,T_2}$ which contradicted our assumptions used in modeling our estimators. This discrepancy predominantly occurred with companies experiencing lower trading volumes on those days, suggesting liquidity issues and inadequate price discovery as the probable cause.

From these summary statistics, we make the following observations. Zoom (ZM) exhibited the highest mean earnings volatility among the analyzed firms at 14.6%, alongside one of the highest volatility for a single earnings announcement at 20.7%.

Firm	**Count**	**Mean**	**SD**	**Min**	**Max**
TSLA	39	10.0	2.8	5.3	15.9
META	40	8.5	2.9	5.0	15.9
AAPL	40	5.1	1.3	2.0	8.6
AMZN	40	7.3	2.0	4.2	10.7
NFLX	40	11.4	2.8	7.1	20.4
AMD	38	11.7	4.1	5.4	23.8
ZM	15	14.6	3.6	9.7	20.7
MRNA	13	9.1	4.2	4.1	20.8
PYPL	30	7.2	2.6	2.1	14.5
GOOGL	35	5.6	1.3	1.9	8.6
Pooled	903	7.3	3.8	0.6	33.6

Table 5.1: Summary Statistics of Earnings Volatility Estimates

This heightened volatility can likely be attributed to the COVID-19 pandemic, which due to a shift towards remote work, increased demand for Zoom's services. This led to increased uncertainty among investors regarding Zoom's future performance forecasts and growth projections. However, it is important to note that Zoom's relatively recent entry into the market resulted in fewer data points for analysis compared to other firms above.

Conversely, Apple (AAPL) recorded the lowest mean volatility at 5.1%, alongside the smallest standard deviation at 1.3%, and one of the lowest minimum estimates at 2.0%. These values suggest that AAPL experiences comparatively lower levels of uncertainty around its earnings than other firms. One plausible explanation for this phenomenon is AAPL's approach of disclosing product information through public showcases and digital publications which may result in a more informed investor base, and consequently, less uncertainty before earnings announcements.

The time series of earnings estimates across the firms is illustrated in Figure 5.1. The yearly mean earnings estimates are shown in Table 5.2. This figure reveals a noticeable increase in volatility for numerous firms around 2020, coinciding with the onset of the pandemic. This connection is logical, given the heightened overall uncertainty regarding market and firm performance during that period. This is also

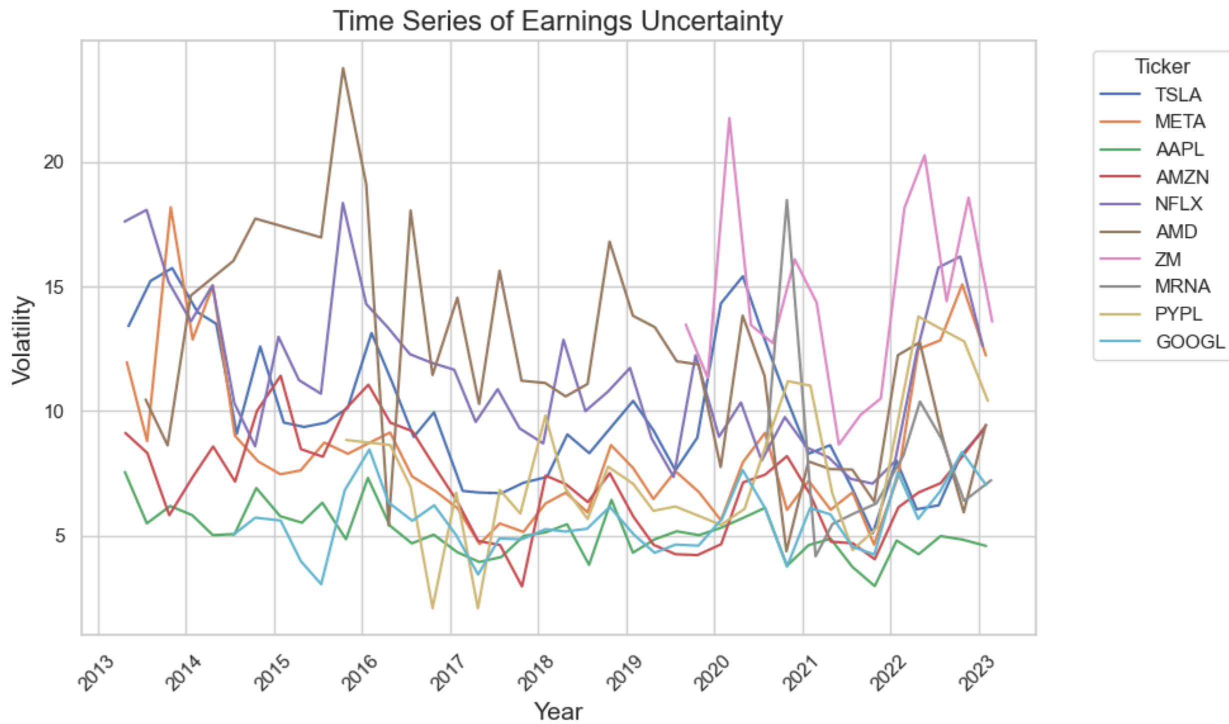

Figure 5.1: Time Series of Earnings Volatility across Firms

further supported by Table 5.2 where we can see a slight jump in mean estimates in 2020 and 2021.

Analyzing the Predictive Power of σ_j

With these estimates, we then delve into the predictive content of the volatility metric σ_j, aiming to examine if a high σ_j predicts if the resulting earnings realizations are more volatile. To do so, our approach involves calculating the magnitude of stock returns, denoted as $|R|$, from right before the earnings announcement and right after the earnings announcement. This value could be thought of as how much the earnings volatility "realized". Subsequently, we record the correlation between a given σ_j and its corresponding absolute return $|R|$ in Table 5.3, conducting this analysis for both firm-specific cases and on a pooled basis.

Interestingly, the firm-specific correlations are generally not statistically significant, and when they are statistically significant, the correlations are negative. This

Firm	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**
TSLA	11.0	6.9	9.4	8.9	13.4	7.0	8.0
META	7.8	5.5	7.1	7.4	7.0	7.1	12.5
AAPL	5.4	4.2	4.6	5.1	4.7	4.1	5.9
AMZN	9.1	5.0	7.5	5.3	6.2	4.9	7.9
NFLX	12.5	9.9	10.7	9.8	10.6	7.8	12.8
AMD	12.7	13.1	12.3	12.7	10.3	7.1	7.9
ZM	—	—	—	14.2	13.9	11.3	18.4
MRNA	—	—	—	9.5	10.2	7.0	10.5
PYPL	5.9	5.1	6.2	6.0	6.8	7.2	11.9
GOOGL	6.3	4.5	5.8	5.0	5.4	4.5	7.3
Pooled	7.2	5.8	6.4	6.3	7.6	6.6	10.1

Table 5.2: Yearly Mean Earnings Volatility Estimates

Firm	**Correlation**	
	All Data	**Before 2020**
TSLA	0.13	0.28
META	0.14	0.26
AAPL	-0.29*	-0.47**
AMZN	-0.01	0.02
NFLX	0.28*	0.24
AMD	-0.07	-0.17
ZM	0.05	—
MRNA	-0.27	—
PYPL	-0.03	-0.46*
GOOGL	-0.32*	-0.19
Pooled	0.05*	0.15***

Table 5.3: Correlation between σ_j and $|R|$
$^{*}\ p < 0.1, ^{**}\ p < 0.05, ^{***}\ p < 0.01$

suggests that on a firm-wide basis, higher σ_j did not necessarily correspond with larger earnings realizations. On a pooled basis, the correlation is 5.1%, exhibiting significance at the 10% significance level. This phenomenon could be due to several factors, most notably the diversification effect, wherein in a broader dataset, the effects of individual firms may cancel out, and the overall trend between earnings volatility and absolute returns becomes more pronounced.

These results are somewhat unexpected as Dubinsky, in his paper [8], reports positive and strongly significant correlations for firms like AAPL at the 1% level, along with a pooled correlation of 28%, which was also highly significant. Not only were our correlations less statistically significant, but the levels of correlation were also lower. This suggests that the predictive power of earnings volatility may have diminished over time since 2004 when Dubinsky did his analysis. A plausible explanation could be that the surge of retail investors in addition to the COVID-19 pandemic may have led to more unpredictable earnings during the 2020s. Further analysis of earnings announcements before March 2020, before the pandemic's onset, reveals that the pooled correlation increases to 15.1%, which is highly significant and aligns closer with Dubinsky's earlier results.

5.1.2 Stochastic Volatility Earnings Estimator

Now, we seek to see how our estimation results change under the stochastic volatility model presented by maximizing the MLE in Equation 3.4. We note our results in Table 5.5. We once again remind what each parameter is in Table 5.4

Parameter	**Description**
κ	Rate of Reversion to Mean Volatility
θ	Mean Reverting Value of Volatility
ξ	Volatility of Volatility
σ_j	Earnings Volatility
σ_ϵ	Standard Deviation of Pricing Error

Table 5.4: Factor Notation

The fourth column of Table 5.5 provides the average estimate of σ_j for each firm. To frame our results, we provide both the stochastic estimator's results and the average estimate from the term-structure estimator, under the modified Black-Scholes model, in Table 5.6. The estimates below are once against expressed in volatility units as before.

In general, the earnings volatilites derived from both models are relatively aligned, with only NFLX displaying significant discrepancy. The volatility estimates under the

Firm	κ	θ	ξ	σ_j	σ_ϵ
AAPL	3.00	0.44	0.63	0.06	0.01
META	2.90	0.47	0.51	0.08	0.01
TSLA	3.00	0.37	0.25	0.10	0.02
AMZN	2.41	0.44	0.67	0.07	0.01
NFLX	2.87	0.48	0.58	0.1	0.01
MSFT	2.18	0.20	0.52	0.06	0.01
INTC	2.90	0.39	0.75	0.06	0.8
AMGN	2.99	0.46	0.28	0.04	0.00

Table 5.5: Parameter Estimates for Firms under Stochastic Volatility with Jumps

Firm	Stochastic Estimator	Mean B-S Estimator
AAPL	5.7	5.1
META	8.0	8.5
TSLA	9.8	9.8
AMZN	7.1	7.1
NFLX	9.6	11.5
MSFT	5.6	4.6
INTC	6.1	5.8
AMGN	3.9	3.6

Table 5.6: Comparison of the Two Earnings Estimates for Each Firm

stochastic model tend to be slightly higher. Dubinsky in [8] suggests that the differences in these values arise from two factors: first, the term-structure estimator relies on only two options, whereas the MLE estimation incorporates every option that is influenced by the earnings announcement; second, the stochastic volatility model assumes constant parameters throughout the entire time series, while the term-structure estimator allows variations in volatility from one earnings to the next. Consequently, a direct comparison of the values may not be entirely appropriate.

When comparing our results to those of Dubinsky's, we observe that our uncertainty values for INTC and AMGN closely matched, but we see noticeably different values for AAPL and MSFT. Specifically, Dubinsky records values of 8.5% and 3.9% for AAPL and MSFT respectively, whereas we obtain values of 5.7% and 5.6% for the same firms. We suggest several reasons for these discrepancies. Intel (INTC) and Amgen (AMGN)'s business models have remained relatively consistent from the early 2000s to now. In contrast, since the early 2000s, Apple (AAPL) has transitioned from a computer manufacturer to a more diversified consumer electronics company. During Dubinsky's study, Apple had just recently launched the iPod, and the iPhone and iPad had yet to be released, and therefore, earnings were more vulnerable to the outcomes of a few product lines. This may have led to the higher value that Dubinsky records. As Apple matured and diversified its product base, in addition to the product showcases Apple holds, the volatility surrounding earnings has dropped. Meanwhile, Microsft (MSFT) has undergone significant technological shifts since the early 2000s moving away from a stable business model focused on software sales, to embracing advancements in mobile computing, artificial intelligence, cloud computing, and subscription-based services. This shift in strategy has led to increased volatility surrounding earnings in the rapidly changing tech landscape.

5.2 Decomposition Results

5.2.1 Results and Discussion of Linear Model

Using the multiple linear model discussed in the previous section, we present the coefficients along with their statistical significance in Table 5.8. In this table, we provide everything in terms of percentage values for readability. In Table 5.7, we once again provide the notation for which we will be describing our features in the following analysis. The analysis was conducted for both firm-specific models and a pooled model aggregating data from all 38 firms. Consistent with our earlier approach, we choose to include only the firm-specific results for the biggest five companies, given that the firm-specific results as a whole tended to be less statistically significant than the pooled model.

Several key insights emerge from our analysis. First, we observe that the biggest predictor of earnings volatility is the previous quarter's earnings volatility. The coefficient of this factor within the pooled regression is 35.9% and highly significant.

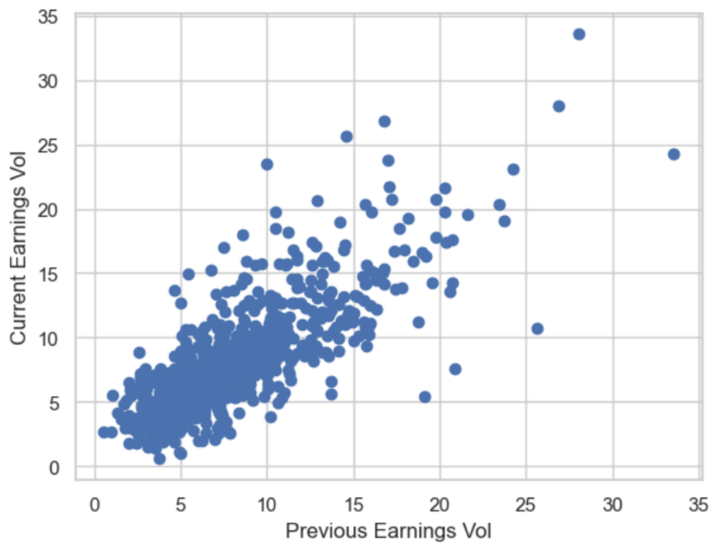

Figure 5.2: Previous Earnings Volatility vs. Current Earnings Volatility

In Figure 5.2, the previous quarter's earnings volatility σ_{j-1} and the current earnings volatility σ_j are plotted against each other. This visual analysis reinforces the statistical findings, illustrating a clear trend where increased earnings volatility in the previous quarter is associated with heightened earnings volatility in the current quarter. This suggests that the uncertainty experienced in one earnings can carry over and influence expectations in subsequent announcements.

We then examine the impact of the analyst forecast variance features on the earnings uncertainty. Surprisingly, the variance around analysts' forecasts appears to exert minimal influence on the uncertainty. Many of the coefficients turn out to be statistically insignificant, or in cases when significance is observed, the relationship is often negative, posing a counterintuitive scenario. This suggests that greater divergence in analysts' predictions about fundamental values translated to reduced uncertainty regarding earnings outcomes. Out of the pooled model, we see that the sole significant variable was the variance in revenue (SAL) forecasts for a given year, denoted by $AN_{SAL,Y}$. This emphasis on revenue as a key fundamental value is logical, given that many of the analyzed firms are valued based on the EV/Sales multiple. Many of the analyzed firms are characterized by high growth and a lack of profitability, and so investors often prioritize revenue growth over profitability in their valuation. Further analysis must be done to explain these results. One potential explanation is that our measure of analyst dispersion may not accurately capture the true variance of opinions over time. For example, one method could be to look into the variance of opinion over time before the earnings quarter for the same analyst as a time series. If an analyst is constantly changing their forecast, this may signal there is uncertainty in how the earnings may turn out.

The analysis of earnings uncertainty across the different quarters yields an interesting result. Initially, we hypothesized that both quarters 3 and 4 would exhibit heightened earnings uncertainty over the other two quarters. Quarters 3 and 4 announcements are closer to the holiday season, leading to a boost in sales which could result in greater volatility on earnings as there is a broader range of earnings outcomes that a firm could report. Additionally, the fourth quarter serves as the end of the fiscal year for many firms, where a company usually summarizes its current year and discusses future guidance and plans for the upcoming fiscal year. Therefore, there tends to be more information that is usually disclosed in the fourth quarter report compared to the other earnings announcements, which may lead to investors pricing in more volatility. However, when we observe the results, it appears that the quarter information is not very statistically significant.

We find that the other factors align with intuitive expectations. The coefficients associated with the market cap factor are predominantly negative, suggesting that larger companies experience less earnings volatility. This trend is logical as bigger firms tend to have more diversified operations and streams of cash flow, and therefore are less susceptible to fluctuations around earnings. We also observe that the current volatility in the market, measured by the VIX, is also a significant predictor of individual earnings volatility. When there is higher volatility in the market, earnings also tend to be more volatile as indicated by the positive coefficients from these results. Both of these results are also highly significant in the pooled model and among the firm-specific models as well. We also see that the number of years since the firm was established is also a significant predictor. In the pooled model, the negative coefficient associated with the firm's age suggests that older firms experienced lower volatility. This is in line with expectations and how other studies have treated this value. Many studies in the field use the age of a firm as a proxy for uncertainty in a firm, with the expectation that newer and younger companies exhibit more earnings volatility.

Notation	**Description**
σ^2_{j-k}	Earnings Estimator from kth Previous Quarter
$\text{AN}_{\text{EPS, Y}}$	Analyst Dispersion around EPS for the Yearly Forecast
$\text{AN}_{\text{EPS, Q}}$	Analyst Dispersion around EPS for the Quarterly Forecast
$\text{AN}_{\text{SAL, Y}}$	Analyst Dispersion around Revenue for the Yearly Forecast
$\text{AN}_{\text{SAL, Q}}$	Analyst Dispersion around Revenue for the Quarterly Forecast
VIX	S&P500 1M Implied Volatility from VIX
$\log \text{Mkt}$	Log of the Market Cap of the Company
Years	Number of Years since IPO
Q2	Indicator Variable for Q2
Q3	Indicator Variable for Q3
Q4	Indicator Variable for Q4

Table 5.7: Notation of Features

5.3 Prediction Results

In this section, given our focus on using these models for prediction and trading, and considering the superior average performance of the pooled models as seen in the previous section, from this section forward, we exclusively detail the coefficients and performances from our pooled models only. For the out-of-sample testing and cross-validation, we split the training set into $n = 30$ partitions.

Feature	TSLA	META	AAPL	AMZN	NFLX	Pooled
const	2.4	27.0***	-0.2	8.3***	8.5***	2.3***
	(2.8)	(6.3)	(1.7)	(2.3)	(2.7)	(0.3)
σ^2_{j-1}	20.8	3.4	8.7	70.4***	23.2	35.9***
	(17.7)	(15.7)	(19.9)	(17.9)	(17.3)	(2.8)
σ^2_{j-2}	6.3	14.4	-22.2	-9.5	-20.6	27.5***
	(18.0)	(13.0)	(19.2)	(21.7)	(18.3)	(3.0)
σ^2_{j-3}	10.9	-31.6**	-15.1	9.2	15.9	-0.8
	(21.4)	(12.4)	(21.0)	(17.5)	(15.9)	(2.9)
$\text{AN}_{\text{EPS, Y}}$	0.0	0.3	-1.3	-0.0	0.6	0.0
	(0.0)	(1.8)	(2.6)	(0.0)	(0.8)	(0.0)
$\text{AN}_{\text{EPS, Q}}$	0.1	4.3*	-0.2	-0.0	0.0	-0.0
	(0.1)	(2.4)	(1.5)	(0.0)	(0.5)	(0.0)
$\text{AN}_{\text{SAL, Y}}$	3.6	1.1	-1.1	7.6***	1.5	1.4***
	(2.2)	(2.6)	(3.8)	(2.7)	(7.0)	(0.4)
$\text{AN}_{\text{SAL, Q}}$	2.8	-8.8**	0.4	-1.6	-8.8	0.2
	(2.1)	(3.7)	(2.3)	(2.1)	(9.4)	(0.3)
VIX	5.9*	-3.9*	1.5*	2.6**	8.4***	3.4***
	(3.2)	(2.2)	(0.8)	(1.1)	(2.7)	(0.5)
log Mkt	-0.1	-1.4***	0.1	-0.6***	-0.5**	-0.1***
	(0.2)	(0.3)	(0.1)	(0.2)	(0.2)	(0.0)
Years	-0.0	0.3***	-0.0	0.1***	0.1	-0.0***
	(0.1)	(0.1)	(0.0)	(0.0)	(0.1)	(0.0)
Q2	0.1	-0.1	-0.1	0.1	0.0	0.0
	(0.3)	(0.2)	(0.1)	(0.1)	(0.2)	(0.1)
Q3	0.2	-0.0	-0.1	0.2	0.3	0.0
	(0.3)	(0.2)	(0.1)	(0.2)	(0.2)	(0.1)
Q4	0.4	-0.4*	-0.0	-0.4*	-0.2	-0.0
	(0.2)	(0.3)	(0.1)	(0.2)	(0.5)	(0.1)

Table 5.8: Coefficients (%) of the First 5 Firms and Pooled
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ Standard errors are presented in parentheses underneath each coefficient.

5.3.1 LASSO

With the LASSO model, we initially normalize our features to achieve a mean of 0 and a standard deviation of 1. Through the cross-validation process, we select a regularization parameter of $\lambda = 0.000215$. The performance of our LASSO model is assessed through its Mean Squared Error (MSE) and R^2 score which are reflected in Table 5.9. On the entire training set, the model achieves an MSE of approximately 1.032×10^{-5}, while on the out-of-sample data, the model achieves an average MSE of 1.2×10^{-5}. The model indicates a high level of explanatory power with an R^2 score of 0.60 on the training set and an out-of-sample R^2 score of 0.52. The R^2 scores, representing the proportion of variance in the dependent variable explained by the predictors, affirmed the model's strong predictive capabilities.

Metric	**Training Set**	**Out-Of-Sample**
Mean Squared Error (MSE)	1.0×10^{-5}	1.2×10^{-5}
R^2	0.60	0.52

Table 5.9: LASSO Model Performance Metrics

Feature	**Coefficient**
const	0.51
σ^2_{j-1}	0.31
σ^2_{j-2}	0.1
σ^2_{j-3}	0.12
$\text{AN}_{\text{EPS, Y}}$	0.0
$\text{AN}_{\text{EPS, Q}}$	0.0
$\text{AN}_{\text{SAL, Y}}$	0.02
$\text{AN}_{\text{SAL, Y}}$	0.0
VIX	0.07
$\log \text{Mkt}$	-0.09
Years	-0.0
Q2	0.0
Q3	0.0
Q4	0.0

Table 5.10: Coefficients (%) of the LASSO Model

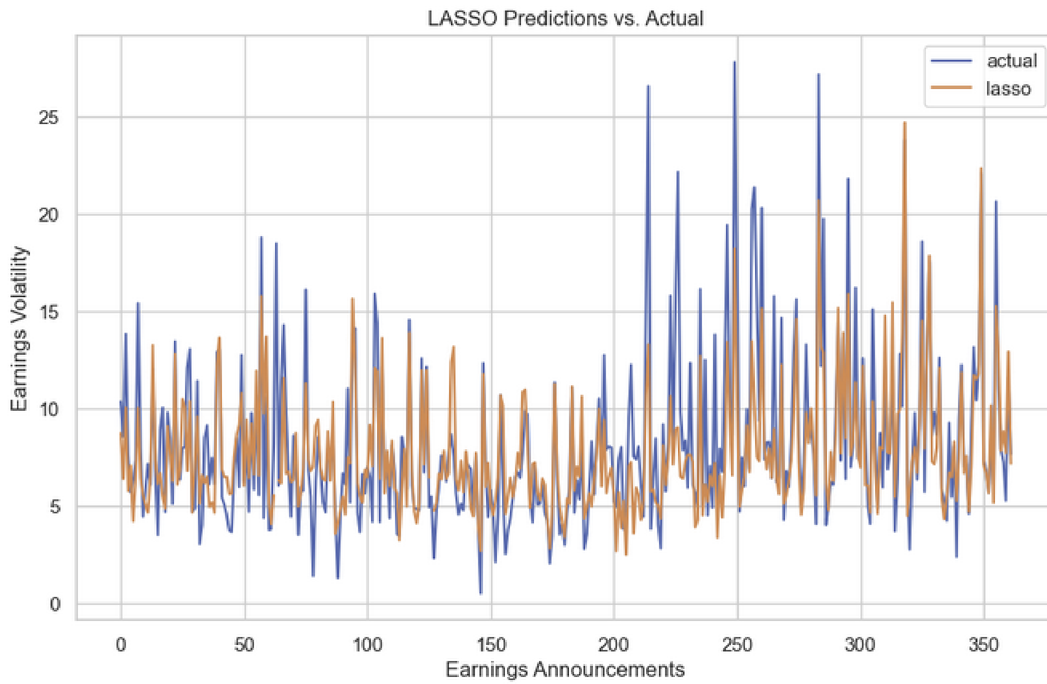

Figure 5.3: Plot of LASSO Predictions and Actual

Furthermore, when observing the coefficients in Table 5.10, we see that several features have been reduced to zero, illustrating the ability of LASSO to perform automatic feature selection. The LASSO model simplified the model complexity and highlighted the features that have the most predictive value. These are in line with our expectations from our earlier discussion on the decomposition of earnings. This method not only further enhances the interpretability around earnings but also improves the model's performance on unseen data. We plot the predicted values and actual values together across time in Figure 5.3 and plot the predicted against the actual in Figure 5.4. The analysis of the scatter plot indicated that generally, the LASSO model neither systematically underpredicted nor overpredicted the volatilities.

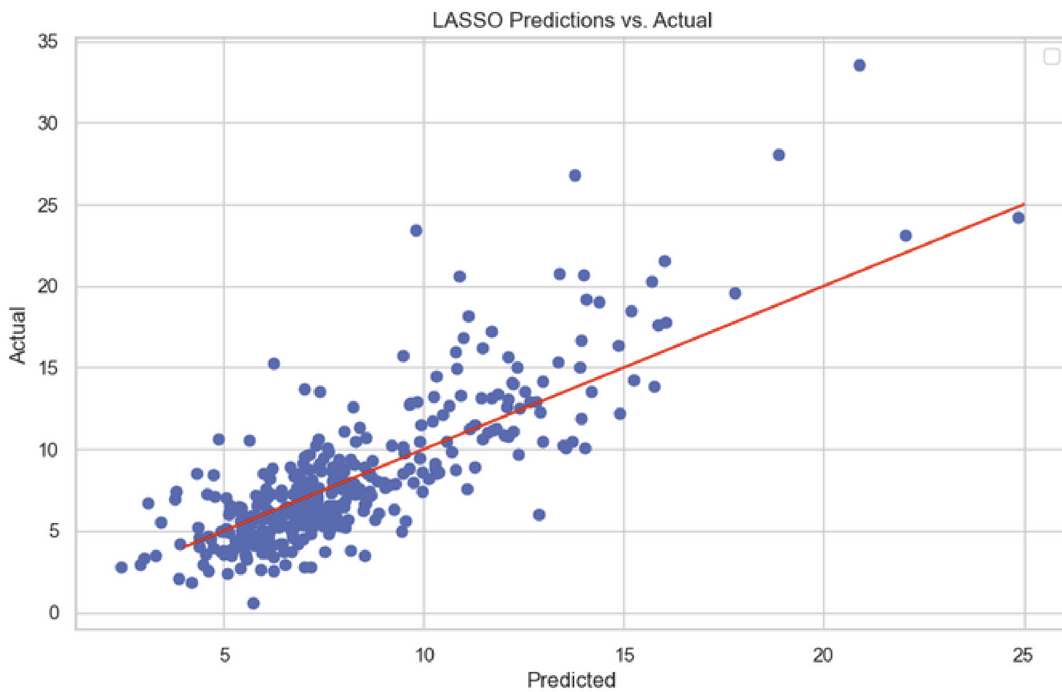

Figure 5.4: Plot of LASSO Predictions against Actual

5.3.2 Random Forest

We do a similar test of performance using the random forest model. Through the randomized search cross-validation process, we select the following hyperparameters for our model: $1,000$ trees, a minimum of 5 samples required to split an internal node, a minimum of 1 sample required at a leaf node, and a maximum depth of 5 per tree.

After training our model, we extract feature importance from our random forest model which is then presented in Figure 5.5. We can clearly observe that the estimators lagged by one quarter and two quarters emerge as the most critical features among all the predictors. Following these, there is a significant decline in importance for the subsequent features. Notably, unlike the other lagged variables, the estimator lagged by three quarters exhibits much less importance. This suggests that the earnings information from the last two quarters plays a more pivotal role in forecasting current volatility, as opposed to information from earlier periods.

Once again, we assess the performance of our random forest model through the

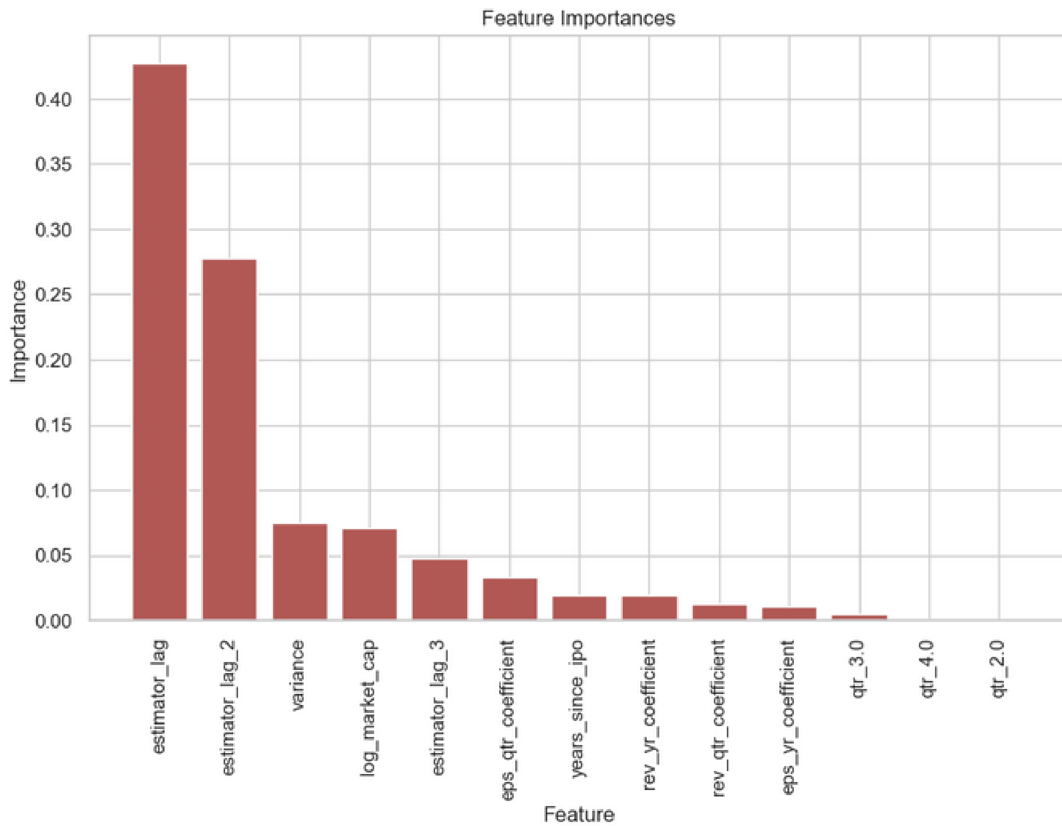

Figure 5.5: Feature Importance from our Random Forest Model

MSE and R^2 score, with the results detailed in Table 5.11. We see that the random forest model achieved a higher R^2 score of 0.85 and lower MSE of 3.8×10^-6 on the entire training set compared to the LASSO model. However, we see that the out-of-sample performance was relatively similar between the two models on both metrics. When we plot the predicted values and actual values in Figure 5.6, we observe that the random forest model has a tendency to underestimate the spikes in volatility. This becomes more apparent in Figure 5.7 where an increased number of points fall above the line as we move leftward, highlighting instances of larger volatility spikes that the model failed to accurately predict.

The comparison of the performance scores between the two models reveals that the two models performed relatively comparably, a somewhat surprising result given the random forest model's reputation for predictive accuracy and its ability to capture

Metric	**Training Set**	**Out-Of-Sample**
Mean Squared Error (MSE)	3.8×10^{-6}	1.2×10^{-5}
R^2	0.85	0.52

Table 5.11: Random Forest Model Performance Metrics

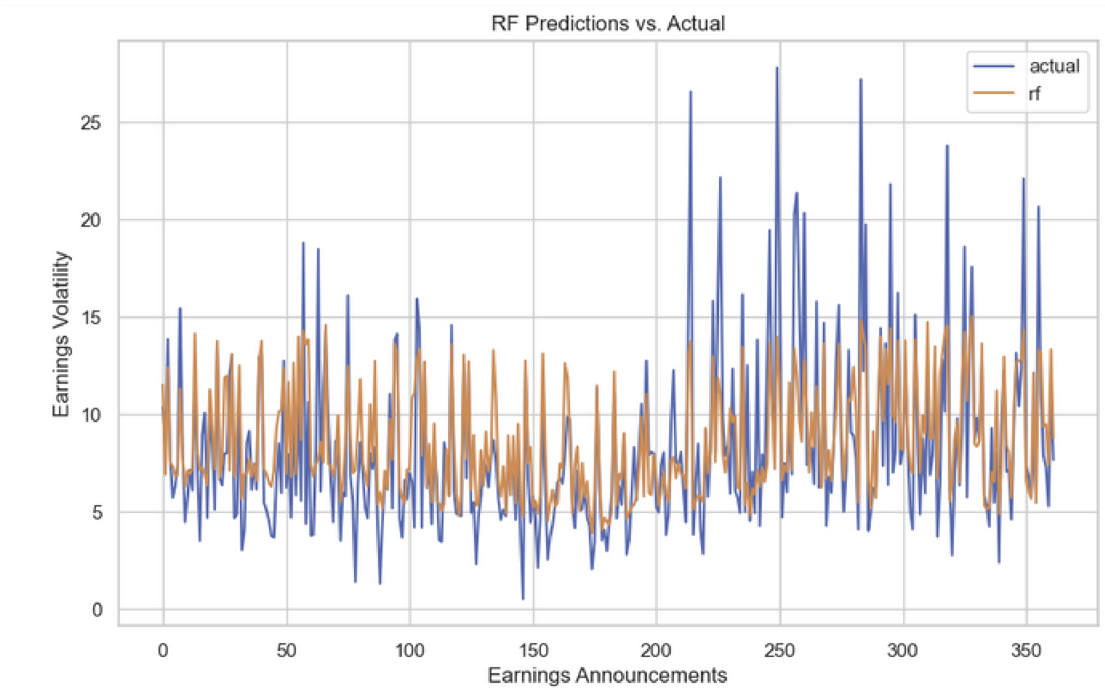

Figure 5.6: Plot of Random Forest Predictions and Actual

complex non-linear relationships. This performance may be attributed to the limited number of features we have available, which restricts the complexity of trees that can be developed in the random forest model. Consequently, a parameterized model like LASSO manages to achieve comparable or even superior performance under these conditions. Given their similar performance, we opt for the simpler LASSO model for our trading strategies in the next chapter.

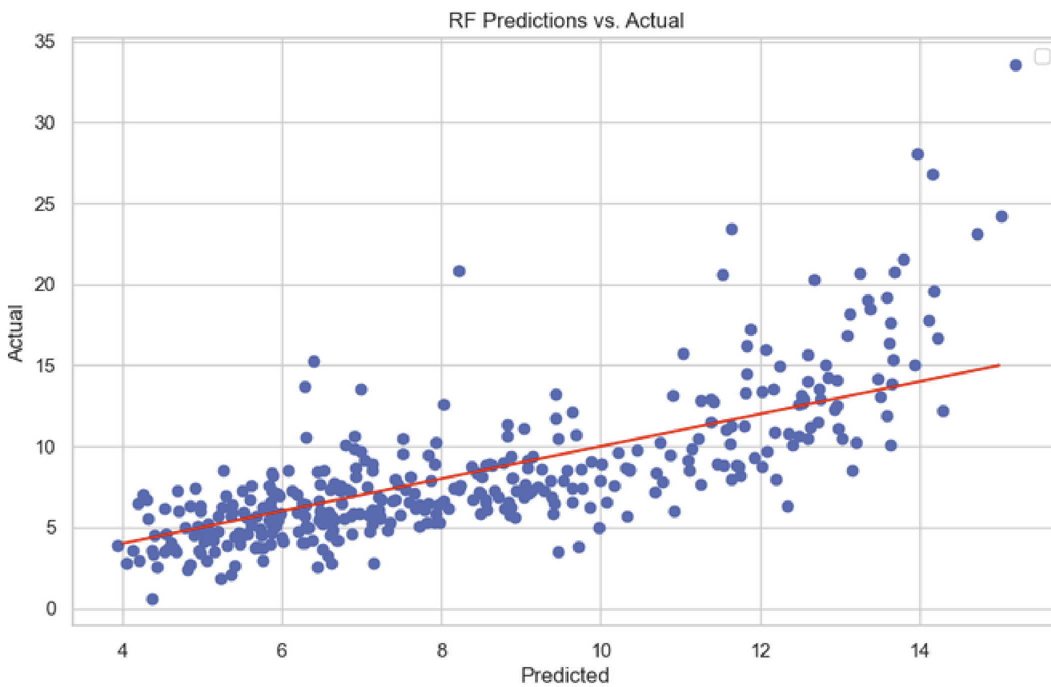

Figure 5.7: Plot of Random Forest Predictions against Actual

Chapter 6

Trading Strategies

6.1 Greeks

Before discussing option trading strategies, it is essential to first begin with an understanding of the Greeks. The Greeks are a set of measures used to assess the level of risk of an option in the options market. A Greek symbol is assigned to each risk, and each measures the sensitivity of the price of the option to one or more underlying parameters. In this thesis, our primary focus will be on the following Greeks: Delta (Δ), Gamma (Γ), and Vega (ν).

6.1.1 Delta (Δ)

Delta (Δ) measures the rate of change of an option's price with respect to a $1 change in the price of the underlying asset. A positive delta indicates a desire for a positive movement in the spot price, while a negative delta indicates a desire for a negative movement in the spot price. For call options, delta values typically range from 0 to 1 with the value becoming more positive as the option becomes more in-the-money (ITM). For put options, delta values range from 0 to -1 with the value becoming more negative as the options become more ITM. Besides measuring price sensitivity, delta also serves as an approximation for the option's probability of expiring ITM [16]. Therefore, an at-the-money (ATM) option typically has a delta whose magnitude is around 0.5, indicating a roughly equal chance of ending ITM or out-of-the-money (OTM). The stock by definition also has a delta of 1. Below is a figure depicting how an option's delta changes in relation to the stock price:

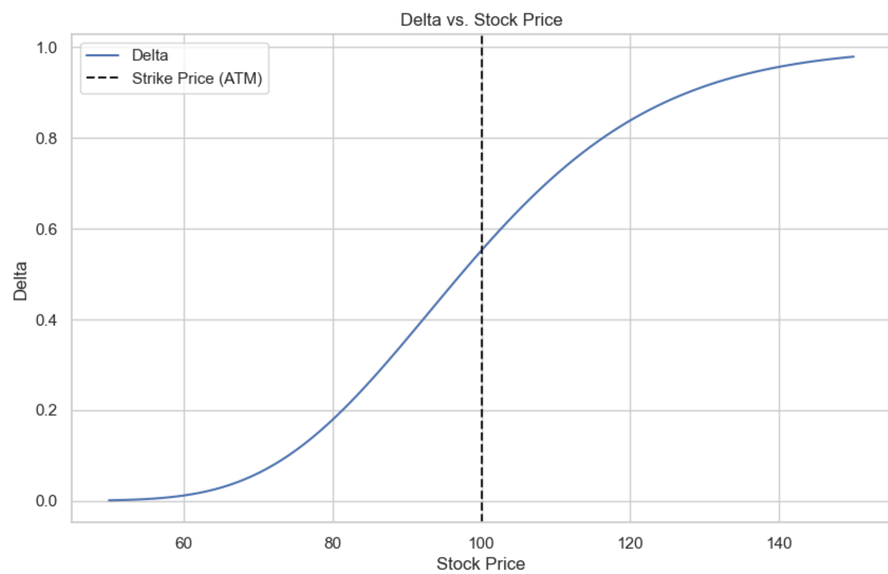

Figure 6.1: Delta vs. Stock Price under Black-Scholes Model

6.1.2 Gamma (Γ)

Gamma (Γ) measures the rate of change of Δ with respect to a \$ 1 change in the price of the underlying asset, reflecting the convexity of an option's value. For any long position, gamma is always positive and typically reaches its peak value for ATM options as these options are most sensitive to changes in the underlying asset price. Gamma plays a role in relation to the underlying asset's realized volatility. Specifically, when the underlying asset exhibits significant realizations of volatility and sharp price fluctuations, gamma can enhance a portfolio's value, notably in strategies that involve delta-hedging. With rapid price fluctuations and a high gamma, the delta of the option changes rapidly, which can be addressed by adjusting our delta-hedge to capture gains from these movements. Below is a figure depicting how an option's gamma changes in relation to the stock price:

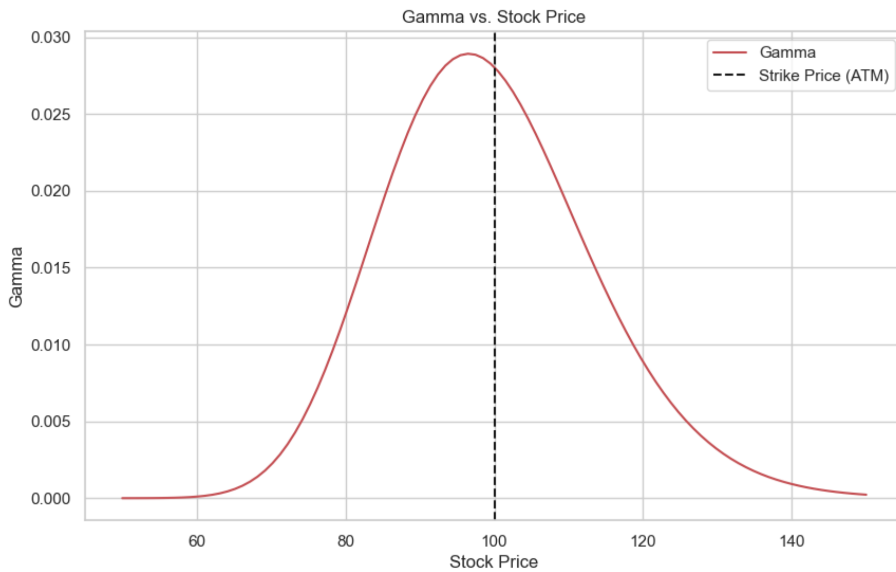

Figure 6.2: Gamma vs. Stock Price under Black-Scholes Model

6.1.3 Vega (ν)

Vega (ν) measures the rate of change of the option's value with respect to a 1% change in the implied volatility. For any long position, vega is always positive, meaning that the option's value increases as implied volatility increases. Like gamma, vega typically reaches its peak for ATM options. Below is a figure depicting how an option's vega changes in relation to the stock price:

Vega and gamma play an integral role in any volatility trading strategy. A positive vega allows us to capture the changes in the expectation of volatility. Similarly, having positive gamma, while maintaining a delta-neutral position, enables us to benefit from the convexity of the option price. This allows us to capture the value created from the realized volatility. When realized volatility is high and the asset price moves significantly, then we can rebalance our portfolio to realize a profit.

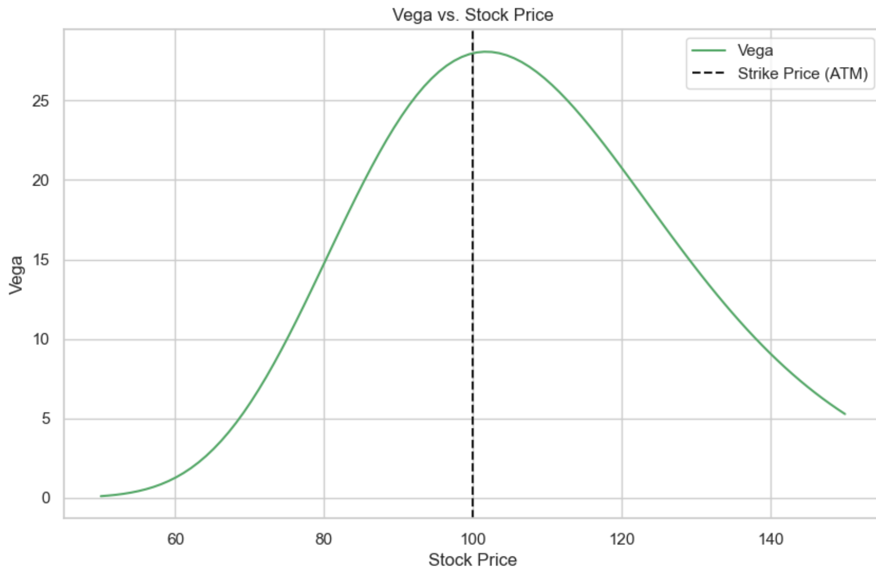

Figure 6.3: Vega vs. Stock Price under Black-Scholes Model

6.2 Option Structures

In the following section, we describe two option structures, the straddle and strangle, that we utilize in our trading strategies. Before we do so, we first adopt specific terminology that goes along with trading options. From here on, we refer to our actions as either taking a "long" or "short" position.

Holding a long position in an option signifies that you own the option contract. For example, if you are long a call option, this gives you the right to buy a stock at a certain price, and if you are long a put option, this gives you the right to sell a stock at a certain price. On the other hand, taking a short position in an option signifies that you have written or sold the option contract. For example, if you are short a call option, you are obligated to sell the stock at a certain price if the counterparty exercises the option, and if you are short a put option, you are obligated to buy the stock at a certain price if the counterparty exercises the option.

Adopting this terminology will help mitigate any confusion about the action that is being taken. For instance, you could be long a put option by "buying" the option contract on the market, which could ultimately lead to "selling" the underlying stock upon exercise. Therefore, to maintain clarity and consistency, we will use the terms "long" and "short". When detailing the structures below, we will describe them from the perspective of holding a long position. However, each structure can be adapted to a short position by inverting each leg of the trade.

6.2.1 Straddle

A long straddle involves being long both a call option and a put option on the same underlying stock with the same strike price and expiration date. The payoff diagram of a straddle at expiration is shown below:

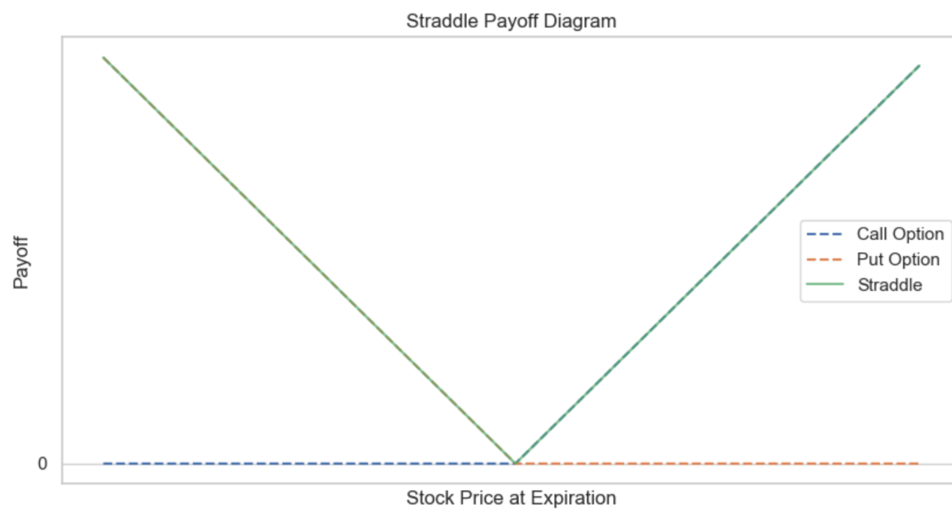

Figure 6.4: Straddle Payoff

In the case of our trading strategies, we choose our strike to be ATM. We focus on this structure and strike for the following reasons:

- Delta (Δ): The delta of an ATM straddle when initially bought is near zero as the delta of the ATM call and the delta of the ATM put offset each other. Therefore, the value of this structure is insensitive to small price movements.
- Gamma (Γ): The gamma of an ATM straddle is high as the gamma is close to its peak for ATM options as shown above. This means that the delta of the

position will become more positive and negative with small movements in the underlying price, which means that large movements in the price will benefit this structure and increase its value.

- Vega (ν): The vega of an ATM straddle is high as the vega is close to its peak for ATM options. Therefore, the value of the straddle increases as the implied volatility increases.

Employing a straddle structure aligns with our trading approach as we focus on predicting earnings volatility and not the directional movement of a jump. Earlier as seen from Table 5.3, we established a significant relationship between earnings implied volatility and how much the volatility of the earnings actually realizes after the announcement. Thus, the decision to buy and sell ATM straddles is strategically informed.

6.2.2 Strangle

A long strangle shares similarities to a long straddle, involving once again being long both a call option and a put option. However, unlike a straddle where the options had the same strike, here the options have different strike prices, with the call strike being higher than the underlying asset price and the put strike being lower. This indicates that both options are OTM. The payoff diagram of a straddle at expiration is shown below:

Like the long straddle strategy, the strikes of the strangle can be chosen symmetrically such that the delta of this structure is near zero, and therefore, once again, we focus on making bets on volatility and not direction. Since the gamma and vega of OTM options are lower than that of ATM options, we see that the gamma and vega of the strangle are lower than that of the straddle. However, this is offset by the fact that OTM options are less expensive, and therefore entering a long strangle

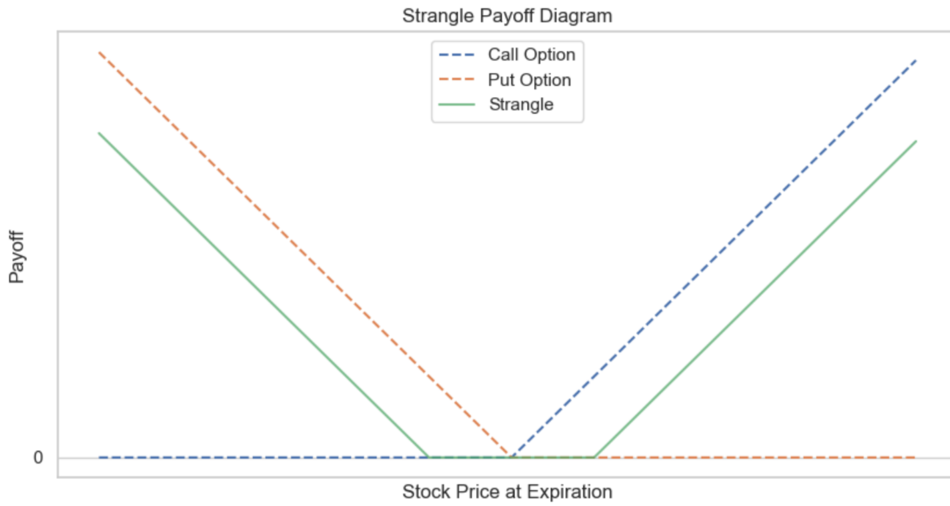

Figure 6.5: Strangle Payoff

is generally less expensive than entering a long straddle.

6.3 Trading Strategies

When back-testing our trading strategies, we must first establish a set of initial assumptions to simplify the model. We list the following assumptions of the market below:

1. We assume there are no trading costs or fees involved whenever we go long or short options. Therefore, transactional fees do not impact our return. We also assume that there are no additional fees from our short position. We also choose to operate at the midpoint of the bid-ask spread.

2. We operate under the assumption of infinite liquidity such that our trades do not influence market prices, regardless of the volume we wish to transact in both the option and underlying asset price.

3. We assume the ability to execute trades immediately at the desired price, assuming an optimal market transaction scenario where our orders are filled instantaneously at the best available price.

4. We consider unlimited depth in our position, meaning there is no cap on the amount we can lose at any given point, allowing us to fully explore the strategy's risk profile without restrictions on losses or drawdowns.

6.3.1 Trading on Past Mean Earnings Volatility

Our most straightforward strategy involves trading based on the mean earnings volatility we obtain from the term-structure estimator. To do so, we first find the mean earnings volatility using the term-structure estimator from our training set. We then evaluate the earnings volatility 7 days prior to the earnings announcement. If the level is below the mean we estimated, we opt to long straddles. Conversely, if it exceeds the mean, we opt to short straddles. This strategy hinges on the premise that when the earnings volatility is below its historical average, the market may be underestimating the volatility of the upcoming earnings and vice versa.

When testing this strategy, we focus on options with the nearest expiration date after the earnings announcement. Our strategy is to hold the positions through the announcement and unwind them the earliest trading day after the earnings are disclosed.

Subsequently, we refine this strategy by replacing the mean earnings volatility with the estimate of earnings volatility derived under the stochastic volatility model. This allows us to assess which approach serves as a more effective baseline for trading around earnings announcements. For the sake of brevity, we label these strategies 1 and 2, respectively.

6.3.2 Trading on Predicted Earnings Volatility

Building on the predictive insights from the previous chapter, where the LASSO model was identified as having superior performance over the random forest model, we utilize all available fundamental information before an earnings and our model to forecast the earnings volatility. We utilize a similar methodology to our trading strategy above but now anchor to this predicted value as our baseline. If the market's current volatility level is below the predicted value, we opt to long straddles and vice versa.

This strategy relies on the premise that our straddle position will gain in value as the discrepancy between the market's priced-in earnings volatility and our model's predicted earning volatility narrows as the earnings date approaches. This approach also implicitly assumes that the diffusive volatility level will remain relatively stable from the time we first initiate our position until the earnings announcement. Therefore, when trading, we only trade on our belief of the volatility of the earnings.

We then further augment this strategy utilizing strangles, again leveraging our predicted earnings volatility. With this new approach, we observe the percentage difference between the predicted value and the current earnings volatility, adjusting the strikes of the strangles based on the magnitude of this discrepancy. A greater percentage difference prompts us to opt for strikes closer to ATM, leveraging options with higher vega to maximize the impact of vega when significant mispricings in volatility expectations are found. Conversely, for smaller discrepancies, we choose strangles with strikes further from ATM, reducing our exposure to vega and gamma while still positioning based on volatility expectations. This strategy allows us to effectively balance both risk and reward.

Our strategy is implemented as follows: For differences exceeding 100%, we opt for ATM straddles to exploit big mispricings of volatility directly. For differences between 50% and 100%, we employ strangles with strikes set approximately 5% from ATM. For any other scenario, our strategy utilizes strangles with strikes set about 10% from ATM. For example, consider an upcoming earnings announcement with the stock currently at $100. If the measured earnings variance yields 0.01, but our model predicts a level of 0.013, indicating a discrepancy of 30%, we opt to long strangles with strikes at \$90 and \$110. If instead, our prediction is at a level of 0.02, we long ATM straddles with strike \$100. We label these strategies 3 and 4, respectively. We summarize all four strategies in Table 6.1.

Strategy	**Description**
1	Long straddles if below mean, short straddles if above
2	Long straddles if below stochastic estimate, short straddles if above
3	Long straddles if market's volatility is below predicted, short straddles if above
4	Long structure if market's volatility is below predicted, short structure if above: ATM Straddles for >100% discrepancy, Strangles 5% from ATM for 50%-100% discrepancy, Strangles 10% from ATM otherwise

Table 6.1: Summary of Trading Strategies Based on Earnings Volatility

6.3.3 Delta-Hedging

For all of the trading strategies above, we construct the following portfolio for back-testing. When it is time to trade, we go long or short N contracts:

$$N = 100,000/(V_t * Q)$$

where V_t represents the market value of the straddle at time t and Q is the contract size. We use $Q = 100$ consistent with the standard delivery of 100 shares for most equity options. This ensures that across all tickers, we maintain an approximately equal nominal value of options for every trade. If positions are taken long, financing is secured by borrowing at the risk-free rate. Conversely, if positions are shorted, then we gain on this cash at the risk-free rate.

Once we initiate our position, the gamma of the straddle causes the delta of our position to deviate from 0 as the underlying stock moves. To mitigate this, we employ "delta-hedging" by adjusting our equity position daily. At the end of the trading day, let us say we have the delta of our portfolio is:

$$\Delta_{\text{Portfolio}} = N * Q * \Delta_{\text{Straddle}} + X$$

where Δ_{Straddle} is the delta of the straddle in the market and X is the number of shares of stock we currently have in our portfolio. Since we want our portfolio to be delta neutral, we adjust by acquiring or liquidating $-\Delta_{\text{Portfolio}}$ shares of stock: selling shares if $\Delta_{\text{Portfolio}}$ is positive and buying shares if $\Delta_{\text{Portfolio}}$ is negative. To finance this process, we borrow at the risk-free rate. We do this process at the end of every trading day until our position is closed. This process ensures that the value of our portfolio gains solely from changes in volatility instead of directional movements. Once the trade ends, we liquidate our positions, converting everything into cash. The return of the trade was determined by comparing the portfolio's value before and after the trade. Following this, we reset the portfolio to its original state before the next trade. Figure 6.6 illustrates the execution of this strategy.

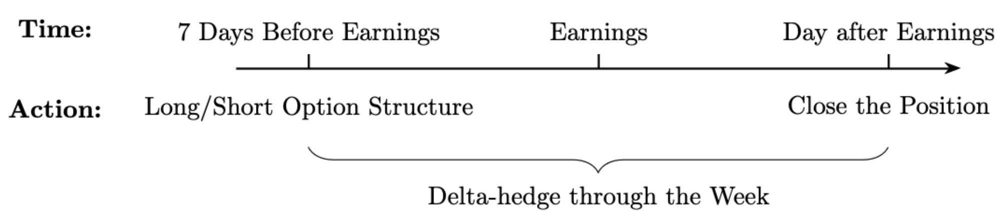

Figure 6.6: Execution of the Strategy

6.4 Back-Testing and Evaluating Trading Performance

To simplify assumptions, we use the current risk-free rate which we obtain from the current 3 Month Treasury Bill Rate across all time periods. Since this is an annualized value, we transform this value to different time periods as follows:

$$r = (1 + r_{\text{Annual}})^{1/N} - 1$$

where N is the number of compounding periods per year. To get the daily risk-free rate, we utilize $N = 365$ while for the weekly value, we utilize $N = 52$.

To evaluate our results, we look at the following benchmarks for each strategy. First, we assess the success rate of the strategy for that a ticker by comparing the instances of accurate buy or sell decisions against the total number of trades executed. We determine if a strategy is accurate if we make a positive return from the execution of the trade. For example, for TSLA, we made in total 12 earnings trades. Under Strategy 1, five of the trades resulted in positive returns, giving us a success rate of 41.7%.

Furthermore, we benchmark our results to two alternative strategies:

1. We examine if the average return of our strategy surpasses the risk-free rate. This benchmark determines if the expected return of the strategy outperforms the returns from holding cash over the same time frame.

2. We evaluate whether or not the Sharpe ratio of our strategy exceeds that of investing in the S&P 500 index. We compute two Sharpe ratios for the following strategies for comparison: first, investment in the S&P during the same time frame as our earnings which we call S&P500 (Week), as we invest through the same week as our strategies. second, investment in the S&P from the start of

2020 and holding until the current period, with its Sharpe ratio adjusted to a weekly measure, which we call S&P500 (Year), as we invest throughout the entire year. The Sharpe ratio, a measure of the excess return per unit risk in a strategy, is calculated as:

$$\text{Sharpe} = \frac{\mu_{\text{strategy}} - r_{\text{risk free}}}{\sigma_{\text{strategy}}}$$

This second benchmark serves to assess whether or not it is better to invest in the S&P500 instead of executing our strategy.

6.4.1 Empirical Results of Back-Test

We back-test our strategies on MSFT, NFLX, TSLA, AAPL, META, AMZN, AMGN, and INTC, which are the same tickers we found our estimates under the stochastic volatility model as well. The strategies are back-tested on data that account for 12 earnings announcements per ticker. After finding our results on individual tickers, we pool our findings, so that we can replicate a "portfolio" of tickers utilizing the strategy.

The results for each strategy are found in Tables 6.2, 6.3, 6.4, and 6.5. The mean return, standard deviation, and success rate are given as percentages. From these results, we make the following observations. Analyzing individual tickers, we note that AAPL has negative mean returns across all strategies. Additionally, TSLA has negative mean returns in all strategies except Strategy 4. This indicates that it is challenging to accurately forecast earnings-related uncertainty for these tickers, even when utilizing a baseline value as in Strategy 1 and 2. On the other hand, AMZN exhibits positive mean returns in every strategy, though with a slight decrease in Strategy 4, and generally has a high Sharpe ratio, indicating high returns relative to the risk taken. Observing the standard deviation of returns, we find it to be particularly high for META across all 4 strategies, highlighting huge swings in returns for this ticker. However, most tickers display return standard deviations within the range of 0.7 to 0.8, suggesting a consistent level of risk across the trades.

Looking at the pooled results, we see that Strategy 2, which leverages the stochastic estimation value, yields the highest returns and highest Sharpe ratio, This lends credence to the idea that our earnings volatility estimation under the stochastic model is a better baseline estimation than the mean value derived from the modified Black-Scholes model. Notably, the first three strategies have positive mean returns, with Strategy 4 being the only strategy with a negative mean return. Interestingly, the standard deviation of Strategy 4 is the lowest, aligning with the expectation that utilizing strangles and straddles would reduce risk. In the future, testing different heuristics on selecting strike prices for the strangles could potentially increase returns while still keeping the risk low.

When evaluating the success rate of each strategy, we note that the success rate of Strategy 1, centered on trading around the mean estimation, boasts the highest success rate at 50%. Interestingly, despite the high frequency of correct predictions, it results in the second-lowest return among the four strategies. Likewise, Strategy 4 has the second highest success rate, yet results in a negative return. This suggests that although these two strategies are able to anticipate market volatility with relative frequency, they underperform in profitability compared to the other two strategies, which despite executing fewer successful trades, achieve higher gains in returns.

Firm	Mean	SD	Sharpe	Success
MSFT	38.13	66.51	0.57	66.67
NFLX	-11.47	81.22	-0.14	33.3
TSLA	-16.2	37.13	-0.44	41.7
AAPL	-6.59	54.12	-0.12	41.7
META	12.15	156.17	0.08	41.7
AMZN	33.54	80.89	0.41	66.67
AMGN	14.16	35.81	0.39	68.3
INTC	-23.16	87.02	-0.27	50.0
Pooled	5.07	85.74	0.06	50.0

Table 6.2: Strategy 1 Performance (%)

Firm	Mean	SD	Sharpe	Success
MSFT	12.83	75.55	0.17	33.3
NFLX	16.54	80.38	0.21	41.7
TSLA	-27.39	20.80	-0.92	16.7
AAPL	-10.98	53.39	-0.21	41.7
META	50.66	148.25	0.34	41.7
AMZN	25.36	83.8	0.30	41.7
AMGN	-6.59	37.88	-0.18	41.7
INTC	53.64	72.45	0.74	83.3
Pooled	14.26	84.71	0.17	42.7

Table 6.3: Strategy 2 Performance (%)

Firm	Mean	SD	Sharpe	Success
MSFT	12.84	75.55	0.17	33.3
NFLX	3.65	81.97	0.04	50.0
TSLA	-16.2	37.13	-0.44	41.7
AAPL	-10.98	53.39	-0.21	41.7
META	11.14	156.24	0.07	33.3
AMZN	25.36	83.8	0.30	41.7
AMGN	-6.59	37.88	-0.18	41.7
INTC	53.64	72.45	0.74	83.3
Pooled	9.11	85.41	0.11	45.8

Table 6.4: Strategy 3 Performance (%)

Firm	Mean	SD	Sharpe	Success
MSFT	-18.09	67.03	-0.27	33.3
NFLX	-30.56	81.85	-0.38	50.0
TSLA	33.61	92.92	0.36	58.3
AAPL	-12.77	67.9	-0.19	33.3
META	-7.0	79.28	-0.09	58.3
AMZN	2.45	76.34	0.03	50.0
AMGN	12.2	76.98	0.16	58.3
INTC	-10.26	76.9	-0.14	33.3
Pooled	-3.80	79.95	-0.05	46.9

Table 6.5: Strategy 4 Performance (%)

Upon stratifying the returns by year, we observe that the highest returns generally occurred in 2022. One possible explanation for this phenomenon is the greater market volatility experienced during that year due to rising inflation and aggressive tightening of monetary policy by the Federal Reserve. Such market conditions allow us in every trade made during that year to capture more volatility through earnings. This environment creates numerous opportunities for our strategies to exploit discrepancies between predicted and actual earnings volatility, which amplifies our ability to achieve higher returns during this time.

Firm	**2020**	**2021**	**2022**	**2023**
MSFT	20.66	23.11	66.37	37.69
NFLX	-45.11	10.79	-19.96	34.41
TSLA	22.49	-23.30	-29.71	-49.86
AAPL	3.78	-65.02	36.37	24.16
META	5.89	-30.31	104.66	-169.27
AMZN	11.34	-14.72	103.96	11.44
AMGN	26.59	4.45	19.90	-7.30
INTC	58.36	-80.10	-51.92	75.08
Pooled	13.0	-21.88	28.71	-5.46

Table 6.6: Annual Returns (%) of Strat 1

Firm	**2020**	**2021**	**2022**	**2023**
MSFT	-22.87	23.11	41.94	-37.46
NFLX	-45.11	10.79	81.20	-34.18
TSLA	-22.26	-23.30	-29.71	-49.86
AAPL	-13.76	-65.02	36.37	24.16
META	1.87	-30.31	138.52	169.50
AMZN	-9.49	-14.72	100.71	-11.21
AMGN	-14.32	12.46	-19.67	-7.30
INTC	62.84	80.33	52.16	-74.85
Pooled	-7.89	-0.83	50.19	-2.65

Table 6.7: Annual Returns (%) of Strat 2

Firm	**2020**	**2021**	**2022**	**2023**
MSFT	-22.87	23.11	41.94	-37.46
NFLX	15.33	10.79	-19.96	34.41
TSLA	22.49	-23.30	-29.71	-49.86
AAPL	-13.76	-65.02	36.37	24.16
META	1.86	-30.31	104.66	-169.27
AMZN	-9.49	-14.72	100.71	-11.21
AMGN	-14.32	12.46	-19.67	-7.30
INTC	62.84	80.33	52.16	-74.85
Pooled	5.26	-0.83	33.31	-36.42

Table 6.8: Annual Returns (%) of Strat 3

Firm	**2020**	**2021**	**2022**	**2023**
MSFT	-42.82	26.30	-39.82	-38.52
NFLX	0.35	-6.84	-85.30	0.83
TSLA	93.35	0.46	-31.48	247.32
AAPL	36.87	-87.36	8.62	51.13
META	36.16	-24.89	29.31	-210.21
AMZN	-40.15	-15.26	47.36	21.44
AMGN	-10.67	54.82	-21.80	46.31
INTC	59.07	-42.85	-28.98	-12.98
Pooled	16.52	-11.95	-15.13	13.16

Table 6.9: Annual Returns (%) of Strat 4

Finally, we benchmark our pooled strategy to the metrics mentioned above. The results are found in Table 6.10. We see that the profitable strategies have mean returns that are both greater than that of the risk-free rate and of both S&P measures. Therefore, it is more profitable to utilize our strategy than to hold cash or invest in the S&P during the times we were trading.

Additionally, we see that the most profitable strategy, Strategy 2, has a Sharpe ratio higher than that of S&P500 (Week), which means that for the level of risk, this strategy provides more return than investing in the S&P500 in that same week. We also observe that Strategy 3 is relatively comparable in Sharpe to the S&P500 (Week) and higher than the S&P500 (Year).

These results indicate that we are able to find strategies that are relatively successful. However, it is important to note the limitation of this strategy imposed by the frequency of earnings announcements. It is somewhat unrealistic to hold cash in between earnings without investing it, making it somewhat impractical to solely rely on the strategies above. A more feasible approach would be to consider these strategies as advantageous overlays, beneficial for portfolio holders to implement during earnings seasons and enhancing their performance selectively.

Strategy	**Mean** (%)	**SD** (%)	**Sharpe**
Strategy 1	5.06	85.73	0.06
Strategy 2	14.26	84.71	0.17
Strategy 3	9.11	85.73	0.11
Strategy 4	-3.80	79.95	-0.05
S&P500 (Week)	0.50	3.31	0.12
S&P500 (Year)	0.26	2.89	0.09
Risk-Free	0.1	—	—

Table 6.10: Comparison of Strategies to Benchmark

Chapter 7

Conclusion

7.1 Conclusion

In this thesis, we develop various methodologies for estimating earnings uncertainty and volatility from options data. We find results of earnings volatility for tickers under both a modified Black-Scholes model incorporating earnings jumps and a more sophisticated stochastic volatility model with earnings jumps. After quantifying the earnings volatility across multiple tickers using these methods, we then decompose these values into various fundamental features. This analysis allows us to build intuition on what features were important in predicting earnings volatility, and we find that previous earnings volatilities from the last two quarters are the most predictive features for the current earnings volatility. Subsequently, we leverage more complex models, such as the LASSO model and a random forest model, to predict earnings volatility and evaluate which model was superior. We find that while the random forest model performs better on the training set, the overall out-of-sample results are relatively comparable between the two models. We opt to utilize the LASSO model for the trading step, seeing as the simpler LASSO model performed as well as the more complex random forest model.

We then develop multiple strategies utilizing options structures like straddles and strangles to trade on the forecasted earnings volatility for a specific earnings announcement. Four distinct strategies were implemented: one based on the mean earnings estimate, another on the stochastic earnings estimate, and two strategies that utilized our predicted earnings estimate derived from our LASSO model. We observe that the strategy leveraging the stochastic earnings estimate had the highest return and highest Sharpe ratio. We affirm that many of our strategies are viable, showing comparable performance to certain benchmarks such as investing in the S&P500 or holding cash at the risk-free rate.

Our work extends upon previous literature by developing profitable trading strategies utilizing previously developed estimators of earnings volatility. This demonstrates the potential value in leveraging discrepancies between our predicted and 'expected' market volatility of earnings announcements.

7.2 Further Work

There are numerous interesting extensions that can be pursued within each aspect of this thesis. We focus on two possible dimensions of future work: enhancing the predictive model and refining the trading strategies.

7.2.1 Enhancing the Predictive Model

In our work, we encountered multiple limitations and challenges that could have influenced our results and the interpretations of our findings. First, the inherent nature of earnings announcements, occurring only four times a year for each ticker, restricts the amount of data available on earnings volatility. We were also limited in the number of features we were able to create and utilize. For example, the scarcity of data on analyst forecasts may have affected our measure of analyst dispersion. In addition, we were only able to leverage a few different features in our models, which may have hurt their predictive capabilities.

One direction for future research lies in the exploration of different stock price models and their effect on the estimated earnings volatility. For example, we could examine a model that allows for the possibility of discontinuous jumps not only in the stock price process but also in the volatility process. Moreover, delving into the statistical properties of our earnings estimators such as finding standard errors could be valuable in thinking about how the uncertainty of our earnings estimates compounds as we go from estimation to predictions to trading.

Further exploration could be extended toward the model features beyond the few that were assessed in this study. One approach could be to incorporate market sentiment analysis through the examination of news and articles, offering a method to measure market expectations and sentiment surrounding earnings. The prevalence of discussion around an earnings announcement may indicate more volatile earnings, providing a more nuanced view of sentiment around the firm's earnings.

Regarding model development and prediction, there exists a considerable opportunity to enhance the complexity and accuracy of our models. In this thesis, we primarily examined two models, the LASSO model and the random forest model. However, future research could explore more sophisticated models such as Long Short-Term Memory (LSTM) Neural Networks. LSTMs are an interesting avenue to research as they are able to model complex dynamics in time series data, making them suited for modeling nuanced behaviors found in financial markets. However, they present challenges in training. Nevertheless, leveraging a more extensive dataset that is comprised of longer historical data and more firms could help mitigate these issues.

7.2.2 Refining the Trading Strategies

When developing our trading strategies, we had to make some strict assumptions about the market. An area of refinement could be to loosen some of those assumptions such as transaction costs, which could impact the performance of our strategies.

The trading strategies that were described in this thesis represented basic strategies that showcased the potential of leveraging our model to make profitable trades. These strategies could be expanded to include alternative options structures such as calendars or experiment with various strike prices that would balance the risk of our strategies. Additionally, exploring strategies that leverage out-of-the-money options could provide ways to exploit volatility skew and offer more insights into the dynamics surrounding earnings. Finally, we could expand our strategies to include directional strategies, in the context of "good" or "bad" sentiment around earnings, in order to open up even more avenues outside of simple volatility bets for leveraging market movements.

A more in-depth exploration of how these strategies could synergize and integrate with more traditional portfolio strategies could be interesting. In this paper, we examined each earnings event independently. In the future, we could extend our study to estimate not only the earnings volatility of a specific company but also the correlation between earnings volatility for pairs of companies. Understanding these correlations may allow us to construct a broader diversified portfolio that takes into account the collective earnings events of various sectors.

Bibliography

[1] G. Bakshi, C. Cao, and Z. K. Zhong. Assessing models of individual equity option prices. *Review of Quantitative Finance and Accounting*, 57(1):1–28, Jul 2021.

[2] R. Ball and P. Brown. An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2):159–178, 2014.

[3] A. Beber and M. W. Brandt. The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, 53(8):1997–2039, 2006.

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

[5] C. Cao and G. Bakshi. Risk-neutral kurtosis, jumps, and option pricing: Evidence from 100 most actively traded firms on the cboe. *Social Science Research Network*, 2003.

[6] T. Crack. *Basic Black-Scholes: Option Pricing and Trading*. Timothy Falcon Crack, 2014.

[7] D. C. Donelson and R. J. Resutek. The predictive qualities of earnings volatility and earnings uncertainty. *Review of Accounting Studies*, 20(1):470–500, Mar 2015.

[8] A. Dubinsky and M. Johannes. Option pricing of earnings announcement risks. *The Review of Financial Studies*, 32(2):646–687, 2019.

[9] C. R. Harvey and R. E. Whaley. Market volatility prediction and the efficiency of the s&p 100 index option market. *Journal of Financial Economics*, 31(1):43–73, 1992.

[10] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343, 1993.

[11] I. I. Inc. I/b/e/s academic research program dataset. Wharton Research Data Services (WRDS), 2023.

[12] S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101, 2002.

[13] J.-K. Lee. An efficient numerical method for pricing american put options under the cev model. *Journal of Computational and Applied Mathematics*, 389:113311, 2021.

[14] R. C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1):125–144, 1976.

[15] M. Mrázek and J. Pospíšil. Calibration and simulation of heston model. *Open Mathematics*, 15(1):679–704, 2017.

[16] S. Natenberg. *Option Volatility & Pricing: Advanced Trading Strategies and Techniques*. McGraw Hill LLC, 1994.

[17] OptionMetrics. Ivy db us. Wharton Research Data Services (WRDS), 2023.

[18] J. M. Patell and M. A. Wolfson. The ex ante and ex post price effects of quarterly earnings announcements reflected in option and stock prices. *Journal of Accounting Research*, 19(2):434–458, 1981.

[19] J. M. Patell and M. A. Wolfson. The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics*, 13(2):223–252, 1984.

[20] N. Privault. *Introduction to Stochastic Finance with Market Examples*. Chapman and Hall/CRC;, 2022.