Energy Investing

Comparing Renewable and Fossil Energy Company Stock Prices Using the Binomial Tree Model

Grayson Hamrick
Honors Thesis

Department of Mathematics
Advisor: Leslie Cheng, Ph.D.
May 13, 2021
Abstract

Starting with the commonly used binomial tree model, we develop an analog to the Hull-White Algorithm with lightened restrictions on the value $q$, the probability that the price of a stock will rise from the beginning to the end of one time period. We apply this new algorithm to a 5-year set of stock price data to determine “optimal” $q$ values ($q$ values which give the model with the best fit to the real data) for a group of energy stocks. These $q$ values, as well as values for the expected day-to-day increase predicted by the new algorithm, are used to compare the investment outlook for renewable energy companies and fossil fuel companies to each other and to the broader stock market.
Acknowledgements

There are a number of people who have made my Bi-Co journey memorable. The Bryn Mawr College Mathematics Department and Haverford College Chemistry Department have taught me how to be a mathematician and scientist in the friendliest and most enriching learning environments I could have imagined. My professors have taught me more than I will ever remember, and I have made memories with my classmates that I will never forget. Thank you to Liz Miličević for always believing in me as a mathematician. Thank you especially to Lou Charkoudian and Casey Londergan for making my chemistry major possible. Thank you as well to the Arnold & Mabel Beckman Foundation for awarding me the Beckman Scholarship and helping to cultivate my love of research.

An excessive amount of thanks is due to Leslie Cheng, the best math advisor and mentor I could have asked for. I am thankful for the hours of support you have provided me, particularly the many meetings we devoted to scheduling to make a second major work. I will always appreciate the effort you put into supporting your students.

Lastly, I would like to thank my family and friends for all they have done for me over the years. I would like to thank my parents, Scott and Krista, for their unconditional love and for their support of my future endeavors. Thank you to my brother, Asher, for pushing me to be the best version of myself (if only to prevent myself from falling too far into your shadow). And to my best friends, Courtney, Joe, Johanna, Colby, Anthony, Brian, Jon, Matt, Ben, and Ethan: thank you for keeping me sane.
## Contents

Abstract i

Acknowledgements ii

1 Introduction 4

1.1 The Global Climate Crisis ............................... 4
1.2 Stocks and Options ........................................ 4
1.3 The Impact of Investing ................................. 5
1.4 Companies of Interest ............................... 6

2 Methodology 7

2.1 The Binomial Model for Stocks and Options ................. 7
2.2 The Generalized Binomial Model .......................... 11
2.3 Derivation of the Hull-White Algorithm .......................................................... 12
2.4 Derivation of an Algorithm for General Q ......................... 17
2.5 Real-World Implementation .............................. 21
2.6 Statistical Analysis ........................................ 22

3 Results 24

3.1 Q and M Values for Classes of Energy Stocks ...................... 24
3.2 Analysis of M ............................................. 34
3.3 Impact of the COVID-19 Pandemic on Energy Investments ............ 35

4 Conclusion 44

5 Digital Appendix 45
List of Figures

1  In the Binomial Model, the share of stock and its associated option begin at the prices $S_0$ and $V_0$, respectively. After one time period $t$, each asset has two price possibilities: $S_u$ or $S_d$ for the share of stock and $U$ or $D$ for the option. ................................................................. 7

2  Call (left) and put (right) option price possibilities in the one-period Binomial Model. ................................................................. 8

3  Depiction of the Generalized Binomial Model over four time periods $t$. The Generalized Binomial Model can be expanded to any number of time periods and the time period, $t$, can be chosen as any length. .............. 12

4  Example of Generalized Binomial Model price movement of a stock over three time periods $t$. ................................................................. 13

5  Stock price branch under the Hull-White Assumptions. .................... 14

6  Stock price tree for Boeing under the Hull-White Assumptions. ............ 17

7  Stock price branch for the binomial tree model with a generalized upwards price movement probability $q$ and downwards price movement probability $1 - q$. ................................................................. 18

8  Scheme of the method used for the real-life application of the Algorithm for General Q, where the stock prices for the week $W_j$ are being predicted using the stock price ratios derived from week $W_{j-1}$. This process was iterated over the entire 5-year data set. ................................. 22

9  Example of best-fitting $q$ values for Weeks 4 and 21 of the ExxonMobil (XOM) data set. The 4th week of the data set (on the left) is best represented by $q = 0.66$, while the 21st week of the data set (on the right) is best represented by $q = 0.30$. Additionally, note the red line of best fit for expected values of each tree. The slope of the 4th week expected value is 0.224, while the slope of the 21st week expected value is 0.058. .......... 23

10 Histograms of the relative frequencies for $q$ values for the weeks of the 5-year data set of the five fossil fuel companies (XOM, CVX, COP, BTU, and ARCH). ................................................................. 26

11 Histograms of the relative frequencies for $q$ values for the weeks of the 5-year data set of the six renewable energy companies (NEE, TSLA, FSLR, RUN, CWEN, and PLUG). ................................................................. 27

12 Histograms of the relative frequencies for $q$ values for the weeks of the 5-year data set of the four most commonly cited market indices (DJIA, NDAQ, RUT, and GSPC). ................................................................. 28

13 Histograms of aggregated $q$ values for stocks in the fossil fuel group (left), renewable energy group (right), and market indices (bottom). Even with an increase in data, the distribution of $q$ values remains noisy. .......... 29
<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Frequencies of weekly $m$ values for each stock in the fossil fuel group over the 5-year time period.</td>
</tr>
<tr>
<td>15</td>
<td>Frequencies of weekly $m$ values for each stock in the renewable energy group over the 5-year time period.</td>
</tr>
<tr>
<td>16</td>
<td>Frequencies of weekly $m$ values for each of the four major indices over the 5-year time period.</td>
</tr>
<tr>
<td>17</td>
<td>Aggregated weekly $m$ values for stocks in the fossil fuel group (left) and the renewable energy group (right). Note that the average (mean) $m$ value for the renewable energy group is roughly four times greater than that of the fossil fuel group. Also note that the maximum $m$ value for the renewable group is more positive than that of the fossil fuel group. Additionally, the minimum $m$ value for the renewable group is less negative than the fossil fuel group.</td>
</tr>
<tr>
<td>18</td>
<td>Frequencies of weekly $m$ values for each stock in the fossil fuel group pre-COVID (before March 11, 2020).</td>
</tr>
<tr>
<td>19</td>
<td>Frequencies of weekly $m$ values for each stock in the fossil fuel group following the onset of COVID in the United States (after March 11, 2020).</td>
</tr>
<tr>
<td>20</td>
<td>Frequencies of weekly $m$ values for each stock in the renewable energy group pre-COVID (before March 11, 2020).</td>
</tr>
<tr>
<td>21</td>
<td>Frequencies of weekly $m$ values for each stock in the renewable energy group following the onset of COVID in the United States (after March 11, 2020).</td>
</tr>
<tr>
<td>22</td>
<td>Frequencies of weekly $m$ values for each of the four major indices prior to the onset of COVID in the United States (before March 11, 2020).</td>
</tr>
<tr>
<td>23</td>
<td>Frequencies of weekly $m$ values for each of the four major indices following the onset of COVID in the United States (after March 11, 2020).</td>
</tr>
<tr>
<td>24</td>
<td>Aggregated weekly $m$ values for stocks in the fossil fuel group (left), the renewable energy group (right), and the broader market indices (bottom) prior to the onset of COVID in the United States (before March 11, 2020).</td>
</tr>
<tr>
<td>25</td>
<td>Aggregated weekly $m$ values for stocks in the fossil fuel group (left), the renewable energy group (right), and the broader market indices (bottom) following the onset of COVID in the United States (after March 11, 2020).</td>
</tr>
</tbody>
</table>

**List of Tables**

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Information relating to the companies of interest. [7]</td>
</tr>
<tr>
<td>2</td>
<td>Stock price history for Boeing (BA).</td>
</tr>
</tbody>
</table>
Metrics relating to the \( m \) values of each group of stocks. Notice that the average \( m \) value for the renewable group is greater than that of the fossil fuel group which is greater than that of the market indices. However, the standard deviations of each of the energy groups is higher than that of the broader market indices. Together, these results indicate that energy, either renewable or fossil, is a higher risk, higher reward investment than the broader market.
1 Introduction

1.1 The Global Climate Crisis

In the midst of the global pandemic of a novel coronavirus, it can be easy to forget that climate change is perhaps the most pressing existential threat that humanity faces today. Beginning with the Industrial Revolutions of the 18th, 19th, and early 20th centuries, the alarming rate of global climate change is largely a result of human action [1]. Over the past 300 years, humans have become increasingly tied to energy. Energy powers our factories, our computers, our lights, and many of the other novelties that have made the modern global standard of living greater than it has ever been. No one of these improvements is a problem in itself; however, the source of the energy that powers these improvements is.

Fossil fuels—such as coal, petroleum, and natural gas—have served as cheap, dense sources of energy for the past 300 years. Easy access to large amounts of transportable energy has propelled human technological development: coal was the source of power for the early steam engines that powered the First Industrial Revolution, while petroleum products such as diesel and gasoline made the Second Industrial Revolution possible [2]. While some fossil fuels are “greener” than others—anthracite coal burns cleaner than bituminous, and natural gas is better than diesel, for example—the burning of all hydrocarbon fossil fuels release greenhouse gases, including the notorious carbon dioxide (CO$_2$), when burned [3]. As schoolchildren have learned for years, the “greenhouse effect” caused by these gases has gradually warmed the Earth, resulting in a cascade of effects such as a loss of biodiversity and sea level rise [1].

1.2 Stocks and Options

There exist a variety of financial securities—assets available to be bought and sold for profit or loss. The most commonly discussed financial instruments are stocks (partial ownership of a company) and bonds (debt issued by a government or a company to be paid back with interest at a later date). Other types of securities include forward, or futures, contracts (a promise to trade a given commodity at some time in the future), swaps (interest rate exchanges), and options (the right to buy or sell an asset). These are all viable investment options and can be combined, with experience, to produce a diversified portfolio. The scope of this paper pertains mostly to stocks, utilizing a model developed for the purchase and sale of options to garner a predictive mechanism for future stock price movement.

As noted above, a stock is a financial security directly related to the value of a fraction of the company that it represents, including its assets and profits. Units of stock are called shares. Shares of publicly-traded corporations—such as Facebook, Inc., The Home Depot, Inc., and Johnson and Johnson—are available to the public and are traded, as if at a digital auction, on exchanges such as the New York Stock Exchange (NYSE) or Nasdaq Stock Market (NASDAQ). Each stock has its own ticker symbol, a
An abbreviation used to identify shares of specific companies. For example, the ticker symbols for the companies above are FB, HD, and JNJ, respectively. The value of a stock is derived from the value of the assets and company profits underlying the share, as well as consumer sentiment with regards to the company.

An option is a more abstract type of financial instrument known as a derivative – the value of an option is based on, or derivative of, the price of a share of stock. An option contract grants the holder the right to buy or sell a particular asset at some specified price at some specified time in the future. There are two main types of options (although many more complicated versions). A call option grants the holder of the contract the right to buy an asset at a certain price (called the strike or exercise price) on a certain date (called the expiration date). A put option grants the holder of the contract the right to sell an asset at a certain price at a certain time in the future. For example, one could purchase a 6-month call option with a strike price of $100 for a share of Apple (AAPL) stock. If, in six months, the price of a share of AAPL is higher than $100, then the holder of the call option will exercise the option to buy the share of AAPL for $100 and gather a profit equal to the difference between the market price of the share and $100. If, in six months, the price of a share of AAPL is lower than $100, then it does not make sense for the holder of the contract to exercise the option to buy a share of AAPL at $100 when it can be bought on the market for a cheaper price, and the option becomes worthless.

1.3 The Impact of Investing

Scientists have been aware of the negative effects of fossil fuel use for years; why do we still use them for the majority of our energy? Money, for several reasons. Fossil fuels have historically been much cheaper than renewable alternatives. Cheap is good. Cheap sources of energy raise standards of living. They also result in large profit margins and lots of money to be made for the large corporations that discover, recover, and distribute those cheap fuels. Several of the largest corporations in American history have been oil companies: Exxon Mobil (the modern descendent of the 19th century oil monopoly Standard Oil), Chevron, Valero Energy, and Conoco Phillips, to name a few. These corporations (and their money) wield extraordinary power and influence over policy decisions, such as environmental regulations and energy restrictions that may otherwise decrease their bottom lines. Because of this influence, fossil fuel companies have maintained wide economic moats for decades. As human society required more energy throughout the past century, demand for fossil fuels increased, profits rose, and the stock prices of fossil fuel companies soared. Fossil fuel behemoths became wise financial investments for individuals interested in participating in the stock market. Increased investment led to increased wealth and power, and the cycle repeated. Money begets money.

Recently, a new method of investing known as Environmental, Social, and Governance (ESG) investing – also known as “sustainable” or “impact” investing – has become popular. An ESG investing strategy does not rely entirely on profit maximization for investment decision-making; instead, investors consider the broader societal impact of
the companies they choose to invest in. These investors contend that while profits are
good, they are not everything in the search for a better, more sustainable society. Com-
panies pursuing the development and deployment of renewable sources of energy fall
into this category of socially-responsible investments. In this paper, I will show, using
the binomial stock model and variations of the Hull–White Algorithm, that investments
in a variety of renewable energy companies are not only socially-responsible, but also
potentially profitable when compared to investments in fossil fuel companies.

1.4 Companies of Interest

<table>
<thead>
<tr>
<th>Name</th>
<th>Ticker</th>
<th>Renewable</th>
<th>Category</th>
<th>Founded</th>
<th>IPO</th>
<th>Cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>NextEra Energy Inc.</td>
<td>NEE</td>
<td>Y</td>
<td>Mixed</td>
<td>1984</td>
<td>June 19, 2014</td>
<td>Large</td>
</tr>
<tr>
<td>Tesla Inc.</td>
<td>TSLA</td>
<td>Y</td>
<td>Vehicles</td>
<td>2003</td>
<td>June 1, 2010</td>
<td>Large</td>
</tr>
<tr>
<td>First Solar Inc.</td>
<td>FSLR</td>
<td>Y</td>
<td>Solar</td>
<td>1999</td>
<td>November 17, 2006</td>
<td>Mid</td>
</tr>
<tr>
<td>Sunrun Inc.</td>
<td>RUN</td>
<td>Y</td>
<td>Solar</td>
<td>2007</td>
<td>August 5, 2015</td>
<td>Small</td>
</tr>
<tr>
<td>Clearway Energy Inc.</td>
<td>CWEN</td>
<td>Y</td>
<td>Mixed</td>
<td>2012</td>
<td>May 11, 2015</td>
<td>Small</td>
</tr>
<tr>
<td>Plug Power Inc.</td>
<td>PLUG</td>
<td>Y</td>
<td>Hydrogen</td>
<td>1997</td>
<td>October 1, 1999</td>
<td>Small</td>
</tr>
<tr>
<td>Exxon Mobil Corporation</td>
<td>XOM</td>
<td>N</td>
<td>Oil/Gas</td>
<td>1870</td>
<td>Before 1962</td>
<td>Large</td>
</tr>
<tr>
<td>Chevron</td>
<td>CVX</td>
<td>N</td>
<td>Oil/Gas</td>
<td>1870</td>
<td>Before 1962</td>
<td>Large</td>
</tr>
<tr>
<td>ConocoPhillips</td>
<td>COP</td>
<td>N</td>
<td>Natural Gas</td>
<td>1875</td>
<td>December 30, 1981</td>
<td>Large</td>
</tr>
<tr>
<td>Peabody Energy Corporation</td>
<td>BTU</td>
<td>N</td>
<td>Coal</td>
<td>1883</td>
<td>2001</td>
<td>Micro</td>
</tr>
<tr>
<td>Arch Resources</td>
<td>ARCH</td>
<td>N</td>
<td>Coal</td>
<td>1969</td>
<td>August 11, 1988</td>
<td>Micro</td>
</tr>
</tbody>
</table>

Table 1: Information relating to the companies of interest. [7]

The above are the eleven stocks that I decided to track. The group includes six
companies focused on renewable sources of energy, while the remaining five mine, pump,
sell, provide and/or distribute fossil fuels. Fossil fuel companies tend to be descendants
of legacy companies, such as Standard Oil. Many of those companies in business today
were originally founded in the 19th century. On the other hand, the renewable energy
industry and its companies are much younger, having been formed as renewable energy
sources first reached economic viability in the late 1900s or early 2000s.

I aimed to select a variety of categories of both renewable and non-renewable sources.
Companies identified as “Mixed” (NEE and CWEN) are both holdings companies with
stakes in multiple renewable energy sources, including solar, wind, and hydroelectric. I
also aimed to cover several market capitalization (Cap) categories, where market capital-
ization is the total value of all shares of a company’s stock, calculated as

\[(\text{Price per Share}) \cdot (\text{Total # of Shares on the Market}) = \text{Market Capitalization}.\]

“Large Cap” companies have a market capitalization of $10 billion or more, “Mid Cap”
between $2 billion and $10 billion, “Small Cap” between $300 million and $2 billion,
and “Micro Cap” between $50 million and $300 million. Each size of company presents
different opportunities and trade-offs for potential investors. For example, large cap com-
panies have established themselves within their respective industries, generally providing
stability and positive long-term returns. Alternatively, small cap companies are new to
their markets or serve niche industries. These stocks provide short-term growth potential
but also risk greater volatility.
In addition to these individual stocks, I analyzed four broader market indices: the Dow Jones Industrial Average (DJIA), Nasdaq (NDAQ), Russell 2000 (RUT) and S&P 500 (GSPC). These indices each measure the cumulative values of a large number of stocks, allowing observers to get a sense of the movement of the stock market as a whole.

2 Methodology

2.1 The Binomial Model for Stocks and Options

We will begin by developing the Binomial Model, a simple model that can be used to predict future stock prices and determine the present value of a corresponding option under the assumption that there exist no arbitrage opportunities – i.e. there are no opportunities for an investor to make a risk-free profit. Suppose we have a share of stock, perhaps Microsoft (MSFT), worth $S_0$ dollars today. Assume that after one time period (of length $t$), the stock price can move up to $S_u$ dollars, or it can move down to $S_d$ dollars. Also consider an option corresponding with MSFT with a current value of $V_0$. After one time period $t$, the option can move to $U$ dollars if the stock price of MSFT moves up to $S_u$ dollars, or it can move to $D$ dollars if the stock price of MSFT moves down to $S_d$ dollars (see Figure 1).

As mentioned earlier, there are two simple types of options: calls and puts. We will consider both of these types of options to construct more descriptive Option Value Trees. Suppose that we have a call option with strike price $X$ dollars such that $S_d < X < S_u$. After one time period $t$, there are two cases.

![MSFT Share Price Tree](image1.png)

![MSFT Option Value Tree](image2.png)

Figure 1: In the Binomial Model, the share of stock and its associated option begin at the prices $S_0$ and $V_0$, respectively. After one time period $t$, each asset has two price possibilities: $S_u$ or $S_d$ for the share of stock and $U$ or $D$ for the option.
Case 1: The stock price moves up to $S_u$ dollars. Then the holder of the option will exercise their right to buy a share of MSFT for $X$ dollars per share, since $X < S_u$. Then the value of the option $U = S_u - X$ dollars.

Case 2: The stock price moves down to $S_d$ dollars. The holder of the option will not exercise their right to buy a share of MSFT for $X$ dollars, since $X > S_d$, so the call option is worthless. Then we have $D = 0$ dollars. These cases are shown in Figure 2.

Now suppose that we have a put option with strike price $X$ dollars such that $S_d < X < S_u$. After one time period $t$, there are two cases in this situation as well.

Case 1: The stock price moves up to $S_u$ dollars. Then the holder of the option will not exercise their right to sell a share of MSFT for $X$ dollars per share, since $X < S_u$, so the put option is worthless. Then $U = 0$ dollars.

Case 2: The stock price moves down to $S_d$ dollars. The holder of the option will exercise their right to sell a share of MSFT for $X$ dollars, since $X > S_d$. Then we have $D = X - S_d$ dollars. These cases are also shown in Figure 2.

![Call Option Value Tree](image1)

![Put Option Value Tree](image2)

Figure 2: Call (left) and put (right) option price possibilities in the one-period Binomial Model.

Now that we have defined the model, we would like to find a fair price for an option in the Binomial Model. Recall that $V_0$ is the value of the option at time $t = 0$. Then, in other words, our goal is to find the value of $V_0$. Also recall that $X$ is the strike price of the option, then let $T$ be the expiration of the option in years and $r$ be the annual interest rate paid by an investment of cash in a bond issued by a government or company, assuming that interest will be compounded continuously (as opposed to monthly, quarterly, annually, etc.).

One way to determine a fair price of an option in the Binomial Model is to use the Replicating Portfolio Method, in which we construct a portfolio, in this case consisting of stocks and bonds, with a value at time $t$ which we assume to be equal to the value of the option at time $t$ [5]. Suppose that the portfolio consists of $a$ shares of stock and $b$ units of a bond worth $1 each, and let $\Pi_t$ be the value of the portfolio at time $t$. Then the
value of the portfolio today, at time $t = 0$, is $\Pi_0 = aS_0 + b$ dollars. By the assumption of the Replicating Portfolio Method, $\Pi_0 = V_0$, so we have

$$V_0 = aS_0 + b. \quad (1)$$

We must determine formulas for $a$ and $b$. At time $t = T$, the Binomial Model gives two possible future values for the portfolio. Call these possible futures the “up” state – when the stock price moves to $S_u$ at time $T$ – and the “down” state – when the stock price moves to $S_d$ at time $T$. Since by assumption the value of the portfolio is equal to the value of the option, the value of the portfolio in the “up” state is equal to the value of the option in the “up” state: $aS_u + be^{rT} = U$. Similarly, the value of the portfolio in the “down” state is equal to the value of the option in the “down” state: $aS_d + be^{rT} = D$. Therefore, we aim to solve the following system of equations for $a$ and $b$, which we can then substitute into Equation (1) to solve for $V_0$:

$$aS_u + be^{rT} = U \quad (i)$$

$$aS_d + be^{rT} = D. \quad (ii)$$

Subtracting (ii) from (i), we get $aS_u - aS_d = U - D$. By factoring out $a$, we get $a(S_u - S_d) = U - D$. Since $S_u > S_d$, we know $S_u - S_d \neq 0$. Dividing by $S_u - S_d$, we arrive at

$$a = \frac{U - D}{S_u - S_d}. \quad (2)$$

We can substitute Equation (2) into (ii) to solve for $b$. By substitution, we have

$$\left( \frac{U - D}{S_u - S_d} \right) S_d + be^{rT} = D. \quad (ii)$$

Subtracting $\left( \frac{U - D}{S_u - S_d} \right) S_d$ from both sides and multiplying by $e^{-rT}$ gives us

$$b = e^{-rT} \left[ D - \left( \frac{U - D}{S_u - S_d} \right) S_d \right]$$

$$= e^{-rT} \left[ \frac{D(S_u - S_d) - US_d + DS_d}{S_u - S_d} \right]$$

$$= e^{-rT} \left[ \frac{DS_u - DS_d - US_d + DS_d}{S_u - S_d} \right]$$

$$= e^{-rT} \left[ \frac{DS_u - US_d}{S_u - S_d} \right].$$

Hence,

$$b = e^{-rT} \left[ \frac{DS_u - US_d}{S_u - S_d} \right]. \quad (3)$$
Substituting these expressions for \(a\) (Equation (2)) and \(b\) (Equation (3)) into our equation for \(V_0\) (Equation (1)), we have:

\[
V_0 = \left( \frac{U - D}{S_u - S_d} \right) S_0 + e^{-rT} \left[ \frac{DS_u - US_d}{S_u - S_d} \right]
\]

\[
= e^{-rT} \left[ e^{rT} \left( \frac{U - D}{S_u - S_d} \right) S_0 + \frac{DS_u - US_d}{S_u - S_d} \right]
\]

\[
= e^{-rT} \left[ \left( \frac{S_0 e^{rT}}{S_u - S_d} - \frac{S_d}{S_u - S_d} \right) U + \left( \frac{S_u}{S_u - S_d} - \frac{S_0 e^{rT}}{S_u - S_d} \right) D \right]
\]

\[
= e^{-rT} \left[ \left( \frac{S_0 e^{rT} - S_d}{S_u - S_d} \right) U + \left( \frac{S_u - S_0 e^{rT}}{S_u - S_d} \right) D \right].
\]

Hence,

\[
V_0 = e^{-rT} \left[ \left( \frac{S_0 e^{rT} - S_d}{S_u - S_d} \right) U + \left( \frac{S_u - S_0 e^{rT}}{S_u - S_d} \right) D \right]. \tag{4}
\]

Let \(q = \frac{S_0 e^{rT} - S_d}{S_u - S_d}\) (the coefficient of \(U\) in Equation (4)). Then:

\[
1 - q = 1 - \frac{S_0 e^{rT} - S_d}{S_u - S_d}
\]

\[
= \frac{S_u - S_d - S_0 e^{rT} + S_d}{S_u - S_d}
\]

\[
= \frac{S_u - S_0 e^{rT}}{S_u - S_d},
\]

which is equal to the coefficient of \(D\) in Equation (4). Thus a fair price, \(V_0\), of an option related to a stock in the Binomial Model is:

\[
V_0 = e^{-rT} \left[ q U + (1 - q) D \right] \quad \text{where} \quad q = \frac{S_0 e^{rT} - S_d}{S_u - S_d}. \tag{5}
\]

Our value \(q\) serves as a probability. In this case, \(q\) represents the chance that the price of the stock will move up to \(S_u\) (from \(S_0\)) after one time period \(t\), since \(q\) is being multiplied by \(U\), the value of the option if the stock price moves up to \(S_u\) after one time period. By convention, probabilities are real numbers in the interval \([0, 1]\). We shall verify that \(0 \leq q \leq 1\).
I. Show $q \geq 0$. Suppose that $q < 0$. Then $S_0e^{rt} - S_d < 0$. Multiplying by $S_u - S_d > 0$ (since $S_u > S_d$), we get $S_0e^{rt} - S_d < 0$. Then $S_0e^{rt} < S_d$. This implies that leaving money in a savings account will always make less money than investing in stock, which is a contradiction. Thus $q \geq 0$.

II. Show $q \leq 1$. Suppose that $q > 1$. Then $S_0e^{rt} - S_d > 1$. Multiplying by $S_u - S_d > 0$, we get $S_0e^{rt} - S_d > S_u - S_d$. Then $S_0e^{rt} > S_u$. This implies that leaving money in a savings account will always make more money than investing in stock, which is a contradiction. Thus $q \leq 1$.

Parts I and II together imply that $0 \leq q \leq 1$, so we can consider $q$ the probability that the price of the stock will increase from $S_0$ to $S_u$ after one time period $t$. Similarly, $1 - q$ is the probability that the price of the stock will decrease from $S_0$ to $S_d$ after one time period $t$.

2.2 The Generalized Binomial Model

We can expand the Binomial Model to $n$ time periods [5]. Let $S_0$ be the price of a share of stock at time $t = 0$, as before. After each time period $t$, assume that the stock price can go up by a factor of $u$ with probability $q$ or down by a factor of $d$ with probability $1 - q$. Then we have a Generalized Binomial Model for stock prices over $n$ time periods, where $n \in \{0, 1, 2, 3, \ldots\}$, and each time period $n$ has $n + 1$ stock price possibilities, or nodes, as shown in Figure 3.

From the simple Binomial Model with one branch, we found that the probability that the price of the stock rises is $q = \frac{S_0e^{rt} - S_d}{S_u - S_d}$. Focusing on one branch of the Generalized Binomial Model with some starting price $S_0 = S_n$ and possible ending prices $S_u = S_nu$ and $S_d = S_nd$, we get:

$$q = \frac{S_n e^{rt} - S_n d}{S_n u - S_n d}$$

$$= \frac{S_n(e^{rt} - d)}{S_n(u - d)}$$

$$= \frac{e^{rt} - d}{u - d}.$$ 

So the probability that the stock price will go up by a factor of $u$ after one time period $t$ in the Generalized Binomial Model is:

$$q = \frac{e^{rt} - d}{u - d} \quad (6)$$

where $r$ is the annual interest rate compounded continuously, $t$ is the length of one time period, $u$ is the upward price movement factor, and $d$ is the downward price movement factor.
Example. Suppose that the current price of a share of Coca-Cola Co. (KO) is $S_0 = 40.00$. Also suppose that there are three time periods with each time period being one month (so \( t = \frac{1}{12} \)) and the annual interest rate compounded continuously is \( r = 0.05 \). A time traveler arrives from three days in the future and tells us that the upward price movement factor \( u = 1.1 \) and the downward price movement factor \( d = 0.9 \). Then we can construct a stock price tree based on the Generalized Binomial Model, as shown in Figure 4, where the probability that the stock price increases is \( q = \frac{e^{rt} - d}{u - d} = \frac{e^{0.05(\frac{1}{12})} - 0.9}{1.1 - 0.9} \approx 0.5209 \) and the probability that the stock price decreases is \( 1 - q \approx 0.4791 \).

2.3 Derivation of the Hull-White Algorithm

At this point, we rely on the appearance of a time traveler to give us our values for \( u \) and \( d \), on which (with a value for \( S_0 \)) we can base our model. We would prefer if this were not the case. Instead, we would like to match \( u \) and \( d \) to two real-world, measurable components of a stock’s behavior:

- Drift \( \mu \), the percent change of the stock’s price over time
- Volatility \( \sigma \), the randomness of the relative return of the stock.

We will do this using the Hull-White Algorithm, first derived by John C. Hull and Alan White in 1990 [6]. We can consider the price of a stock after a short time period \( \Delta t \) to be a random variable \( S \). Then the relative return—the return as a fraction of the original price \( S_0 \)—of the stock is \( \frac{S - S_0}{S_0} = \frac{S}{S_0} - 1 \). We hypothesize:
A. \( \mu \Delta t = E[\frac{S_k}{S_{k-1}} - 1] \) (the expected, or average, relative return)

B. \( \sigma^2 \Delta t \) is the variance of the relative returns.

Assume that for \( k = 1, 2, 3, ..., n \), the stock price \( S_{k-1} \) will go up by a factor of \( u \) with probability \( q = \frac{1}{2} \) or down by a factor of \( d \) with probability \( 1 - q = \frac{1}{2} \), as shown in Figure 5. These are known as the Hull-White Assumptions [5].

Then after one time period \( t \), \( S_k = uS_{k-1} \) or \( S_k = dS_{k-1} \) for \( k = 1, 2, 3, ..., n \). Hence \( \frac{S_k}{S_{k-1}} = u \) or \( \frac{S_k}{S_{k-1}} = d \). For \( k = 1, 2, 3, ..., n \), let \( X_k = \frac{S_k}{S_{k-1}} \). Additionally, to avoid confusion with \( \mu \), let \( u = a \) and \( d = b \). Then:

\[
X_k = \begin{cases} 
  a & \text{with probability } \frac{1}{2} \\
  b & \text{with probability } \frac{1}{2}.
\end{cases}
\]

By Hypothesis A (from above):

\[
\mu \Delta t = E \left[ \frac{S_k}{S_{k-1}} - 1 \right] = E \left[ X_k - 1 \right] = E[X_k] - E[1]
\]

using the expected value property that \( E[X \pm Y] = E[X] \pm E[Y] \) for any random variables \( X \) and \( Y \)

\[
= \frac{1}{2} \cdot a + \frac{1}{2} \cdot b - 1 \quad \text{with definition of } E[X] \text{ and } E[k] = k \text{ for } k \in \mathbb{R}.
\]
Hence:

\[ \mu \Delta t = \frac{1}{2}(a + b) - 1. \]  

(7)

By Hypothesis B, we have:

\[ \sigma^2 \Delta t = \text{variance of relative returns of stock} \]

\[ = \text{Var} \left( \frac{S_k}{S_{k-1}} - 1 \right) \]

\[ = \text{Var} \left( X_k - 1 \right) \]

\[ = 1^2 \cdot \text{Var}(X_k) \text{ using } \text{Var}(aX + b) = a^2 \text{Var}(X) \text{ for any } a, b \in \mathbb{R} \text{ and random variable } X \text{ with } a = 1 \text{ and } b = -1 \]

\[ = \text{Var}(X_k) \]

\[ = \mathbb{E}[(X_k - \mathbb{E}(X_k))^2] \text{ using the definition of variance} \]

\[ = \mathbb{E}[(X_k - \mathbb{E}(X_k))(X_k - \mathbb{E}(X_k))] \]

\[ = \mathbb{E}[X_k^2 - 2X_k \mathbb{E}(X_k) + [\mathbb{E}(X_k)]^2] \]

\[ = \mathbb{E}[X_k^2] - \mathbb{E}[2X_k \mathbb{E}(X_k)] + \mathbb{E}[[\mathbb{E}(X_k)]^2] \]
\[
E[X_k^2] - 2E[X_k]E[k] + (E[X_k])^2 \quad \text{using } E[cX] = c E[X] \text{ for any } c \in \mathbb{R}
\]

and \(E[k] = k\) for \(k \in \mathbb{R}\)

\[
= E[X_k^2] - 2(E[X_k])^2 + (E[X_k])^2
\]

\[
= E[X_k^2] - (E[X_k])^2
\]

\[
= \frac{1}{2}a^2 + \frac{1}{2}b^2 - \left( \frac{1}{2}a + \frac{1}{2}b \right)^2
\]

\[
= \frac{1}{2}a^2 + \frac{1}{2}b^2 - \left[ \frac{1}{2}(a + b) \right]^2
\]

\[
= \frac{1}{4}[2a^2 + 2b^2 - (a + b)(a + b)]
\]

\[
= \frac{1}{4}[2a^2 + 2b^2 - (a^2 + 2ab + b^2)]
\]

\[
= \frac{1}{4}(a^2 + b^2 - 2ab)
\]

\[
= \frac{1}{4}(a - b)(a - b).
\]

Hence:

\[
\sigma^2 \Delta t = \frac{1}{4}(a - b)^2. \quad (8)
\]

Let’s solve for \(a\) and \(b\). Multiplying Equation (8) by 4, we get \(4\sigma^2 \Delta t = (a - b)^2\). Taking the square root of both sides, we have \(2\sigma \sqrt{\Delta t} = a - b\). Hence:

\[
a = 2\sigma \sqrt{\Delta t} + b. \quad (9)
\]

Now by Equation (7), \(\mu \Delta t = \frac{1}{2}(a + b) - 1\). Adding 1, we have \(\mu \Delta t + 1 = \frac{1}{2}(a + b)\). Multiplying by 2, we obtain \(2(\mu \Delta t + 1) = a + b\). Subtracting \(a\), we get \(b = 2(\mu \Delta t + 1) - a\). Using the expression for \(a\) in Equation (9), the last equation becomes \(b = 2(\mu \Delta t + 1) - [2\sigma \sqrt{\Delta t} + b]\), or \(b = 2(\mu \Delta t + 1) - 2\sigma \sqrt{\Delta t} - b\). Adding \(b\), we obtain \(2b = 2(\mu \Delta t + 1) - 2\sigma \sqrt{\Delta t}\). Hence:

\[
b = \mu \Delta t + 1 - \sigma \sqrt{\Delta t}.
\]

Substituting this into Equation (9), we get \(a = 2\sigma \sqrt{\Delta t} + (\mu \Delta t + 1 - \sigma \sqrt{\Delta t})\). Thus:

\[
a = \mu \Delta t + 1 + \sigma \sqrt{\Delta t}.
\]

Then, to summarize:

\[
a = \mu \Delta t + 1 + \sigma \sqrt{\Delta t} \quad (10)
\]
\[ b = \mu \Delta t + 1 - \sigma \sqrt{\Delta t}. \]  \hspace{1cm} (11)

Using Hypothesis A, we found that \( \mu \Delta t = E[X_k] - 1 = \bar{X} - 1 \), and Hypothesis B tells us that \( \sigma \sqrt{\Delta t} = \sqrt{\text{Var}[X_k]} = \) sample standard deviation of \( X_k = S \), where:

\[
\bar{X} - 1 = \text{sample mean} - 1 \\
= \frac{X_1 + X_2 + \ldots + X_n}{n} - 1
\]  \hspace{1cm} (12)

and

\[
S = \text{sample standard deviation of stock ratios } X_k \\
= \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n - 1}}.
\]  \hspace{1cm} (13)

Plugging these into Equations (10) and (11) (and switching back to \( u \) and \( d \) notation), we can most succinctly state the Hull-White Algorithm as:

\[ u = \bar{X} + S \]  \hspace{1cm} (14)

\[ d = \bar{X} - S \]  \hspace{1cm} (15)

where \( \bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \) and \( S = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n - 1}} \).

**Example.** Suppose that Boeing (BA) stock for the past several days has been priced as shown in Table 2. We will find estimates for \( u \) and \( d \) using the Hull-White Algorithm.

<table>
<thead>
<tr>
<th>Day</th>
<th>BA Price at Closing ($)</th>
<th>Stock Price Ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( S_0 = 29.56 )</td>
<td>( X_1 = \frac{S_1}{S_0} = \frac{30.125}{29.56} \approx 1.019 )</td>
</tr>
<tr>
<td>2</td>
<td>( S_1 = 30.125 )</td>
<td>( X_2 = \frac{S_2}{S_1} = \frac{28.56}{30.125} \approx 0.948 )</td>
</tr>
<tr>
<td>3</td>
<td>( S_2 = 28.56 )</td>
<td>( X_3 = \frac{S_3}{S_2} = \frac{27.50}{28.56} \approx 0.963 )</td>
</tr>
<tr>
<td>4</td>
<td>( S_3 = 27.50 )</td>
<td>( X_4 = \frac{S_4}{S_3} = \frac{28.125}{27.50} \approx 1.023 )</td>
</tr>
<tr>
<td>5</td>
<td>( S_4 = 28.125 )</td>
<td>( X_5 = \frac{S_5}{S_4} = \frac{27.44}{28.125} \approx 0.976 )</td>
</tr>
<tr>
<td>6</td>
<td>( S_5 = 27.44 )</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Stock price history for Boeing (BA).

First, we find the stock price ratios, \( X_n \), as shown in Table 2. Next, we find the sample mean of the stock price ratios:

\[
\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} = \frac{4.929}{5} \approx 0.9858.
\]

Now we find the sample standard deviation of the stock price ratios:

\[
S = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + (X_4 - \bar{X})^2 + (X_5 - \bar{X})^2}{5 - 1}}
= \sqrt{\frac{(1.019 - 0.9858)^2 + (0.948 - 0.9858)^2 + (0.963 - 0.9858)^2 + (1.023 - 0.9858)^2 + (0.976 - 0.9858)^2}{4}}
\approx 0.0337.
\]
Now we can find $u$ and $d$:

$$u = \overline{X} + S = 0.9858 + 0.0337 = 1.0195$$

$$d = \overline{X} - S = 0.9858 - 0.0337 = 0.9521.$$

And now we can use these values for $u$ and $d$ to predict BA stock prices—according to the Hull–White Assumptions—for Day 7 and Day 8, as shown in Figure 6.

![Boeing (BA) Stock Price Prediction](image)

Figure 6: Stock price tree for Boeing under the Hull-White Assumptions.

### 2.4 Derivation of an Algorithm for General $Q$

When we initially derived the Hull-White Algorithm, we assumed that the stock price would move either up with a probability $\frac{1}{2}$ or down with a probability $\frac{1}{2}$. However, we can derive an algorithm that generalizes the Hull-White Algorithm to be able to include any probability $q$ ($0 < q < 1$) that we choose, removing the restriction that $q = \frac{1}{2}$. As before, we hypothesize:

A. $\mu \Delta t = E\left[\frac{S}{S_0} - 1\right]$ (the expected, or average, relative return)

B. $\sigma^2 \Delta t$ is the variance of the relative returns.

Now, we assume that for $k = 1, 2, 3, ..., n$, the stock price $S_{k-1}$ will go up by a factor of $u$ with probability $0 < q < 1$ or down by a factor of $d$ with probability $0 < 1 - q < 1$, as shown in Figure 7. Our generalization of $q$ differs from the Hull-White Assumptions.

As before, after one time period $t$, $S_k = uS_{k-1}$ or $S_k = dS_{k-1}$ for $k = 1, 2, 3, ..., n$. Hence $\frac{S_k}{S_{k-1}} = u$ or $\frac{S_k}{S_{k-1}} = d$. For $k = 1, 2, 3, ..., n$, let $X_k = \frac{S_k}{S_{k-1}}$. Additionally, to avoid confusion with $\mu$, let $u = a$ and $d = b$. Then:

$$X_k = \begin{cases} a & \text{with probability } q \\ b & \text{with probability } 1 - q. \end{cases}$$
By Hypothesis A:

\[
\mu \Delta t = \mathbb{E}\left[ \frac{S_k}{S_{k-1}} - 1 \right] = \mathbb{E}\left[ X_k - 1 \right] = \mathbb{E}[X_k] - \mathbb{E}[1]
\]

using the expected value property that \( \mathbb{E}[X \pm Y] = \mathbb{E}[X] \pm \mathbb{E}[Y] \) for any random variables \( X \) and \( Y \)

\[
= q \cdot a + (1 - q) \cdot b - 1
\]

with definition of \( \mathbb{E}[X] \) and \( \mathbb{E}[k] = k \) for \( k \in \mathbb{R} \).

Hence:

\[
\mu \Delta t = qa + (1 - q)b - 1. \tag{16}
\]

By Hypothesis B, we have:

\[
\sigma^2 \Delta t = \text{variance of relative returns of stock}
\]

\[
= \text{Var}\left( \frac{S_k}{S_{k-1}} - 1 \right)
\]
\[ \text{Var}(X_k - 1) \]
\[ = 1^2 \cdot \text{Var}(X_k) \]
\[ = \text{Var}(X_k) \]
\[ = \text{E}[(X_k - \text{E}(X_k))^2] \]
\[ = \text{E}[(X_k - \text{E}(X_k))(X_k - \text{E}(X_k))] \]
\[ = \text{E}[X_k^2 - 2X_k\text{E}(X_k) + \text{E}^2(X_k)] \]
\[ = \text{E}[X_k^2] - \text{E}[2X_k\text{E}(X_k)] + \text{E}^2[\text{E}(X_k)] \]
\[ = \text{E}[X_k^2] - 2\text{E}[X_k]\text{E}[X_k] + (\text{E}[X_k])^2 \]
\[ = \text{E}[X_k^2] - 2(\text{E}[X_k])^2 + (\text{E}[X_k])^2 \]
\[ = \text{E}[X_k^2] - (\text{E}[X_k])^2 \]
\[ = qa^2 + (1 - q)b^2 - (qa + (1 - q)b)^2 \]
\[ = qa^2 + (1 - q)b^2 - (q^2a^2 + 2q(1 - q)ab + (1 - q)^2b^2) \]
\[ = qa^2 - qa^2 + (1 - q)b^2 - (1 - q)^2b^2 - 2q(1 - q)ab \]
\[ = (q - q^2)a^2 - 2q(1 - q)ab + [(1 - q) - (1 - q)^2]b^2 \]
\[ = q(1 - q)a^2 - 2q(1 - q)ab + q(1 - q)b^2 \]
\[ = q(1 - q)(a^2 - 2ab + b^2) \]
\[ = q(1 - q)(a - b)^2. \]

Hence:
\[ \sigma^2 \Delta t = q(1 - q)(a - b)^2. \] (17)
Now we will solve the $2 \times 2$ system of equations given by Equations (16) and (17) for $a$ and $b$. Multiplying Equation (17) by $\frac{1}{q(1-q)}$, we get $\frac{\sigma^2 \Delta t}{q(1-q)} = (a - b)^2$. Remember that $0 < q < 1$, so neither $q$ nor $1-q$ are equal to zero. Taking the square root of both sides, we have $\sqrt{\frac{1}{q(1-q)}} \sigma \sqrt{\Delta t} = a - b$. Hence:

$$a = \sqrt{\frac{1}{q(1-q)}} \sigma \sqrt{\Delta t} + b. \quad (18)$$

Now by Equation (16), $\mu \Delta t = qa + (1-q)b - 1$. Adding 1, we have $\mu \Delta t + 1 = qa + (1-q)b$. Subtracting $(1-q)b$, we get $qa = \mu \Delta t + 1 - (1-q)b$. Dividing by $q$ ($q \neq 0$, since $0 < q$), we get $a = \frac{(\mu \Delta t + 1) - (1-q)b}{q}$. Combining this with Equation (18), we have:

$$\frac{(\mu \Delta t + 1) - (1-q)b}{q} = \sqrt{\frac{1}{q(1-q)}} \sigma \sqrt{\Delta t} + b$$

$$\Rightarrow (\mu + 1) - (1-q)b = q \sqrt{\frac{1}{q(1-q)}} \sigma \sqrt{\Delta t} + qb$$

$$\Rightarrow (1-q)b + qb = (\mu \Delta t + 1) - q \sqrt{\frac{1}{q(1-q)}} \sigma \sqrt{\Delta t}$$

$$\Rightarrow b = (\mu \Delta t + 1) - \sqrt{\frac{q}{1-q}} \sigma \sqrt{\Delta t}. \quad (19)$$

Substituting Equation (19) into Equation (18), we get:

$$a = \sqrt{\frac{1}{q(1-q)}} \sigma \sqrt{\Delta t} + (\mu \Delta t + 1) - \sqrt{\frac{q}{1-q}} \sigma \sqrt{\Delta t}$$

$$= (\mu \Delta t + 1) + \left(\sqrt{\frac{1}{q(1-q)}} - \sqrt{\frac{q}{1-q}}\right) \sigma \sqrt{\Delta t}$$

$$= (\mu \Delta t + 1) + \left(\sqrt{\frac{1}{q\sqrt{1-q}}} - \sqrt{q} \frac{1}{\sqrt{1-q}}\right) \sigma \sqrt{\Delta t}$$

$$= (\mu \Delta t + 1) + \left(\sqrt{\frac{1}{q\sqrt{1-q}}} - \frac{\sqrt{q}}{\sqrt{1-q}} \cdot \frac{\sqrt{q}}{\sqrt{1-q}}\right) \sigma \sqrt{\Delta t}$$

$$= (\mu \Delta t + 1) + \left(\sqrt{\frac{1-q}{q\sqrt{1-q}}} - \frac{\sqrt{q} \sqrt{1-q}}{q\sqrt{1-q}}\right) \sigma \sqrt{\Delta t}$$

$$= (\mu \Delta t + 1) + \left(\sqrt{\frac{1-q}{\sqrt{q\sqrt{1-q}}} \cdot \sqrt{q\sqrt{1-q}}} \cdot \sqrt{\Delta t}\right)$$

i.e. $a = (\mu \Delta t + 1) + \left(\sqrt{\frac{1-q}{q}}\right) \sigma \sqrt{\Delta t}$. 

20
So to summarize:

\[ a = \mu \Delta t + 1 + \left( \sqrt{1 - q} \right) \sigma \sqrt{\Delta t} \]  
(20)

\[ b = \mu \Delta t + 1 - \left( \sqrt{\frac{q}{1 - q}} \right) \sigma \sqrt{\Delta t}. \]  
(21)

We showed in the last section that \( \mu \Delta t + 1 = X \) and that \( \sigma \sqrt{\Delta t} = S \), so we can now state our new algorithm with a general \( q \) (switching back to \( u \) and \( d \) notation) as:

\[ u = X + \sqrt{\frac{1 - q}{q}} S \]  
(22)

\[ d = X - \sqrt{\frac{1}{1 - q}} S \]  
(23)

where \( \bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \) and \( S = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \ldots + (X_n - \bar{X})^2}{n-1}} \), as before.

### 2.5 Real-World Implementation

The purpose of financial mathematics is largely to make more-educated financial decisions. As noted earlier, the aim of this project is to utilize the Binomial Tree Model to analyze and predict the price movement of energy stocks. Such analysis can be used to inform future investment decisions related to this group of assets. Using historical stock prices (over the last five years) as our source of stock price ratios and the Algorithm for General Q derived in the previous section, we can compare the binomial trees produced by a variety of \( q \) values to the historical data to see which value of \( q \) produces the best-fitting tree. Over many weeks, one might expect one or a few \( q \) values to dominate in their ability to predict stock price outcomes. One could also imagine that the nodes of the predicted binomial trees could, using their attached prices and probabilities, provide expected values for the stock price throughout the week. These expected values can be fit with a line, the slope of which (which we will call “m”) can be used as an indicator of future price activity, where \( m > 0 \) indicates upward future price movement and \( m < 0 \) indicates downward future price movement. We divide the initial value of \( m \) given by this method by the beginning stock price for the week, \( S_0 \), so that \( m \) values are comparable from week to week and between stocks, which are likely priced quite differently. Financial managers could then weight their portfolios according to the predictions provided by these best-fitting \( q \) values and their corresponding expected value slopes \( m \).

The last five years (March 7, 2016 to March 5, 2021) of daily closing prices for the stocks listed in Table 1 were scraped from Yahoo! Finance (a media outlet for financial news and data) and organized into weeks according to the five business days [7]. All scraped data can be found in the Digital Appendix. Call these weeks \( W_1, W_2, \ldots, W_n \) where \( n \in \mathbb{N} \). An algorithm was written which used the stock price data from \( W_{j-1} \) as the input for \( \bar{X} \) and \( S \) in the Algorithm for General Q. Then, \( q \) values ranging from 0.01 to 0.99 in steps of 0.01 were combined with these values for \( \bar{X} \) and \( S \) to produce the upward price movement factors \( u \) and the downward price movement factors \( d \) for each different \( q \) value. These movement factors were combined with the last day of \( W_{j-1} \) (i.e.
Figure 8: Scheme of the method used for the real-life application of the Algorithm for General Q, where the stock prices for the week $W_j$ are being predicted using the stock price ratios derived from week $W_{j-1}$. This process was iterated over the entire 5-year data set.

Friday) to create the Generalized Binomial Tree for the next week predicted by each value of $q$. The nodes of the resulting trees were compared with the stock price data of week $W_j$ to determine which value for $q$ produced the best-fitting binomial tree (“best-fitting” describes the binomial tree with the smallest error, as defined in the next paragraph). This scheme is shown in Figure 8.

The best-fitting binomial tree models for each week were determined using the Python function “error_fun(q, stock, real_life)” written in error.py (see the Digital Appendix). The algorithm checks the error between each node of the binomial tree and the real-life price for that day using the following formula:

$$\text{Error} = \frac{\text{Real-Life Price} - \text{Theoretical Price}}{\text{Real-Life Price}}$$

(24)

where the Theoretical Price is the price specified by the node. The minimum errors for each day are added together to give the total error for the week.

The slopes, $m$, of the expected value linear fits of each best-fitting binomial tree were calculated and recorded. An example is shown in Figure 9. Similarly, the deviations (percent error) of the real-life stock price from the expected value at the end of the week (given by the value $S_0 + 4m$) were calculated and recorded using the following formula:

$$\text{Percent Error} = \frac{\text{Expected Value} - \text{Real-Life Price}}{\text{Expected Value}} \times 100\%.$$  

(25)

See the Digital Appendix (week_by_week_analysis.ipynb) for the Python code that was used to implement this strategy.

2.6 Statistical Analysis

Statistical analysis of the resulting data sets was carried out using the numpy Python package mean(), percentile(), and std() functions to determine the average, 25th and 7th
Figure 9: Example of best-fitting $q$ values for Weeks 4 and 21 of the ExxonMobil (XOM) data set. The 4th week of the data set (on the left) is best represented by $q = 0.66$, while the 21st week of the data set (on the right) is best represented by $q = 0.30$. Additionally, note the red line of best fit for expected values of each tree. The slope of the 4th week expected value is 0.224, while the slope of the 21st week expected value is 0.058.

percentile, and standard deviation for the data sets, respectively, as shown in week_by_week_analysis.ipynb. Statistical significance was determined using a two-tailed Student’s t-test. In all cases, the null hypothesis was that there was no significant difference between the data sets being compared. The alternative hypothesis was that the data sets were significantly different from each other. The standard deviation of Group $N$, $s_n$, was determined using the numpy.std() function mentioned previously. The standard error, $s_d$, between the groups was determined using the following formula:

$$
\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}
$$

(26)

where $s_1$ is the standard deviation of Group 1, $s_2$ is the standard deviation of Group 2, $N_1$ is the sample size of Group 1, and $N_2$ is the sample size of Group 2. The $t$-score, $t$, was calculated using the following formula:

$$
t = \frac{|\mu_1 - \mu_2|}{s_d}
$$

(27)

where $\mu_1$ is the sample mean for Group 1 (calculated with numpy.mean()), $\mu_2$ is the sample mean for Group 2, and $s_d$ is the standard error between the two groups. The degrees of freedom, $df$, were determined by adding the number of samples in each group and subtracting two:

$$
df = N_1 + N_2 - 2
$$

(28)

where $N_1$ and $N_2$ are the number of samples in Group 1 and Group 2, respectively. The calculated values for $t$ and $df$ were used with a standard t-table (such as that found at www.tdistributiontable.com) to determine a p-value for the comparison of the pair of data sets. A p-value less than 0.05, corresponding to a less than 5% probability of the differences in the data being a result of random chance, was chosen as the cutoff for statistical significance.
3 Results

3.1 Q and M Values for Classes of Energy Stocks

Using the algorithm described in Section 2.5, the weekly \( q \) and \( m \) values for the stocks listed in Table 1 were determined from the 5-year data set. See Figures 10 and 11 for the relative frequencies with which each value of \( q \) produced the best fitting binomial tree model when compared to real data for each stock. Also see Figure 13 for the frequency of \( q \) values when aggregated into the larger fossil fuel, renewable energy, and market indices groups. Note the markedly random distribution of \( q \) values in all cases. Perhaps these results are surprising, as we might expect one or a few \( q \) values to dominate in their predictive power, at least over the course of several years. Interestingly, most, but not all, stocks and indices studied have a high frequency of “extreme” optimal \( q \) values, at either the lower \( q \) limit (0.01) or the upper \( q \) limit (0.99), or both. However, these peaks are not large enough to consistently predict an optimal \( q \) value for any particular stock or index. The failure of the model to converge on any small number of preferred \( q \) values is further evidence that the traditional Hull-White Algorithm, with \( q = \frac{1}{2} \), is insufficient for predicting stock prices of energy companies. It is interesting to note, however, that the distributions of optimal \( q \) values for the four market indices studied were random as well. This result could indicate that identifying an optimal \( q \) value for a stock or a set of stocks is not a feasible strategy for more intelligent future price prediction using the Binomial Tree Model and the Algorithm for General Q.

Weekly \( m \) values provide somewhat clearer insight into stock price movement between fossil fuel and renewable energy stocks. Similarly to weekly \( q \) values, weekly \( m \) values were computed and recorded for each week within the 5-year data set for each stock, using the identified optimal \( q \) value for that week to calculate expected values. See Figures 14, 15, and 12 for the frequency at which the binomial model produced expected slopes \( m \) for individual stocks in the fossil fuel and renewable energy groups, and the market indices, respectively. All of the stocks analyzed had \( m \) value distributions concentrated around zero, while almost all of the stocks had positive average (mean) \( m \) values (the lone exception was BTU, in the fossil fuel group, with an average \( m \) slightly less than zero). See Figure 17 for aggregated group \( m \) values.

Both energy groups have positive average \( m \) values. However, the renewable energy group average \( m \) value is about five and a half times that of the fossil fuel group average \( m \) value. Practically, this translates to a day-to-day expected gain five and a half times as large for a portfolio made up of renewable energy stocks than for one made up of fossil fuel stocks. Additionally, the renewable energy group \( m \) values had greater upside (a more positive maximum \( m \) value) and also less downside (a less negative minimum \( m \) value), as shown in Table 3. It is interesting to note that the 25th percentile \( m \) value in the renewable group (-0.00517) is less than the 25th percentile \( m \) value in the fossil fuel group (-0.00491), the 75th percentile \( m \) value in the renewable group (0.00826) is greater than the 75th percentile \( m \) value in the fossil fuel group (0.00575), and the standard deviation of the \( m \) values in the renewable group (0.01756) is greater than that of the fossil fuel group (0.0157). These would all indicate that, while the stocks in the renewable
group have greater upside than those in the fossil group, they are more volatile. Similarly, the stocks in both energy groups, fossil and renewable, are more volatile than the broader market, as evidenced by their larger standard deviations (0.01756, 0.0157, and 0.00678 for renewable, fossil, and the market, respectively). Volatility equates to risk, so investors in renewable energy stocks should be aware that they take on additional risk in order to potentially achieve larger gains. This is typical of almost any investing strategy. An exception to this rule, however, is in the average (mean) performance of the stocks in the fossil fuel group: whereas the standard deviation for the fossil fuel m values is greater than the standard deviation of the broader market m values implying greater volatility, the average m value is actually greater for the broader market than for the fossil fuel group. Hence, fossil fuels may be higher risk, lower reward investments than the broader market.

It should be noted that the m value data sets for the fossil fuel group and the renewable energy group are not statistically different as determined by a Student’s t-test. Hence the differences in minimum, average, and maximum m values are within the possible range of error. However, with further sampling, it is possible that these results could tend further towards significance.

<table>
<thead>
<tr>
<th>Group</th>
<th>Minimum</th>
<th>25th Percentile</th>
<th>Mean</th>
<th>75th Percentile</th>
<th>Maximum</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Renewable</td>
<td>-0.08081</td>
<td>-0.00517</td>
<td>0.00271</td>
<td>0.00826</td>
<td>0.19575</td>
<td>0.01756</td>
</tr>
<tr>
<td>Fossil</td>
<td>-0.10077</td>
<td>-0.00491</td>
<td>0.00049</td>
<td>0.00575</td>
<td>0.11687</td>
<td>0.0157</td>
</tr>
<tr>
<td>Market Indices</td>
<td>-0.05800</td>
<td>-0.00229</td>
<td>0.00062</td>
<td>0.00375</td>
<td>0.05023</td>
<td>0.00678</td>
</tr>
</tbody>
</table>

Table 3: Metrics relating to the m values of each group of stocks. Notice that the average m value for the renewable group is greater than that of the fossil fuel group which is greater than that of the market indices. However, the standard deviations of each of the energy groups is higher than that of the broader market indices. Together, these results indicate that energy, either renewable or fossil, is a higher risk, higher reward investment than the broader market.
Figure 10: Histograms of the relative frequencies for $q$ values for the weeks of the 5-year data set of the five fossil fuel companies (XOM, CVX, COP, BTU, and ARCH).
Figure 11: Histograms of the relative frequencies for $q$ values for the weeks of the 5-year data set of the six renewable energy companies (NEE, TSLA, FSLR, RUN, CWEN, and PLUG).
Figure 12: Histograms of the relative frequencies for $q$ values for the weeks of the 5-year data set of the four most commonly cited market indices (DJIA, NDAQ, RUT, and GSPC).
Figure 13: Histograms of aggregated $q$ values for stocks in the fossil fuel group (left), renewable energy group (right), and market indices (bottom). Even with an increase in data, the distribution of $q$ values remains noisy.
Figure 14: Frequencies of weekly $m$ values for each stock in the fossil fuel group over the 5-year time period.
Figure 15: Frequencies of weekly $m$ values for each stock in the renewable energy group over the 5-year time period.
Figure 16: Frequencies of weekly $m$ values for each of the four major indices over the 5-year time period.
Figure 17: Aggregated weekly $m$ values for stocks in the fossil fuel group (left) and the renewable energy group (right). Note that the average (mean) $m$ value for the renewable energy group is roughly four times greater than that of the fossil fuel group. Also note that the maximum $m$ value for the renewable group is more positive than that of the fossil fuel group. Additionally, the minimum $m$ value for the renewable group is less negative than the fossil fuel group.
3.2 Analysis of M

To use \( m \) as an important metric for portfolio allocation, it is important to understand what \( m \) is a function of. In Section 2.5, we defined \( m \) to be the slope of the best fit line through the expected values of the stock price on each day of the binomial tree model divided by the starting stock price. Let’s consider one branch of the binomial tree model, where we have initial day \( n \) with an initial stock price \( S_n \) and two price possibilities after one time period, \( uS_n \) and \( dS_n \), where \( u \) is the upward price movement multiplier and \( d \) is the downward price movement multiplier. Then the expected value of the stock price after one time period is the sum of the possible prices multiplied by their probabilities:

\[
E = q \cdot uS_n + (1 - q) \cdot dS_n
\]

In Section 2.4, we found that \( u = \bar{X} + \sqrt{1 - q} \cdot S \) and \( d = \bar{X} - \sqrt{1 - q} \cdot S \). So:

\[
E = q \cdot uS_n + (1 - q) \cdot dS_n
\]

\[
= S_n \left[ qu + (1 - q)d \right]
\]

\[
= S_n \left[ (\bar{X} + \sqrt{1 - q} \cdot S)q + (\bar{X} - \sqrt{1 - q} \cdot S)(1 - q) \right]
\]

\[
= S_n \left[ q\bar{X} + \sqrt{q(1 - q)}S + (1 - q)\bar{X} - \sqrt{q(1 - q)}S \right]
\]

\[
= S_n \left[ (q\bar{X} + (1 - q)\bar{X}) + (\sqrt{q(1 - q)}S - \sqrt{q(1 - q)}S) \right]
\]

\[
= S_n\bar{X}.
\]

Then the expected value of the stock price after one time period is dependent on the initial stock price \( S_n \) and \( \bar{X} \). Now we have a line segment with endpoints \((n, S_n)\) and \((n + 1, S_n\bar{X})\). Then the slope of this line, \( m^* \), is:

\[
m^* = \frac{S_n\bar{X} - S_n}{(n + 1) - n}
\]

\[
= \frac{S_n(\bar{X} - 1)}{1}
\]

\[
= S_n(\bar{X} - 1).
\]

Dividing \( m^* \) by \( S_n \) gives us \( \bar{X} - 1 \). So the slope of the line from \( S_n \) on day \( n \) to the expected value for the stock price on day \( n + 1 \) is \( \bar{X} - 1 \) for each individual branch in the generalized binomial model. Then the average slope for the entire model, including all branches together, gives us \( m = \frac{k(\bar{X} - 1)}{k} = \bar{X} - 1 \), where \( k \) is the number of branches in the model. Then our value \( m \) is simply a restatement of the drift, \( \mu \Delta t \), of the stock or index in question as defined in Section 2.3. Notably, our value \( m \) is independent of \( q \). Thus fitting the historical data with an optimal \( q \) value offers no advantage when estimating the future expected value of a stock according to the Binomial Model. Instead, the predicted future value relies only on the average of the stock price ratios \( \bar{X} \) calculated from the previous week’s data.
3.3 Impact of the COVID-19 Pandemic on Energy Investments

The global outbreak of the SARS-CoV-2 virus has had obviously far-reaching effects, not the least of which has included a historic economic downturn [8]. In the early days of the pandemic, the Dow Jones Industrial Average (DJA) tumbled nearly 35% from a then-all-time high of 29,398 points on February 10, 2020, to 19,174 points on March 16, 2020, only one month later [7]. Similarly, the broader market S&P 500 index shed over 30% of its value, to fall from 3,337 points to 2,304 points in less than a month. This short stretch of time included several of the steepest daily descents in stock market history. On March 9, the DJA fell a record 2,013.76 points to close at 23,851.02. That record would be broken twice in the next week, with a 2,352.60 point loss on March 12 and a 2,997.10 point loss on March 16 each setting new records. As one might expect, energy stocks did not escape unscathed. For example, Exxon Mobil (XOM) plummeted nearly 50% from $60.34 per share to $31.45 during the month from February 19 to March 26, 2020. Meanwhile, Plug Power (PLUG) also halved in value, falling from $5.72 per share on February 19 to $2.76 on March 16, 2020.

While the market slumped, the broader economic slowdown provided opportunities for well-positioned companies to take advantage. With lockdowns in effect, some companies took the time to upgrade their infrastructure. For example, Amazon ordered over half a billion dollars worth of hydrogen fuel cells to outfit their warehouse forklifts and meet increased demand in an environmentally sustainable way [9]. Other renewable energy companies have benefited from pandemic relief and infrastructure bills that include funding for renewable energy projects across the country [10]. Meanwhile, decreased travel meant decreased demand and dropping prices for fossil fuels. Using the same approach as before, we can evaluate how these factors have impacted renewable and fossil fuel energy company stock prices following the onset of the pandemic in the United States.

The 5-year data sets used previously were divided into a pre-pandemic group (March 7, 2016 to March 10, 2020) and a post-pandemic onset group (March 11, 2020 to March 5, 2021) and the same analyses were completed as in Section 3.1. These results are shown in Figures 18, 19, 20, 21, 22, 23, 24, and 25.

Perhaps the most noticeable trend following the onset of the pandemic is the broadening of the distributions of \( m \) and a subsequent increase in standard deviations from pre- to post-pandemic. This difference is particularly stark in the fossil fuel group, which saw its standard deviation nearly triple from its pre-pandemic value (0.00959) to its post-pandemic value (0.02736), although this is true of the renewable group and broader market (whose standard deviations each nearly doubled) as well. The time period since March 11, 2020, though volatile, has been a particularly cheery stretch for stocks across the board, as the average \( m \) value of each group more than quadrupled from the time period before the outbreak of the pandemic. The uncertainty of much of the past year has bled over into the financial world, leading to large swings in both directions and big gains for patient investors.
Figure 18: Frequencies of weekly \(m\) values for each stock in the fossil fuel group pre-COVID (before March 11, 2020).
Figure 19: Frequencies of weekly $m$ values for each stock in the fossil fuel group following the onset of COVID in the United States (after March 11, 2020).
Figure 20: Frequencies of weekly \( m \) values for each stock in the renewable energy group pre-COVID (before March 11, 2020).
Figure 21: Frequencies of weekly $m$ values for each stock in the renewable energy group following the onset of COVID in the United States (after March 11, 2020).
Figure 22: Frequencies of weekly $m$ values for each of the four major indices prior to the onset of COVID in the United States (before March 11, 2020).
Figure 23: Frequencies of weekly $m$ values for each of the four major indices following the onset of COVID in the United States (after March 11, 2020).
Figure 24: Aggregated weekly $m$ values for stocks in the fossil fuel group (left), the renewable energy group (right), and the broader market indices (bottom) prior to the onset of COVID in the United States (before March 11, 2020).
Figure 25: Aggregated weekly $m$ values for stocks in the fossil fuel group (left), the renewable energy group (right), and the broader market indices (bottom) following the onset of COVID in the United States (after March 11, 2020).
4 Conclusion

Beginning with the commonly used binomial tree model, we developed an analog to the Hull-White Algorithm with lightened restrictions on the value $q$, the probability that the price of a stock will rise from the beginning to the end of one time period. This new “Algorithm for General Q” showed significant improvement and flexibility in the fitting of the generalized binomial model to real world data than the canonical Hull-White Algorithm when applied to a 5-year set of stock closing price data. However, the optimal $q$ value from week was not robust for any single stock or any group (renewable, fossil, or broader market) of stocks. Thus, while the adapted algorithm allows improved analysis of past results, it fails to provide significant predictive power.

Additionally, using the slope ($m$) of a linear fit of the expected values predicted by the general binomial model from week to week, the day-over-day expected growth was calculated for each stock. These $m$ values revealed that the stocks in the renewable group were consistently expected to outpace both their rivals in the fossil fuel group and the broader market, albeit with somewhat more volatile results. These results indicate that, in addition to providing a social good, directed investment in next-generation renewable energy companies can be profitable when compared to investment in old-school fossil fuel companies and can even beat the broader market by a substantial margin.

Lastly, the effects of the COVID–19 pandemic on energy stocks were studied. Stocks in all categories (renewable, fossil, and broad market) were impacted. The most noticeable effect on energy stocks was the increase in volatility, represented by widened $m$ value distributions and larger standard deviations of $m$ values. Despite heightened volatility (risk), the market—and renewable energy stocks especially—have been very successful over the course of the pandemic, spurring hope that the post-pandemic recovery will be short.

There are several possible future directions for this work. The results given by the binomial tree model in this work could be compared to the predictions made by other models. One model of particular interest is Geometric Brownian Motion (GBM), the continuous variable model at the heart of the Black-Scholes model [11]. Perhaps GBM will provide the predictive insights that the binomial model was unable to. Another possible line of inquiry could be the investigation of how the present values of options given by the Hull-White Algorithm compare to current options prices on the market.

The development of renewable energy systems will be essential for human civilization to persist into the coming centuries. As the premier threat facing us today, climate change has and will continue to motivate investment into these technologies. So, it turns out, might profit.
5 Digital Appendix

All data and code used in this project can be found at the following link:

GitHub

or

https://github.com/ghamrick34/Math-Senior-Thesis-Digital-Appendix
References


