2022

How Punxsutawney Phil’s Predictions Affect the Stock Market: A Groundhog Day Analysis

Jessica Cuna Zamora

Claremont McKenna College

How Punxsutawney Phil's Predictions Affect the Stock Market: A Groundhog Day Analysis

Submitted to
Professor Hughson

By
Jessica Cuna Zamora

For
Senior Thesis
Spring 2022
April 25, 2022

Acknowledgements

First and foremost I dedicate this thesis and my entire degree to my parents, Jose & Josefina Cuna. I'm only here today because of your sacrifices and unwavering faith in me. Que Dios les bendiga y les repage por todo lo que me has dado, gracias por siempre estar ahí conmigo y por ayudarme a lograr mis sueños. To my sisters, Celia & Diana, for a variety of reasons, including but not limited to: paying for my phone bill, occasionally treating me to boba, sending memes, rating my outfits, and most importantly, sending me pictures of our beloved dog Bambi. Speaking of which, a huge shoutout to my dog for being the most adorable angel alive that never fails to brighten my day. My family has always been my biggest cheerleaders and I will never be able to completely express how much they have impacted me and how much I love them.

Thank you to all the people I've met at CMC in these four years including professors, staff, faculty, friends and strangers. In particular, thank you Professor Hughson for inspiring and guiding me through this thesis; you have the patience of a saint and I couldn't imagine a better topic to write about. My deepest gratitude to my friends that have been there through thick and thin; you all have given me the most precious memories and have put up with so many of my shenanigans and general tomfoolery.

Abstract

With its roots dating back to 1887, Groundhog Day has cemented itself as a beloved American holiday where people gather to see if Punxsutawney Phil will predict 6 more weeks of winter, or an early arrival of spring. Utilizing the data and methodology framework from Shanaev, Savva, and Fedorova (2021) to test if Groundhog Day predictions have any effect on S&P 500 returns, this paper revisits and revises the analysis in attempt to replicate and improve the original findings with a dummy-variable regression while controlling for other calendar anomalies. Additionally, this study expands the original analysis by including two new tests: a weather effect analysis and a steel industry specific analysis. It is found that Groundhog Day predictions do not create statistically significant returns under either prediction, despite the findings of the base paper that find significant returns for early spring predictions.

Table of Contents

1. Introduction

The American stock market is a representation of its investors and its market-makers. There are an immense amount of factors that determine the day-to-day movement of the market including war, economic conditions and prospects, international transactions, and more. All of these large factors are naturally important and thought-provoking, but based on financial literature there may be other underlying factors to the market that can cause abnormal returns even if they are not common, for example, Groundhog Day.

The legend of Groundhog day dates back to the Christian religious holiday of Candlemas, where Christians would take their candles to church on February 2nd to be blessed for the remaining winter. Later, the Germans introduced a weather-predicting animal, the hedgehog, to see if there would be a "second winter" in conjunction with the same holiday. When the European settlers arrived to the New World, they wanted to continue their same traditions, but could not find a hedgehog. As a replacement, the overgrown, hibernating rodent of the groundhog was chosen to continue the holiday in 1886 and has since formed the Groundhog day we know today.

The original groundhog, Punxsutawney Phil, will be the focus in this analysis due to his long history of predictions. Each prediction falls into one of two categories: "Early Spring" since Phil does not see his shadow, or "Long Winter" because Phil does see his shadow and expects 6 more weeks of winter. The focus of this paper is to analyze returns surrounding Groundhog Day events. More specifically, the paper analyzes the difference in returns between Punxsutawney Phil's prediction of either an early spring or a long winter. This choice in holiday may seem random or a bit too niche, but a paper written by Shanaev, Savva, and Fedorova (2021) found that there are statistically significant results surrounding early spring prognoses for the S&P 500, hence the motivation to revisit and revise the original study. Additionally, a deep thank you to the original authors for not only providing a riveting topic to study, but also for providing the original data set and code to help build this revised analysis.

Shanaev, et al. (2021) utilize S&P 500 returns from 1928-2021 to find that there are statistically significant positive returns to the market for early spring predictions of 1.85%, 1.15%, and 2.78% for the event windows of [-10; 10], [-10; 1], and [-30; 30] respectively. As for long winter predictions, they find no significant returns, but they do observe a trend that long winter drives moderately negative returns under almost all event windows in their analysis, and vice versa for early spring. Their methodology for measuring returns uses a buy-and-hold strategy instead of the more usual average returns over the event window, in addition to "normalizing" returns using a threshold garch model (TGARCH).

This study is a revision of a previous paper that studied the same question. Techniques, methodology, and nearly the same dataset surrounding S&P 500 results are based on the original paper, however this study also includes brand-new analysis to understand the relationship between market returns and "correct" weather predictions and market returns for the steel industry. Using market return data for the S&P 500 between 1928-2021 and a dummy-variable regression yielded very few significant results for early spring prognoses at the 10% significance level, but this significance went away once the full-control regressions were introduced. There was also no significance in results found in sub-periods of the dataset.

These results are in direct opposition to previous literature which lists that early spring prognoses should have statistically significant positive results. Although Shanaev, et al. (2021) utilized a buy-and-hold method to calculating returns instead of the more usual average returns over the event window, there is a greater issue that lies within the coding in their dataset. Dr. Shanaev, a co-author of the original paper generously provided the original code and dataset for this paper. While analyzing and attempting to re-run their code, it was discovered that their dummy variable for prediction was coded in a way that may explain their statistically significant results. Using only one variable, they designated a -1 value for long winter predictions and a +1 value for early spring predictions. This forces one prediction type, like long winter, to have a symmetric effect on results. Meaning both predictions have an equal and opposite effect in the dataset, causing the standard error to not have the proper (higher) value it should have for early spring predictions (since there are considerably less observations and should exhibit a larger standard error), and therefore causing the early spring predictions to show significance.

The key difference in this analysis is the change of coding format for the prediction variable. Two separate long winter "LW" and early spring "ES" variables were created to equal 1 only for the the years and event window days where appropriate and 0 otherwise for every other day. This way, the dummy variable is only capturing the effects of only the appropriate predictions and years, so one prediction type is not influencing the effect of the other. A t-test comparison between the prediction dummy variable coefficients is also utilized to test if each return is statistically significantly different from zero, which it is found that they are. The finding of the returns being statistically significant prove that the original coding format was incorrect because one cannot assume a symmetric effect for both prediction types.

For the main regression results, there is evidence of significance between groups at the 10% significance level and in only two of three event windows. This is not the same result as the original paper, which found statistically significant differences at the 1% and sometimes the 5% level in almost all of their 7 event windows. Though there are some promising trends, it is not enough and not strong enough to form the conclusion that Groundhog Day predictions drive results.

The additional tests conducted serve to add on to the first set of results and to broaden the scope of how, where, or why Groundhog day may have an effect on the market. First, a weather-effect framework is considered to test if the years in which Phil was correct with his weather prediction drive any results. Although there is one instance of statistical significance with controls for an early spring prediction under the [0; 1] event window, unfortunately, there continued to be no significant results for every other window and prediction. Additionally, instead of seeing if there was an effect on the overall market, the steel industry returns replaced the S&P returns to test if an industry that is important to Pennsylvania (Phil's home state) would show different results. Once again, there were no statistically significant results from this analysis for either combination of event window, prediction type, or sub-period returns.

For the format of this study, section 2 will outline important previous literature and background to the holiday. Section 3 presents details on the different data sets compiled and information on the tests run for the study. Section 4 presents results and discussion for all the tests. Finally, Section 5 presents a summary for all the findings to the paper and ideas for future studies. Additional tables and pictures are found in the section 7 appendix.

2. Literature Review

2.1. Main references

There are very few studies addressing the question of whether or not the market reacts to a groundhog's prediction, therefore my work will revisit and build upon a recently completed study on the same topic.

Shanaev, Savva, and Fedorova (2021) analyzed stock market returns from 1928-2021 around the event period to see if there were any abnormal returns. Their data included daily U.S. stock market data (S&P 500 returns) and had various event periods ([-1; 1], [-10; -1], [1; 10], [-10; 10], [-30; -1], [1; 30], and [-30; 30]) surrounding the event date of February 2nd. Through their research, they were able to find a market anomaly surrounding Punxsutawney Phil's predictions when focusing on the early spring prediction (no shadow) years and under almost all event periods, minus [1;30]. The authors then focus their analysis down to the [-10; 10], [-10; 1], and [-30; 30] windows for early spring predictions, which showed a statistically significant market appreciation of 1.85%, 1.15%, and 2.78% respectively. On the other hand, when Phil predicted a long winter (did see his shadow), there is an implied negative return of -0.644%, -0.043%, for the [-10; 10] and [-10; 1] windows.

Both of these scenarios and returns may help outline persistent investor and market irrationality surrounding calendar events because of potential investor anticipation surrounding Groundhog Day. This may be driven because of excitement surrounding the holiday and the potential to either be able to "bet" correctly on the prediction, or investors expecting good news of an early spring (hence the significance since no one wants to suffer through more winter). They also claim their robust findings cannot be explained by the January effect, the "Halloween Indicator," turn-of-the-month effect, or other seasonalities.

Additionally, Shanaev, et al. (2021) draws upon the event study method as explained in MacKinlay (1997). Specifically, the conventional constant return model methods in the paper drive the creation of the S&P 500 quasi-portfolio to calculate buy-and-hold abnormal returns. Essentially, they estimate the buy-and-hold returns for their given event window and prediction to estimate significance. Additionally, the paper uses the threshold garch (TGARCH) model to normalize returns and control for heteroskedasticity. Since Shanaev, et al. (2021) heavily uses this method for their statistical models and calculations, it is important to note since this paper will attempt to mimic their method and results for a first-round of results, with a few changes and improvements.

Klopfenstein (2017) also attempts to measure the impact of Groundhog's Day -and the weather effect- on market returns. Their data includes daily returns on the S&P 500, Dow 30, and NASDAQ 100 from 1965 to 2015. Their analysis shows "a significant difference in annual performance for the S&P 500 (10.2 percent stronger), Dow 30 (13.8 percent stronger) and NASDAQ 100 (13.7 percent stronger) when Punxsutawney Phil predicts, "spring is just around the corner.'' Their findings may suggest that the weather effect, where investor's moods are positively influenced by good weather, may be at play due to the dramatically higher and statistically significant returns for the six week event period. Additionally, they tested to ensure that the January effect did not influence results, which is believed to influence the whole year. Though this paper is not central to the analysis presented here, it is important to note that this is a valid research question that has been tackled by professionals and students alike.

2.2. Groundhog Day History and Important Details

The famous, all-american holiday is the center of the analysis for this paper, therefore it's critical to understand its roots and actual purpose before applying any statistical tests to it. Based on information from The Punxsutawney Groundhog Club (2022), who are responsible for organizing the annual event, the tradition of Groundhog day began in 1887 in Punxsutawney, Pennsylvania and has a rich history surrounding the birth of the day.

Figure 1: Punxsutawney Phil's handler A.J. Dereume holds the famous groundhog during the 136th Groundhog Day, at Gobblers Knob in Punxsutawney, Pennsylvania, U.S.

Freed (2022)

The legend began as a celebration of the Christian religious holiday of Candlemas Day in Europe, which is held on February 2nd and involves Cristians taking candles to church to have them blessed. Families wanted to bring blessings to their home for the remaining winter and over time an aspect of weather prediction was added from an old English folk song (Punxsutawney Groundhog Club, 2022):

If Candlemas be fair and bright,

Come, Winter, have another flight;

If Candlemas brings clouds and rain,

Go Winter, and come not again.

This song and interpretation of the day became the norm to follow. Germany later introduced weather-predicting animals into the mix to richen the lore. According to their updated interpretation, on February 2nd if the chosen hedgehog saw its shadow there would be a "second winter." Once settlers arrived in the United States, the absence of hedgehogs caused the groundhog, another hibernating animal, to become the new symbol.

Therefore, since 1886, Punxsutawney has had the pleasure of hosting the tradition and Punxsutawney Phil (the groundhog), was named the "Weather Predictor Extraordinaire" according to the official Punxsutawney Groundhog Club (2022) and Inner Circle (the group of local dignitaries responsible for carrying on the tradition and taking year-round care of Phil). To reiterate, every year on February 2nd, Phil emerges from his burrow and either does or doesn't see his shadow. If Phil does see his shadow, he is predicting six more weeks of winter. If he does not see his shadow, he is predicting an early spring. Additionally, according to legend, Phil is the exact same groundhog that was chosen that fateful year in 1886. In other words, Phil is immortal and has prediction abilities, making him superior to every other wild or in-captivity groundhog, who only have lifespans of approximately 6-14 years (depending if they are in captivity or not). If there were to be rumors of another Phil, or another groundhog to be named Phil, they are all imposters since only the real Phil is sustained by the "elixir of life."

Exactly how Phil came across either the recipe or a lifetime supply of the elixir is as mysterious as the dark side of the moon. However, if the legend as written in Harry Potter is correct, using the power of the Philosopher's Stone allowed for the creation of said elixir, which grants the user immortality as long as they continue consuming the concoction (Rowling, 1999). It is reasonable to assume that the great witch or wizard that created the elixir, like Nicholas

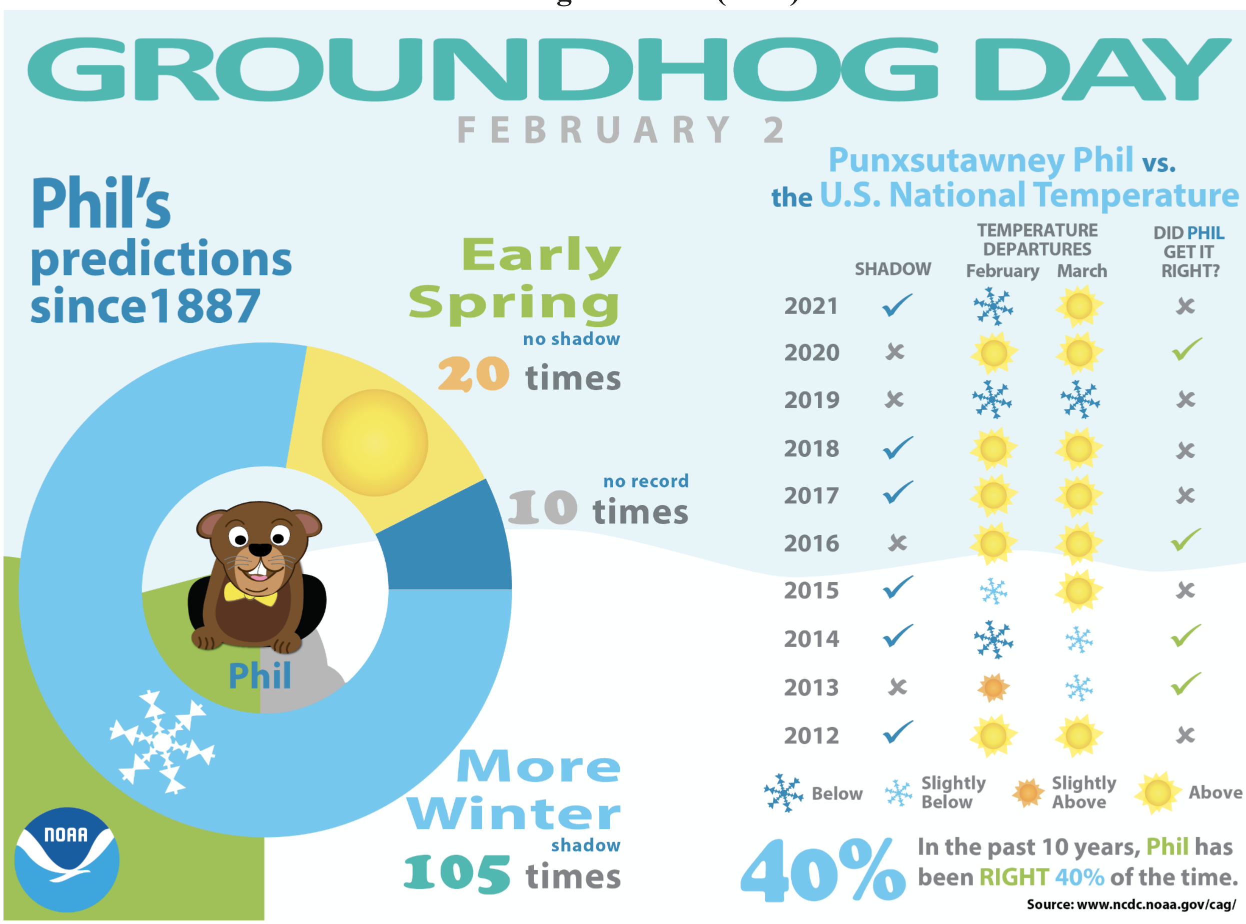
Flamel and his wife, would want to grant one very special entity a chance to live forever. Frankly, there was no better choice than the symbol of the New World and weather predictor extraordinaire himself to have the honor of continuing his legacy for centuries to come.

Over the years, fascination and jealousy have brewed around Phil's celebrity status in the U.S. that has drawn many copycat groundhogs. The Reader's Digest (Bryant, 2021) list includes, but not limited to: Dunkirk Dave (New York, 58 years predicting), Buckeye Chuck (Ohio, predicting since the 1970s), General Beaureguard Lee (Georgia), Staten Island Chuck (New York), Chuckles IX (Connecticut), Chattanooga Chuck (Tennessee), Thistle the Whistle-pig (Ohio), Sir Walter Wally (North Carolina), Pierre C. Shadeaux (Louisiana), and Grover the Groundhog (Pennsylvania, since 2006). Though all of these other groundhogs exist and each have their own track record of predictions and accuracy rates, it's important to note that this paper only focuses on the OG Punxsutawney Phil since the other groundhogs don't have nearly the same amount of years on the job and they are replaced by another groundhog after death (though most continue sharing the same name, only Chuckles adds the suffix of which exact number groundhog he or she is).

Within Phil's long reign as the top animal with predicting powers, society has seen a rise in the popularity of other animals that claim to be able to predict events. For example, there is a wooly bear caterpillar that also predicts how long and harsh winter might be, Catfish in Japan that can sense earthquakes, Oscar the cat in Rhode Island that can predict an elderly person's death, and even Paul the German Octopus who has correctly predicted the winners of the World Cup 8 times (Lombardi, 2012). Animals that can predict sporting event winners are particularly popular in the press since people actively and openly place bets on their favorite team, which allows these animals the power to influence people's decisions. Though humorous in nature, the existence of these animals and trends is an important value-add to understanding how potential investors may react to the idea of "placing a bet" on the occurrence of an event or outcome. Hence the strong motivation behind researching the question of how, or if, Phil the Groundhog moves the market.

2.3. Predictions and Accuracy Rate

Figure 2: Example of Punxsutawney Phil's accuracy rate and past predictions according to NOAA (2022).



Barker (2020)

As with any famous entity, Phil is subject to scrutiny and doubt by the general public. They doubt if Phil can really predict the weather, or if his prognosis is just left up to chance or extraneous factors (e.g. camera flashes). The Inner Circle confidently reports that Phil has a

100% accuracy rate with his predictions, though third-party sources most commonly rate him with a 39% accuracy rate (Stormfax Weather Almanac, 2022). However, how such estimates were calculated are difficult to determine. There is no formal site listing for which years Phil was correct or not, or any information as to how to determine if Phil was correct. There only exists a small study using data from the (NOAA) that determines that in the past 10 years, Phil has been correct 50% of the time.

As a better idea of Phil's entire history of predictions, Table A.1. lists them all out. From 1887-2022, Phil has predicted six more weeks of winter (shadow) 105 times, early spring (no shadow) 19 times, and no record 11 times (no records usually happened during war periods since Phil is a true patriot and did not find it appropriate to predict during dire times). These predictions were sourced from the Stormfax Weather Almanac (2022) and cross-referenced with the past predictions from The Punxsutawney Groundhog Club (2022) site. 1942 was listed as "partial shadow" on both sites, but since it was not definitive it will be listed as no record. The 1943 prediction is listed as long winter on the almanac site and as no record on the club site, therefore it will also be listed as no record so it will not sway results later on. 2022 will not be included for the statistical tests, but is included in the table.

3. Data and Methodology

3.1. Main Study

This study will utilize the past Groundhog Day predictions from 1928-2021 for Punxsutawney Phil exclusively. To reiterate why, Phil is the most reported and followed groundhog and has the longest recorded record. Therefore, it is reasonable to assume that investors will follow and be most moved by Phil's predictions, so it is the most important for investor sentiment and mood. This subset of predictions will be paired with daily return data from the S&P500 for the same period of time since market data only goes back as far as 1928, even though prediction information goes as far back as 1887. In this summarized set of data, Phil predicted long winter 75 times, early spring 17 times, and there was no record 2 times (no record years were dropped from the dataset).

The original dataset was sourced from the original author of the first and only published Groundhog Day paper (Shanaev, et al., 2021) and was modified to work with STATA. An extra binary variable was created to separately indicate early spring (ES) and long winter (LW) in order to separately measure the effect of both. Additionally, the exact prediction for the year 1955 was changed from early spring to long winter since both original sources, like the Punxsutawney Groundhog Club and the Stormfax Weather Almanac report that Phil saw his shadow that year.

This study attempts to estimate the average abnormal returns surrounding the event on February 2nd (or the next trading day after). Therefore, an event study-like methodology is applied. To assess the impact of Groundhog Day on the market, the following main dummy variable regression models were estimated, which regression 3 & 4 also control for other calendar anolmolies:

(1) $$R_t = \mu + \beta_1 ES_t + \epsilon_t$$

(2) $$R_t = \mu + \beta_1 LW_t + \epsilon_t$$

(3) $$R_t = \mu + \beta_1 ES_t + \beta_2 M_t + \beta_3 F_t + \beta_4 J_t + \beta_5 ToM_t + \beta_6 H_t + \beta_7 SiM_t + \epsilon_t$$

(4) $$R_t = \mu + \beta_1 LW_t + \beta_2 M_t + \beta_3 F_t + \beta_4 J_t + \beta_5 ToM_t + \beta_6 H_t + \beta_7 SiM_t + \epsilon_t$$

Where R_t is the daily average abnormal returns for the S&P 500 for early spring (ES_t) and long winter (LW_t) indicator prognostications for the event windows [-1; 1], [0; 1] and [-10; 10]. The control variables of M_t, F_t, J_t, ToM_t, H_t, and SiM_t are dummy variables for Monday, Friday, January, turn-of-the-month, holiday, and Halloween ("Sell in May and go away") effects respectively (Shanaev, et al., 2021). μ and $_i$ represent the intercept and the dummy variable coefficients, which will assess the impact of Groundhog Day prognostics on the stock market. $_1$ will be the coefficient of interest for both indicator variables within the results section.

The event windows chosen for this study of [-1; 1], [0; 1] and [-10; 10] are not in-line with more traditional event studies with windows of around 30 days post and pre announcement (MacKinlay, 1997). However, the new windows are a better representation of when and for how long we would expect to see the announcement effects. Previous literature utilized up to a [-30; 30] event window, though it is unclear if the significance it was picking up was truely from Groundhog Day alone. It's difficult to exactly determine when or for how long investors begin and stay interested in Phil's predictions, but it arguably is not for very long considering Valentine's Day is held within the same month, and frankly Phil is most likely not a huge discussion point too much outside of his holiday.

Therefore, this study re-defines what the "upper bound" of event windows should be by designating [-10; 10] as the longest one. The previous literature for Groundhog stock market anomalies highlight how the 10 days before and after the event capture the greatest effect of the announcement and the pre-event movements. That is, even if Phil's prediction isn't necessarily a "surprise," 10 days before the event seems like a reasonable amount of time for investor build-up and bet placing since the event involves a longer-term post-event bet, and therefore they need time to research and investigate. Meaning that two trading weeks before the event is not an unreasonable amount of time to start checking the weather to make an informed bet. That being said, two short-term event windows are also included to measure the impact of the prediction itself and its subsequent reaction from investors.

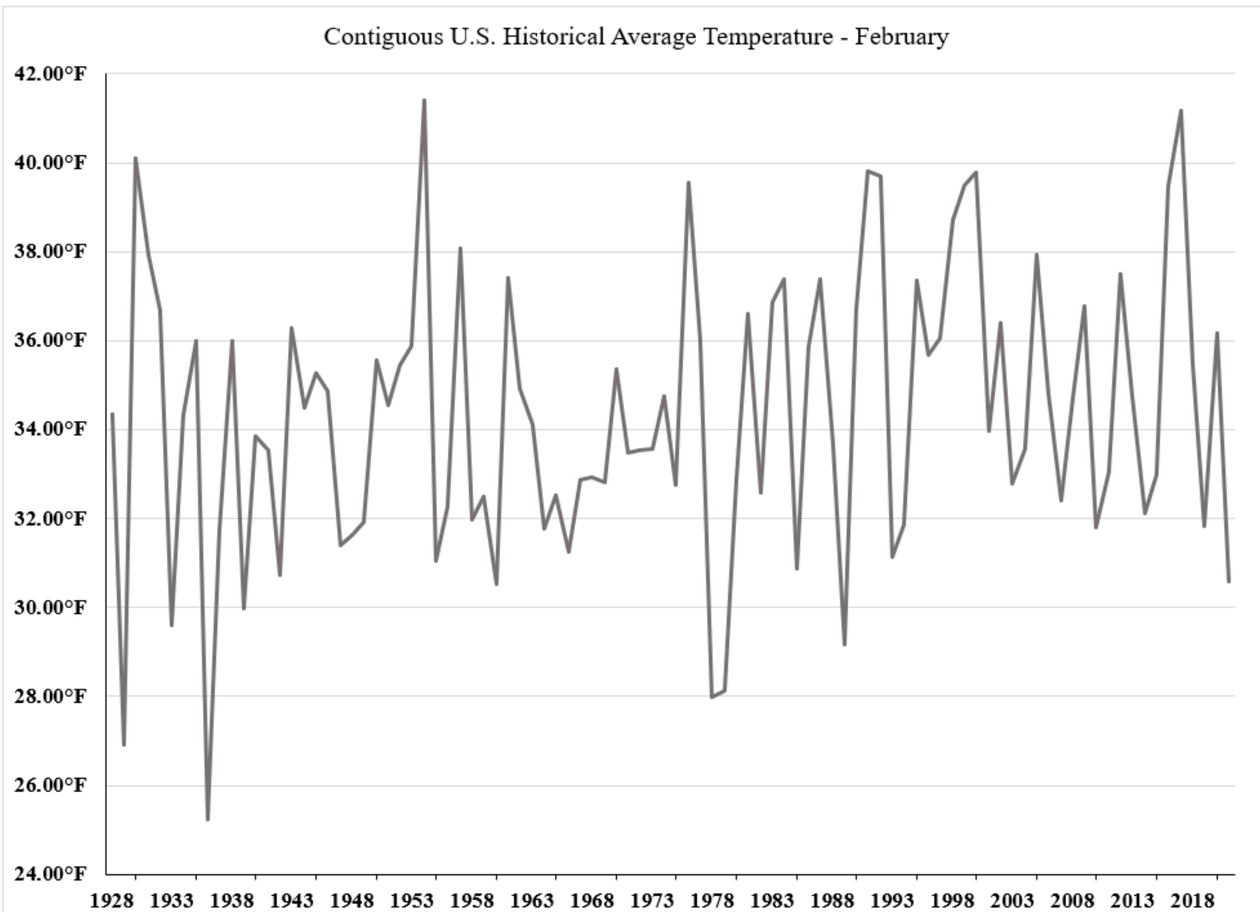
Additional changes from previous literature include changing the coding of the dummy variables and the measurement used to capture returns. First, from the provided data set belonging to the original author, the dummy variable for a long winter prediction was equal to -1, while early spring was given a +1 value. This coding was changed because the -1/+1 suggests that one prediction type, like long winter, has a symmetric effect on results. Meaning both predictions have an equal and opposite effect in the dataset, causing the standard error to not have the proper (higher) value it should have for early spring predictions (since there are considerably less observations and should exhibit a larger standard error), and therefore causing the early spring predictions to show significance.

The revised data set uses two dummy variables, "ES" for early spring and "LW" for long winter, where either variable equals +1 if that specific year had a long winter/early spring prediction for the given event period. T-tests of the coefficients of the no-control regressions (1 & 2) are also done to confirm the difference in-between group significance and effect. Additionally, the previous paper used Python to calculate buy-and-hold TGARCH returns, however this study will utilize STATA to calculate average returns over the event window. The benefits of using TGARCH model are unclear in the scope of this analysis, so using unaltered day-to-day market returns will be used instead.

3.2. Supplemental Data, Tests, & Assumptions - Weather Effects

For additional statistical tests to supplement the main results, data for weather is needed and required.

Figure 3: Historical U.S. average temperature for the contiguous states for the month of February 1928-2021.

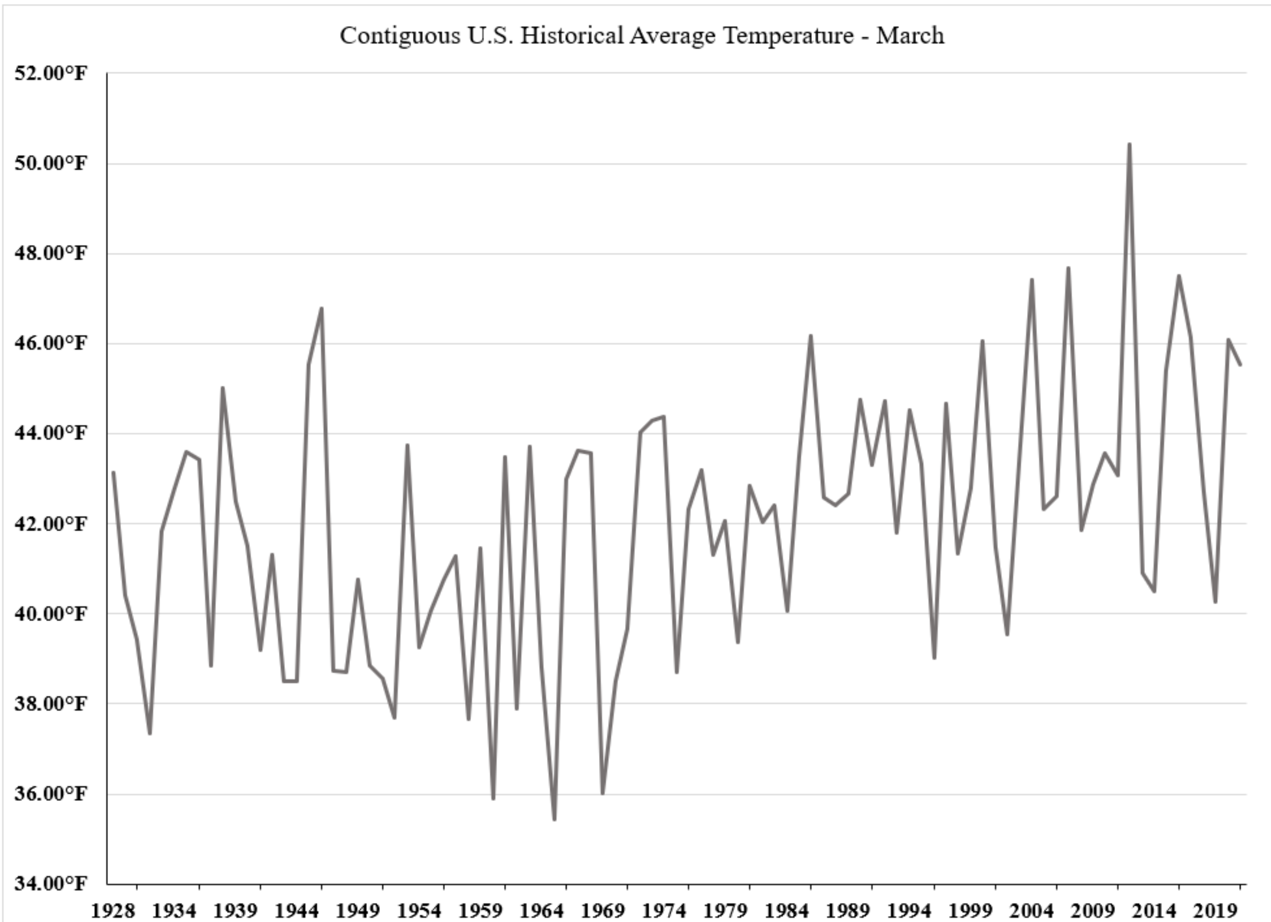


NOAA (2022)

For the tests that attempt to measure the relationship between Phil's "correct" predictions and market returns, weather data for the months of February and March of every year are needed.

National average weather data from the National Oceanic and Atmospheric Administration (NOAA) from 1928-2021 were collected for the contiguous 48 states. Graphically, the 1-month average temperature for the months of February and March for the time period of 1928-2021 are depicted in Figures 3 and 4.

Figure 4: Historical U.S. average temperature for the contiguous states for the month of March 1928-2021.



NOAA (2022)

These temperature averages are used to mimic the process from Figure 2 to decide if Phil was "correct" or not from his original prognosis according to Table 2. The general average from 1928-2021 for February and March were first separately calculated. "Correctness" was determined if either February or March of a given year was above average in temperature and if

Phil predicted an early spring. Or, if both months were calculated to be below average in temperature and Phil predicted a long winter, then it was determined that he was correct. Table A.2. outlines the years that Phil was deemed to be correct.

However, this study will deviate from the strategy used in Figure 2 in order to better fit the event windows used. Though it is important to acknowledge March for the overall prediction, in a retrospective view, or "perfect foresight," with event windows of either [-1; 1], [0; 1] and [-10; 10] it is not the best fit to determine "correctness." For the revised version only February will be considered. Using weather and climate data from the National Centers for Environmental Information (2022) for February from 1928-2021, a moving average is used to determine if a particular year was above or below average. The moving average adds 0.15℉ per decade (0.015℉ per year) to account for global warming and create a better fit for realistic expectations versus using a straight average from such a long period of time. Additionally, the added trend helps account for the wide variation in average temperature for a specific month year to year. The inclusion of the trend makes the total number of "correct" years 54 instead of 52 (without trend). Although this does not seem like a drastic change, with such a small dataset every year counts, additionally it proves that the inclusion of an adjustment for global warming does make a difference. The exact rate of global warming can also be adjusted higher/lower depending on personal preference or previous knowledge.

Table 2: Revised details and conditions to determine Punxsutawney Phil's correct predictions.

Correct Long Winter Prediction: February - Below average temperature for a given year
Correct Early Spring Prediction: February - Above average temperature for a given year

In order to fully apply the weather data and correctness decisions, one important assumption must be made: Punxsutawney Phil has full agency over each of his prognoses. There are no other studies that have a framework of how to approach this specific weather-related study this paper will explore, therefore the question of whether or not Phil has control over his predictions is an important component to consider.

In more contemporary versions of the holiday, Phil emerges from his burrow at approximately 7:25am on February 2nd to give his prediction. Although being forcefully awoken by the sound of singles shaking right outside your front door does not necessarily suggest that Phil has a choice in all of his decisions, an older version of the tradition provides better insight to what process Phil usually takes before predicting. Before the mid-1960s, Phil's time of emergence was a lot more sporadic, emerging as early as 7am and as late as 11am. This means that Phil, while in his burrow, could wait for the best time to emerge based on his knowledge of the future and current temperature conditions on that morning. For his later appearances, it could suggest that he was waiting for the day-of weather to change in order to match his prediction, and similarly for his early appearances.

The important takeaway from this assumption is that Phil has complete agency over his decisions and makes all of them himself. That is, the members of the Inner Circle who care for him, or anyone else around him are not the ones influencing or changing Phil's expert predictions.

Specifically, the data set reflects only the years that Phil predicted either an early spring or long winter correctly, all other years are left out of the analysis to attempt to conclude if investors care if he was correct or not. That is, if investors believe in Phil's prediction powers, the years that they are more confident in their weather predicting powers given the current conditions of February, should show a greater movement within the market.

3.3. Supplemental Data, Tests, & Assumptions - Steel Industry

The final additional test for this paper revolves around testing to see if there are any statistically significant returns for the steel industry. The motivation behind this choice is that, since steel is one of the largest industries in Pennsylvania, we could potentially see changes in returns over the same event window because a large number of people who might invest in this industry might also live close to Phil, or at least pay more attention to his predictions.

The exact same methodology from section 3.1 apply here, the only difference being that the dependent variable of returns shifts from returns on the market (R_t) to returns of the steel industry (SR_t). The exact same regressions and control apply here.

Steel industry data and returns is sourced from the Fama French definition of the Steel Industry (industry number 19) from the period of 1928-2021. Control variables relating to a specific day of the week or month are also adjusted to fit the new dependent variable. Results are presented for the entire period of 1928-2021 in addition to two sub periods of 1928-1944 and 1945-2021. Since the steel industry began to move westward to the Midwest post WWII, a sub-period analysis is appropriate to see if there are any drastic differences in returns due to the economic shift of the center of the steel industry.

4. Results and Discussion

4.1. Main Results

Using the previously mentioned event windows and dummy variable regression models, the results for early spring and long winter predictions are summarized in Table 4.

Table 4: Average daily returns via dummy variable regressions 1-4 around Groundhog day predictions within different event windows for time period 1928-2021.

Regression Prediction	(1) Early Spring	(2) Long Winter	(3) Early Spring	(4) Long Winter
Event Window	[-1; 1]	[-1; 1]	[-1; 1]	[-1; 1]
Coefficient	0.2693	-0.0210	0.2251	-0.0716
Std. Errors	(0.1680)	(0.0803)	(0.1683)	(0.0814)
Event Window	[0; 1]	[0; 1]	[0; 1]	[0; 1]
Coefficient	0.3387*	-0.0860	0.3013	-0.1291
Std. Errors	(0.2057)	(0.0982)	(0.2058)	(0.0992)
Event Window	[-10; 10]	[-10; 10]	[-10; 10]	[-10; 10]
Coefficient	0.1068*	-0.0291	0.0890	-0.0551
Std. Errors	(0.0637)	(0.0313)	(0.0647)	(0.0338)

*Coefficient results are presented in percent form, standard errors reported in parentheses, and statistical significance at the 10% level denoted with *.*

Table 5: T-test results for the difference between long winter vs early spring coefficients (no control regression coefficients).

T-test: $\beta_{LongWinter} - \beta_{Early\ Spring} = 0$

Event Window	[-1; 1]	[0; 1]	[-10; 10]
T-stat	1.559	1.860	1.887

When considering the differences between early spring and long winter predictions, running a no-control regression (columns 1 and 2) does give us a statistically significant result for early spring prediction returns for the [-10; 10] and [0; 1] windows. Table 5 also conveys similar information by showing the significant t-values for the same windows at the 10% level. This means that the two groups are different from each other, which proves that the original coding in the data set is incorrect given that long winter and early spring do not have the same effect on the market. For clarity, Table 5 t-tests do not test the basic significance of the coefficient, what it does is test if the difference between the coefficient's effects are statistically significant. That being said, when a full-control regression is run, the early spring prediction loses its significance and the long winter returns continue to be insignificant in all event windows.

These results become more interesting when we consider the previous literature results. The authors found that early spring predictions always had statistically significant results under every event window except [1; 30]. They received a positive return of 0.7289% in the [-1; 1] window and 2.5384% in the [-10; 10] window, both being significant at the 5% level. What the authors fail to do is give details as to which regression they are running to get those results. Although we do not get the same percent return since we each calculate them differently, the [-10; 10] event window is significant at one point in both analyses (the [0; 1] window included here is not tested in the previous paper, although still significant).

As further analysis, the sub-period of 1945-2021 is studied using the same regressions and methodology as before. Table 6 outlines the results of the regressions and provides an additional check that shows the significance in the [-10; 10] event window persists (with no controls). Interestingly enough, the [-10; 10] provides another surprise: regression 4 results are statistically significant at the 10% level with a negative market return of -0.0608% in the sub-sample. Table 7 T-test results confirm that the [-10; 10] event window is considerably "worse" when comparing the two dummy variable coefficients so much so that it is significant. However, these results alone are not enough to drive the conclusion that Groundhog Day has a very strong, statistically significant effect on the S&P 500.

Table 6: Average daily returns via dummy variable regressions 1-4 around Groundhog day predictions within different event windows for time sub-period 1945-2021.

Regression Prediction	(1) Early Spring	(2) Long Winter	(3) Early Spring	(4) Long Winter
Sub-period: 1945-2021				
Event Window	[-1; 1]	[-1; 1]	[-1; 1]	[-1; 1]
Coefficient	0.1657	0.0070	0.0977	-0.0656
Std. Errors	(0.1428)	(0.0734)	(0.1430)	(0.0744)
Event Window	[0; 1]	[0; 1]	[0; 1]	[0; 1]
Coefficient	0.2619	-0.0272	0.1985	-0.0930
Std. Errors	(0.1748)	(0.0897)	(0.1749)	(0.0906)
Event Window	[-10; 10]	[-10; 10]	[-10; 10]	[-10; 10]
Coefficient	0.0987*	-0.0319	0.0798	-0.0608*
Std. Errors	(0.0544)	(0.0286)	(0.0553)	(0.0309)

*Coefficient results are presented in percent form, standard errors reported in parentheses, and statistical significance at the 10% level denoted with *.*

Table 7: T-test results for the difference between long winter vs early spring coefficients (no control regression coefficients) for sub-period 1945-2021.

T-test: $\beta_{LongWinter} - \beta_{Early\ Spring} = 0$			Sub-period 1945-2021
Event Window	[-1; 1]	[0; 1]	[-10; 10]
T-stat	0.990	1.470	2.093

As opposed to the base paper that suggests that Groundhog day predictions drive significant market returns, we cannot defend that finding with this revised analysis, but there is still some information conveyed through these results. There are trends in the data specific to the prediction type that are important to point out. All long winter prediction event windows point to moderately negative returns (or extremely low, barely positive returns), while the early spring prediction event windows point to moderately positive returns, which coincide exactly with the paper this analysis is based on. These trends point to the possibility that a superstitious investor would pay attention to the prognosis and react accordingly within the market. For example, even if none of the results are statistically significant, the early spring predictions are closer to being significant under both event windows, meaning that the stronger spring effect may reveal some sort of information to the market that drives investors to update their portfolios.

As previously mentioned, the difference in coding method may be a potential reason why this analysis was unable to achieve the same level of significance as the base paper. The use of a symmetric -1/+1 code within the same variable forces the early spring predictions to have the same effect as a long winter prediction. However it's incorrect to assume this relationship when early spring prognoses have significantly less observations than long winter, even if their economic magnitude is larger. So when calculating the buy-and-hold returns, the standard error that goes with it is not fully capturing how large it should be, and therefore showing significance. Although the t-tests run here show a very low level of significance (at the 10% level), it nonetheless confirms the difference between early spring and long winter predictions, which provides further proof that the original dataset should not have been coded in the way it was.

Even with mainly statistically insignificant results, the main results point to a variety of hypotheses and points of expansion for this analysis. The largest being what if investors are being influenced by some sort of weather effect before Phil makes his prediction. It could explain why there is a clear positive/negative average relationship between the potential prognosis because investors are attempting to anticipate what Phil will predict depending on the weather in the days leading up to the event. Shanaev, et al. (2021) imply that this may be a possibility given their graphs depicting the S&P 500's movement in the [-10; 10] event window, however they do not explicitly test this hypothesis. Furthermore, a more compelling way to test a potential weather effect is to see how the returns over a longer period respond if and when Phil is "correct" or not. This following analysis is now moving past just revisiting the original analysis, and adding new tests to improve the focus of the analysis.

4.2. Weather Effect Results

Using only the hindsight "correct" prediction years, as outlined in Table A.2., for both long winter and early spring predictions paired with the same main event windows and dummy variable regression models, the results for early spring and long winter predictions are summarized in Table 8.

Table 8: Average daily returns for correct predictions via dummy variable regressions 1-4 around Groundhog day predictions within different event windows for time period 1928-2021.

Regression Correct Prediction	(1) Early Spring	(2) Long Winter	(3) Early Spring	(4) Long Winter
Event Window	[-1; 1]	[-1; 1]	[-1; 1]	[-1; 1]
Coefficient	0.3018	0.0156	0.2552	-0.0331
Std. Errors	(0.2088)	(0.1071)	(0.2089)	(0.1078)
Event Window	[0; 1]	[0; 1]	[0; 1]	[0; 1]
Coefficient	0.5041**	-0.0316	0.4643*	-0.7580
Std. Errors	(0.2556)	(0.1310)	(0.2556)	(0.1316)
Event Window	[-10; 10]	[-10; 10]	[-10; 10]	[-10; 10]
Coefficient	0.0608	-0.0466	0.0426	-0.0698
Std. Errors	(0.0793)	(0.0411)	(0.0799)	(0.0428)

*Coefficient results are presented in percent form, standard errors reported in parentheses, and statistical significance at the 10% and 5% level denoted with * and ** respectively.*

Table 9: T-test results for the difference between long winter vs early spring coefficients (no control regression coefficients).

T-test: $\beta_{LongWinter} - \beta_{Early\ Spring} = 0$			
Event Window	[-1; 1]	[0; 1]	[-10; 10]
T-stat	1.221	1.865	1.187

The perfect forecast model yields statistically significant results only under the [0; 1] event window. Regression 1 shows significant average returns at the 5% level of .5041%. Regression 3 (full controls) continue showing significant results at the 10% level for the early spring prediction, yielding a positive average market return of 0.4643%. Table 9 t-statistics further confirm that the early spring prediction returns are particularly better for the event window of [0; 1].

Once again, we do see the same trends as the main regression results where early spring years tend to bring moderately positive returns compared to long winter years. Although we cannot say with absolute certainty that the years that investors anticipated would be correct actually drive any results, it's still an important test to run.

The pre-event period trends shown in the Shanaev, et al. (2021) study pointed to the possibility that weather effects may be a driving factor behind the rise in return. In other words, for the years that were warmer/colder leading up to the event, and then the subsequent matching early spring/long winter prediction we would hope to see stronger movements in the market. That however is not always the case; the event windows of [-1; 1] and [0; 1] do have stronger positive/negative changes for early spring/long winter predictions, with one exception, but that same trend does not hold for the [-10; 10] event window. This suggests that in the years that Phil is "correct" there is a stronger return to the market, suggesting that investors are somewhat paying attention to the weather and Phil's prediction in the very short term around the event date. However, due to the lack of significance across the board, it's difficult to make any definitive statements from the results.

This specific test could however be improved upon to align the research question more closely to the actual test. The trouble here is that weather data for the contiguous 48 states is only available in monthly values. If there were to be average daily data for February or March, the test may be improved by instead testing to see if the pre-event period drove any results in anticipation to Phil's prediction depending on the weather then. This would provide a more focused view of what investors are seeing or anticipating and subsequently how they react, if at all. Additionally, the actual conditions to determine a "correct" prediction may be modified to include some March data (to fit the "6 weeks" after prediction timeline) or in general can be improved upon considering that the current conditions necessary to determine "correctness" are not available. Including these changes and improvements could create a particularly strong and convincing test to combine the effect of potential Groundhog Day returns with the potential behavioral changes investors exhibit according to the weather.

4.3. Steel Industry Results

Here we test the hypothesis that steel industry returns might show a stronger response to Phil's prediction since the steel industry is central to Pennsylvania and should have many steel investors. While using the steel industry returns instead of the S&P 500, the same event windows and dummy variable regression models 1-4, the results for early spring and long winter predictions are summarized in Table 10.

Table 10: Average daily returns for Fama French Steel Industry via dummy variable regressions 1-4 around Groundhog day predictions within different event windows for time period 1928-2021.

Regression	(1)	(2)	(3)	(4)
Prediction	Early Spring	Long Winter	Early Spring	Long Winter
Event Window	[-1; 1]	[-1; 1]	[-1; 1]	[-1; 1]
Coefficient	0.0687	-0.0805	0.0572	-0.0996
Std. Errors	(0.3353)	(0.1602)	(0.3357)	(0.1623)
Event Window	[0; 1]	[0; 1]	[0; 1]	[0; 1]
Coefficient	0.2746	-0.1051	0.2778	-0.1039
Std. Errors	(0.4105)	(0.1959)	(0.4105)	(0.1976)
Event Window	[-10; 10]	[-10; 10]	[-10; 10]	[-10; 10]
Coefficient	0.0147	-0.0087	0.0106	-0.0129
Std. Errors	(0.1275)	(0.0623)	(0.1292)	(0.0670)

Coefficient results are presented in percent form and standard errors reported in parentheses.

As an attempt to revise and add-on to the statistical tests run in Shanaev, et al. (2021), a focus on the steel industry is presented because it is an important industry in Pennsylvania and may exhibit changes in results depending on Punxsutawney Phil's predictions. However, even this test fails to provide any statistically significant results for our furry friend. Although the same aforementioned positive/negative trends in early spring/long winter prognoses are once again evident, it fails to provide results strong enough to argue that Groundhog Day predictions move the market.

Table 11: Average daily returns via dummy variable regressions 1-4 around Groundhog day predictions within different event windows for time sub-period 1928-1943.

Regression	(1)	(2)	(3)	(4)
Prediction	Early Spring	Long Winter	Early Spring	Long Winter
Sub-period: 1928-1943				
Event Window	[-1; 1]	[-1; 1]	[-1; 1]	[-1; 1]
Coefficient	1.3810	-0.5805	1.2256	-0.6150
Std. Errors	(1.9955)	(0.5555)	(1.9910)	(0.5629)
Event Window	[0; 1]	[0; 1]	[0; 1]	[0; 1]
Coefficient	-0.0798	-0.4028	-0.2294	-0.4054
Std. Errors	(2.4438)	(0.6795)	(2.4374)	(0.6849)
Event Window	[-10; 10]	[-10; 10]	[-10; 10]	[-10; 10]
Coefficient	0.0719	-0.0455	0.0370	-0.0825
Std. Errors	(0.7557)	(0.2154)	(0.7562)	(0.2289)

*Coefficient results are presented in percent form, standard errors reported in parentheses, and statistical significance at the 10% level denoted with *.*

Table 12: Average daily returns via dummy variable regressions 1-4 around Groundhog day predictions within different event windows for time sub-period 1945-2021.

Regression	(1)	(2)	(3)	(4)
Prediction	Early Spring	Long Winter	Early Spring	Long Winter
Sub-period: 1945-2021				
Event Window	[-1; 1]	[-1; 1]	[-1; 1]	[-1; 1]
Coefficient	-0.0135	0.0347	-0.0228	0.0197
Std. Errors	(0.2986)	(0.1534)	(0.2991)	(0.1555)
Event Window	[0; 1]	[0; 1]	[0; 1]	[0; 1]
Coefficient	0.2968	-0.0351	0.3037	-0.0318
Std. Errors	(0.3655)	(0.1876)	(0.3658)	(0.1893)
Event Window	[-10; 10]	[-10; 10]	[-10; 10]	[-10; 10]
Coefficient	0.0111	-0.0019	0.0097	-0.0012
Std. Errors	(0.1137)	(0.0597)	(0.1156)	(0.0644)

*Coefficient results are presented in percent form, standard errors reported in parentheses, and statistical significance at the 10% level denoted with *.*

A secondary, sub-period regression set is re-run to account for the fact that Pennsylvania's one strong steel industry has moved westward overtime to the Midwest. Therefore, the sub-periods of 1928-1943 and 1945-2021 were chosen because they represent the shift in the steel industry. The ideal results would be that the 1928-1943 sub-period had strong results in favor of either an early spring or long winter prediction as proof of how Phil's prediction could affect the steel's industry's outcome/outlook. However, this period split-up yielded dramatically different results depending on the event window and a deviation from the positive/negative trends that were present in previous regressions. These results are driven by the fact that the sub-periods made the scarce early spring predictions even scarcer. The standard errors for some of the event windows is further proof of the sample problem since there is only one instance of an early spring prediction between 1928-1943. Unfortunately no hypothesis can be defended by the results of either sub-period or complete time period.

Further analysis is appropriate in this scenario because it may be the case that even though Pennsylavainians love and adore Phil, the weather simply does not affect the steel industry. It's difficult to determine how much the weather, or the predicted weather would apply to the industry since the Fama French industry definition includes all parts of steel manufacturing. Another industry of interest may be agriculture, but once again it's difficult to remove or know exactly what type of agricultural activities are important to this study. Furthermore, analyzing the effect on specific companies may provide other results. For example, companies that are headquartered around Punxsutawney or simply located within Pennsylvania would be prime targets for further analysis.

Table 13: Sample of publicly traded companies headquartered within Pennsylvania.

Company Name	Ticker	Industry	Headquarter City	Market Cap.
Comcast	CMCSA	Telecom	Philadelphia	$212.7 B
AmerisourceBergen	ABC	Pharmaceuticals	Conshohocken	$34.4 B
Aramark	ARMK	Restaurants/Food Services	Philadelphia	$9.99 B
PNC Financial Services	PNC	Regional Banks	Pittsburgh	$74.6 B
Rite Aid	RAD	Drug Retail	Camp Hill	$444.0 M
American Eagle Outfitters	AEO	Apparel Retail	Pittsburgh	$2.99 B
Dick's Sporting Goods	DKS	Specialty Stores	Coraopolis	$8.6 B
Kraft Heinz Company	KHC	Packaged Food & Meats	Pittsburgh	$53.1 B
Hershey Company	HSY	Packaged Food & Meats	Hershey	$47.1 B
UGI Corporation	UGI	Gas Utilities	King of Prussia	$7.9 B

For example, Table 13 demonstrates a small list of top names of public firms with headquarters within Pennsylvania. The breadth and depth of different industries and companies within the state could spir an analysis centering on these specific companies or a more formal "event-study" comparing returns of a company/portfolio to the overall market.

5. Conclusion

The motivation for this study was to provide evidence for the possible calendar anomaly in the U.S. associated with Punxsutawney Phil predictions during the period of Groundhog Day. This paper revised a dummy-variable regression framework in conjunction with an event study-like method from a previous paper, created a weather trend related regression to test Phil's accuracy, and provided an steel-industry focused analysis to test the strength of a Pennsylvanian key industry.

In the main regression results from 1928-2021, the S&P 500 results did not present any statistically significant results that were able to support the previous findings that there was an abnormal return anomaly for Groundhog Day. Instead, it did support the finding that the original methodology from the base paper was flawed due to incorrect coding of the indicator early spring/long winter binary variable. However, the results did present some possibility of the weather effect, which fueled subsequent statistical tests that improved upon the statistical analysis of the original base paper. Unfortunately, the weather effect test also did not provide any fruitful results, though it did reveal potential improvements in the base conditions/assumptions necessary for a test of that nature. Finally, the steel-industry returns became the central focus to see if Groundhog Day traditions affected a key Pennsylvaian industry, but this also did not provide any strong results. Even when attempting to split the time period to reflect the change in the structure of the industry, it provided even less promising results.

Although all tests failed to provide statistically significant results, it does not mean that the research question is insignificant itself. There is still a peculiar trend that was evident within all tests from providing positive average returns for early spring predictions, and negative average returns for long winter predictions. Additionally, there are many improvements that can be done to the analysis in order to further shape this new area of literature. If anything, the drastically different and opposite results presented here warrant further investigation because there is evidence of contention. Despite the fact that the original data set from Shanaev, et al. (2021) was used, this paper presented opposite results to the strong statistical significance found in the original base paper.

Not only does the process for determining significance need to be streamlined, but there are a variety of additional tests of interest that can be performed. Section 4 results provided some initial ideas to improve each test. Additionally, for future analysis it would be interesting to test if a particular year's press or media coverage surrounding Phil causes higher/lower returns in the market. For example, some years Google doodle sometimes highlights Phil on their main page, which involuntarily forces people to remember the holiday and see what Phil predicted. The larger amount of coverage may or may not drive market returns.

In summary, although the Weather Predictor Extraordinaire was unable to drive results in the stock market, he was able to drive happiness into our hearts. This framework can also be applied to other potential calendar anomalies, or any "event" that could potentially have the power to move the market. But for now, I hope the shaman of shadows, the prognosticator of all prognosticators, Springer of the Spring, The Prophet of PA a very happy extra few centuries of predictions and health (Karl and Doss, 2017).

6. References

Barker, Aaron. "How accurate is Punxsutawney Phil really?" *Fox Weather*, 1 Feb. 2022, www.foxweather.com/learn/how-accurate-is-punxsutawney-phil-really.

Bryant, Kelly. "10 Famous Groundhogs Besides Punxsutawney Phil." *Reader's Digest*, 14 July 2021, www.rd.com/article/famous-groundhogs-besides-punxsutawney-phil/.

Edwards, Phil. "Punxsutawney Phil: incompetent — or evil?" . , VOX News, 2 Feb. 2017, www.vox.com/2015/2/2/7963267/punxsutawney-phil-accuracy-evil.

Freed, Alan. *Punxsutawney Phil*. 2022, Reuters. www.pbs.org/newshour/nation/dont-stow-your-snow-boots-yet-punxsutawney-phil-saw-his-shadow-predicting-six-more-weeks-of-winter.

Karl, Andy, and Barrett Doss, performer. "Day One." *Original Broadway Cast of Groundhog Day*, produced by Chris Nightingale, Michael Croiter, and Tim Minchin, Avatar Studios, 2017. https://genius.com/Original-broadway-cast-of-groundhog-day-day-one-annotated.

Klopfenstein, Ashley. "Student, Faculty research shows Groundhog Effect on stock prices." *News Center*, edited by Tom Perry, Marietta University, 9 Jan. 2017, www.marietta.edu/article/student-faculty-research-shows-groundhog-effect-stock-prices.

Lombardi, Linda. "Punxsutawney Phil and 7 other animals who predict the future." *Today News*, TODAY.com, 1 Feb. 2012, www.today.com/news/punxsutawney-phil-7-other-animals-who-predict-future-wbna46223522.

MacKinlay, A. Craig. "Event Studies in Economics and Finance." *Journal of Economic Literature*, vol. 35, no. 1, American Economic Association, 1997, pp. 13–39, http://www.jstor.org/stable/2729691.

Melore, Chris. "Groundhog Day: Punxsutawney Phil saw his shadow — and 3 in 5 Americans trust his prediction!" *Oklahoma News 4*, Nexstar Media Inc., 2 Feb. 2022, https://kfor.com/news/u-s-world/groundhog-day-punxsutawney-phil-saw-his-shadow-and-3-in-5-americans-trust-his-prediction/.

National Oceanic and Atmospheric Administration (NOAA). *Groundhog Day - Punxsutawney Phil*. 2022, National Oceanic and Atmospheric Administration. www.silive.com/news/2022/02/groundhog-day-2022-prediction-how-accurate-is-punxsutawney-phil.html.

NOAA National Centers for Environmental information, Climate at a Glance: National Time Series (February & March, 1928 - 2022), published March 2022, retrieved on March 15, 2022 from https://www.ncdc.noaa.gov/cag/

Punxsutawney Groundhog Club. "Groundhog Club's Inner Circle." *The Punxsutawney Groundhog Club* , 2022, www.groundhog.org/inner-circle.

Punxsutawney Groundhog Club. "History & Past Predictions." *The Punxsutawney Groundhog Club* , 2022, www.groundhog.org/history-past-predictions.

Punxsutawney Groundhog Club. "Legend & Lore." *The Punxsutawney Groundhog Club* , 2022, www.groundhog.org/legend-and-lore.

Rowling, J.K. Harry Potter and the Sorcerer's Stone. New York: Scholastic, 1999. Print.

Shanaev, Savva, et al. "The Groundhog Day Stock Market Anomaly." *Finance Research Letters*, 22 Dec. 2021, doi:https://doi.org/10.1016/j.frl.2021.102641.

Stormfax Weather Almanac. "Groundhog Day." *Stormfax Weather Almanac*, 2022, www.stormfax.com/ghogday.htm.

The Associated Press. "Groundhog Day prediction is six more weeks of winter then 'beautiful' spring." *The Guardian*, 2 Feb. 2021, www.theguardian.com/us-news/2021/feb/02/groundhog-day-prediction-winter-punxsutawney-phil.

WHECTV. "Punxsutawney Phil: Bundle up!" *News10 NBC*, The Associated Press, 2 Feb. 2022, www.whec.com/news/punxsutawney-phil-bundle-up/6377537/.

Wheeler, Andrew P. "Testing the equality of two regression coefficients." Academic Blog Posts. , WordPress.com, 19 Oct. 2016, https://andrewpwheeler.com/2016/10/19/testing-the-equality-of-two-regression-coefficients/.

7. Appendix

Table A.1.: Punxsutawney Phil predictions (1887-2022).

Year	Prediction	Year	Prediction	Year	Prediction	Year	Prediction
1887	Long Winter	1921	Long Winter	1955	Long Winter	1989	Long Winter
1888	Long Winter	1922	Long Winter	1956	Long Winter	1990	Early Spring
1889	No Record	1923	Long Winter	1957	Long Winter	1991	Long Winter
1890	Early Spring	1924	Long Winter	1958	Long Winter	1992	Long Winter
1891	No Record	1925	Long Winter	1959	Long Winter	1993	Long Winter
1892	No Record	1926	Long Winter	1960	Long Winter	1994	Long Winter
1893	No Record	1927	Long Winter	1961	Long Winter	1995	Early Spring
1894	No Record	1928	Long Winter	1962	Long Winter	1996	Long Winter
1895	No Record	1929	Long Winter	1963	Long Winter	1997	Early Spring
1896	No Record	1930	Long Winter	1964	Long Winter	1998	Long Winter
1897	No Record	1931	Long Winter	1965	Long Winter	1999	Early Spring
1898	Long Winter	1932	Long Winter	1966	Long Winter	2000	Long Winter
1899	No Record	1933	Long Winter	1967	Long Winter	2001	Long Winter
1900	Long Winter	1934	Early Spring	1968	Long Winter	2002	Long Winter
1901	Long Winter	1935	Long Winter	1969	Long Winter	2003	Long Winter
1902	Early Spring	1936	Long Winter	1970	Early Spring	2004	Long Winter
1903	Long Winter	1937	Long Winter	1971	Long Winter	2005	Long Winter
1904	Long Winter	1938	Long Winter	1972	Long Winter	2006	Long Winter
1905	Long Winter	1939	Long Winter	1973	Long Winter	2007	Early Spring
1906	Long Winter	1940	Long Winter	1974	Long Winter	2008	Long Winter
1907	Long Winter	1941	Long Winter	1975	Early Spring	2009	Long Winter
1908	Long Winter	1942	No Record	1976	Long Winter	2010	Long Winter
1909	Long Winter	1943	No Record	1977	Long Winter	2011	Early Spring
1910	Long Winter	1944	Long Winter	1978	Long Winter	2012	Long Winter
1911	Long Winter	1945	Long Winter	1979	Long Winter	2013	Early Spring
1912	Long Winter	1946	Long Winter	1980	Long Winter	2014	Long Winter
1913	Long Winter	1947	Long Winter	1981	Long Winter	2015	Long Winter
1914	Long Winter	1948	Long Winter	1982	Long Winter	2016	Early Spring
1915	Long Winter	1949	Long Winter	1983	Early Spring	2017	Long Winter
1916	Long Winter	1950	Early Spring	1984	Long Winter	2018	Long Winter
1917	Long Winter	1951	Long Winter	1985	Long Winter	2019	Early Spring
1918	Long Winter	1952	Long Winter	1986	Early Spring	2020	Early Spring
1919	Long Winter	1953	Long Winter	1987	Long Winter	2021	Long Winter
1920	Long Winter	1954	Long Winter	1988	Early Spring	2022	Long Winter

Punxsutawney Groundhog Club (2022), Stormfax Weather Almanac (2022)

Table A.2.: Punxsutawney Phil prediction accuracy (1928-2021) based on revised conditions for "correctness."

Year	Correct?	Year	Correct?	Year	Correct?
1928	No	1960	Yes	1991	No
1929	No	1961	No	1992	No
1930	No	1962	No	1993	No
1931	No	1963	No	1994	No
1932	No	1964	Yes	1995	Yes
1933	Yes	1965	Yes	1996	No
1934	Yes	1966	No	1997	Yes
1935	No	1967	No	1998	No
1936	No	1968	Yes	1999	No
1937	Yes	1969	Yes	2000	No
1938	No	1970	No	2001	Yes
1939	No	1971	Yes	2002	No
1940	Yes	1972	No	2003	No
1941	Yes	1973	Yes	2004	No
1942	Yes	1974	No	2005	Yes
1943	No	1975	Yes	2006	No
1944	No	1976	No	2007	No
1945	No	1977	No	2008	No
1946	No	1978	Yes	2009	Yes
1947	Yes	1979	No	2010	No
1948	No	1980	Yes	2011	Yes
1949	Yes	1981	Yes	2012	No
1950	No	1982	Yes	2013	No
1951	No	1983	No	2014	No
1952	No	1984	Yes	2015	No
1953	Yes	1985	No	2016	No
1954	No	1986	Yes	2017	Yes
1955	Yes	1987	No	2018	Yes
1956	Yes	1988	Yes	2019	Yes
1957	No	1989	No	2020	Yes
1958	Yes	1990	No	2021	No
1959	Yes				

Figure A.1.: Collage of Punxsutawney Phil's magnificent aura and presence.

Sources from left to right: Melore (2022), Edwards (2017), The Associated Press (2021), WHECTV (2022)