Analysis of S&P Instruments Their Price Pinning Tendencies

Christopher Lam
Portfolio Website
Github Repo Link

Overview of Study

Objective

The objective of this study is to analyze differing instruments of the S&P 500 (SPY, ^XSP, and ^SPX) to observe and compare their sensitivity to price pinning behavior. Specifically, this study isolates the day of the week to observe if an instrument has a statistically significant tendency to showcase price pinning behavior when compared to theoretical random chance and relative to another.

Background

While all intended to represent the S&P 500 index, the mathematical calculations that are utilized in price calculation in SPY, ^XSP, and ^SPX differ (e.g. dividend calculations and expense fees) . Further, the overall intention of each instrument to investors and traders also differ (e.g. ^XSP is intended for options trading). One byproduct of these differences is the ability for a price pin to occur. A **price pin** is where an underlying asset settles at or near a specified strike price at the time of an option contract's expiration. This natural phenomena often occurs due to market makers hedging their positions in order to remain delta neutral (unbiased towards the general movements of the underlying market).

This phenomena is often observed during or around important days within the year (e.g. FOMC and major economic news days, major holidays, OPEX and quad-witching days, earning reports, etc.). However, due to recent advancements in options trading, the amount of data and available opportunities for price pinning has increased. Thus, this study aims to analyze these distributional price biases, specifically for S&P 500 assets.

Dataset Information & Pipeline

Data Pipeline Overview

The data for this study is divided into three distinct parts: Raw Data, Cleaned Data, and Analyzed Data. **Raw Data** refers to data pulled via Python through the extension yfinance. **Cleaned Data** refers to the raw data after it has been processed, removing unnecessary features and adding new core features required for analysis. **Analyzed Data** refers to a dataset built from the cleaned data that aims to derive key insights revolving around price pins. This pipeline is utilized for each of the underlying assets observed (SPY, ^XSP, ^SPX).

The data for each asset is under a daily interval, and spans from 2023-03-01 to 2026-01-09. This allows the data to encompass both bull and bear markets, while maintaining a sense of recency with the data. Further, ^XSP data exists only up to 2023-03-01, so keeping the timeframe standardized ensures fair comparison of the data.

Data Storage & Architecture

Throughout the data pipeline, information is stored onto a PostGreSQL 18 server. Specifically, each distinct part of the overall data (raw, cleaned, and analyzed) is stored in order to keep a stable record of all data. However, actual analysis and filtration was conducted primarily through Python via the Pandas and NumPy extensions. In essence, the PostGreSQL server acts as a persistent read only storage for the data in this study. As such, a corresponding upload & download Python file is utilized in order to push and pull each distinct part of the overall data.

Raw Data

This data is sourced from the yfinance extension in Python on the applicable underlying assets (SPY, ^XSP, ^SPX). Here, the columns stored are as follows:

- **Date**
  - YYYY-mm-dd HH:MM:SS
- **Open**
  - Official opening price of the ticker
- **High**
  - Official daily high of the ticker
- **Low**
  - Official daily low of the ticker
- **Close**
  - Official closing price of the ticker
- **Volume**
  - Volume traded during the day(if calculable)

Although not every column is applicable in further calculations, they are intentionally left in the schema due to their importance in general price analysis, as well as for potential extension.

Each file is saved under the format {**ticker**}_daily_price_raw, where **ticker** is the applicable underlying asset.

Cleaned Data

This data is the cleaned version of the raw data, applicable for each of the underlying assets. In addition to the columns found within the clean data, additional columns are included:

- **Weekday**
  - Trading day of the week ["Monday", "Tuesday", … , "Friday"]
- **True_rounded_close**
  - Closing price according to standard rounding, 2 decimals
- **True_dollar_rounded_close**
  - Closing price according to standard rounding, nearest dollar
- **True_point_5_rounded_close**

  - Closing price according to standard rounding, to the nearest 0.5
- **True_point_25_rounded_close**
  - Closing price according to standard rounding, to the nearest 0.25
- **Floor_rounded_close**
  - Floored closing price, to the nearest dollar
- **Ceil_rounded_close**
  - Ceilinged closing price, to the nearest dollar
- **Is_dollar_rounded_true_multiple_of_5**
  - Binary variable [0, 1] to see if variable **True_dollar_rounded_close** is an exact multiple of 5
- **Is_dollar_rounded_true_multiple_of_10**
  - Binary variable [0, 1] to see if variable **True_dollar_rounded_close** is an exact multiple of 10
- **Is_point_5_rounded_true_multiple_of_1**
  - Binary variable [0, 1] to see if variable **True_point_5_rounded_close** is an exact multiple of 1
- **Is_point_25_rounded_true_multiple_of_1**
  - Binary variable [0, 1] to see if variable **True_point_25_rounded_close** is an exact multiple of 1
- **Is_dollar_rounded_tolerance_1_multiple_of_5**
  - Binary variable [0, 1] to see if variable **True_dollar_rounded_close** is a multiple of 5, with a tolerance of 1 (i.e. **True_dollar_rounded_close** rounded value ends in 0, 1, 4, 5, 6, or 9)
- **Is_dollar_rounded_tolerance_1_multiple_of_10**
  - Binary variable [0, 1] to see if variable **True_dollar_rounded_close** is a multiple of 10, with a tolerance of 1 (i.e. **True_dollar_rounded_close** rounded value ends in 0, 1, or 9)

Each file is saved under the format {**ticker**}_daily_price_clean, where **ticker** is the applicable underlying asset.

Analyzed Data

This data is an analyzed version of the cleaned data using statistical analysis. For each applicable asset, a **p-value threshold** and **Wilson confidence interval (CI) threshold** is utilized for statistical analysis. The (p-value, wilson) threshold combinations are from the subset [(0.05, 0.05), (0.1, 0.1), (0.2, 0.1), (0.2, 0.2)]. For each (p-value, wilson) threshold combination for each of the applicable assets, the columns are as follows:

- **Index**
  - Index of the entry
- **Day**
  - Trading day of the week ["Monday", "Tuesday", … , "Friday", "absolute"]
    - "Absolute" refers to the feature analyzed regardless of day value
- **Feature**

- Feature in question being observed; Belongs to the subset of the *cleaned data's* Is_* columns. (i.e. a column for **Is_dollar_rounded_true_multiple_of_5** is represented here).
    - Of the subset ["dol_round_true_mult_5", "dol_round_true_mult_10", "point_5_round_true_mult_1", "point_25_round_true_mult_1", "dol_round_tol_1_mult_5", "dol_round_tol_1_mult_10"]
- **P_hat**
    - Point estimate for observed sample data
    - In the range [0,1]
- **P0**
    - Null hypothesis; Built under the assumption of random chance
        - E.g. "dol_round_true_mult_5" holds a p0 of 0.20, in accordance with the random chance probability of the event occurring. Below is the full p0 information for applicable features:

          EXPECTED_P = {
        - **"dol_round_true_mult_5": 0.20,**
        - **"dol_round_true_mult_10": 0.10,**
        - **"point_5_round_true_mult_1": 0.50,**
        - **"point_25_round_true_mult_1": 0.25,**
        - **"dol_round_tol_1_mult_5": 0.60,**
        - **"dol_round_tol_1_mult_10": 0.30,**
        - **}**

    - In the range [0,1]
- **Delta**
    - Difference in the observed point estimate (p_hat) and the assumed random chance (p0)
    - In the range [0,1]
- **N_obs**
    - Number of observations taken into consideration
- **Z**
    - Z-score
- **P_value**
    - Measured probability value of the feature
    - In the range [0,1]
- **Ci_low**
    - Lower bound for the Wilson confidence interval (CI)
- **Ci_high**
    - Upper bound for the Wilson confidence interval (CI)
- **Is_significant**
    - Boolean variable (False, True) for if p_value is significant enough to reject the null hypothesis
- **Ticker**

- Ticker symbol analysis is conducted under
- **Pval_threshold**
  - Threshold for which p-value significance is tested to
  - Of the subset [0.05, 0.1, 0.2]
- **Wilson_threshold**
  - Threshold alpha for Wilson confidence interval (CI)
  - Of the subset [0.05, 0.1, 0.2]

Each file is saved under the format {**ticker**}_analyzed_data_pval_{**pval_threshold**}_wilson{**Wilson_threshdol**}, where:

- **ticker** is the applicable underlying asset.
- **Pval_threshold** is the p-value threshold used, using "_" in place of decimal points (i.e. 0.05 -> 0_05)
- **Wilson_threshold** is the Wilson alpha threshold used, using "_" in place of decimal points (i.e. 0.05 -> 0_05)

Statistical Analysis of Data

As described, several p-value - wilson alpha CI combinations were utilized; This section will describe the rationale of the design choices, as well as the overall process in calculating.

Wilson CI vs Wald CI

Due to the nature of the study (identifying and comparing price pining phenomena across different S&P 500 vehicles) as well as the structure to which this is measured (binary variables indicating if closing price was featured), a Wilson CI was chosen over the traditional Wald CI.

This design decision was made explicitly, due to a Wilson CI being superior to Wald CI calculations for binary variables, which is what the features studied are measured in. Further, this study spliced the data into days of the week and analyzed them independently in addition to in an absolute sense. The result of this is a sample size of roughly 300 — a moderate amount given the context which a Wilson CI benefits from.

P-value - Wilson CI combinations

In total, there are 4 combinations that this study focused on when it came to (**p-value**, **Wilson CI alpha**):

- (0.05, 0.05)
- (0.1, 0.1)
- (0.2, 0.1)
- (0.2, 0.2)

Each of these thresholds serve a different purpose when it comes to addressing the overall objective of price pinning.

The p-value Wilson threshold combination (0.05, 0.05) serves as a strict, core classification of the data. For any row whose variable **Is_significant** is equal to true, the following are satisfied:

- Distinguishable from the null (random chance) at 5%
- Effects large enough to bypass sampling noise
- Stable enough to survive conservative CIs

Ultimately, rows that pass this threshold combination are robust enough for action. However, this threshold does miss weaker, yet persistent, signals, as well as potential regime-controlled behaviors. As such, failure to pass this threshold does not immediately constitute absence of any actionable effect.

The p-value Wilson threshold combination (0.1, 0.1) serves as a looser, more suggestive classification of the data. For any row whose variable **Is_significant** is equal to true, the magnitude of effect may not be as significant, but still worth exploring. Ultimately, rows that pass this threshold combination are again robust enough for action, though the overall effect will be noticeably weaker than that of a (0.05, 0.05) combination

The p-value Wilson threshold combination (0.2, 0.1) serves as a screening classification of the data, identifying instances that may require further analysis in order to justify action. For any row whose variable **Is_significant** is equal to true, the magnitude of effect will likely not be as significant, but is likely stable enough to justify further data analysis in the future. Ultimately, rows that pass this threshold combination are good candidates for a second layer of analysis.

The p-value Wilson threshold combination (0.2, 0.2) serves as an exploratory map.. For any row whose variable **Is_significant** is equal to true, the magnitude of effect is unlikely to be significant, but is able to be visualized to show discrepancy of some capacity. Ultimately, rows that pass this threshold combination compare relative magnitudes and consistency, but not areas where significance can be found.

Visualization of Data

This section describes the three major types of charts utilized in this study, as well as the core question each aims to answer/visualize

Point-Estimate vs Null w/ Wilson CI

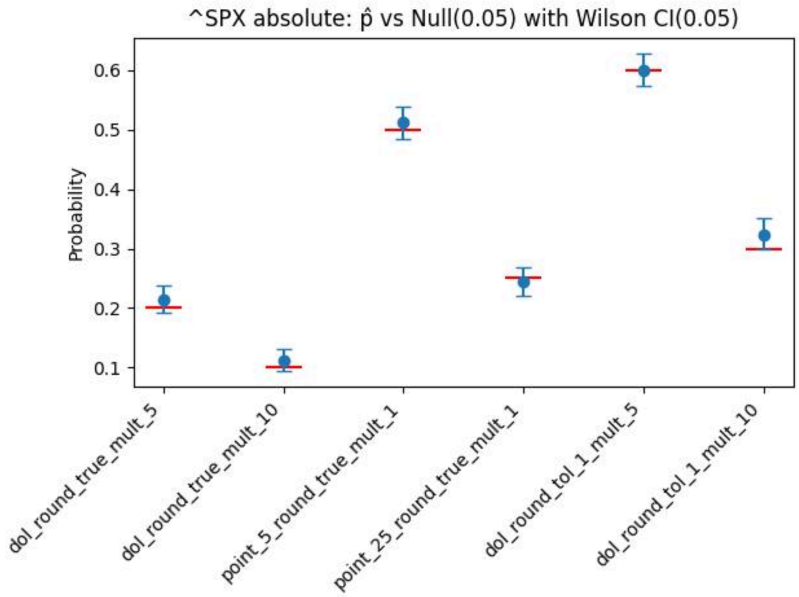

Fig 1A:^SPX absolute: p̂ vs Null(0.05) with Wilson CI(0.05).jpg

This type of chart aims to answer if for a given day and feature, the observed probability is more significant than that of the null hypothesis (random probability). For each feature on a given day:

- The observed probability of the feature p_hat is represented through a dot
- The feature's Wilson confidence intervals are represented via the vertical error bars
- The null hypothesis is represented through the red horizontal bar

Here, the objective of each chart is to determine if there is enough evidence to statistically reject the null hypothesis of price pinning being a random distribution. That is, a feature whose observed interval and confidence interval are both outside of the null hypothesis showcase a statistically meaningful deviation in price pinning effects. For instances where the observed probability and its respective confidence intervals are above the null hypothesis, price pinning is shown to have an upward bias. On the contrary, instances where the observed probability and its respective confidence intervals are below the null hypothesis, price pinning is shown to have a downward bias. Lastly, the overall distance from the null hypothesis indicates the strength of the effect.

See **Fig 1A** as an example point estimate plot, configured for a p_value threshold of 0.05 and Wilson alpha threshold of 0.05 on the ^SPX index. For each underlying asset (SPY, ^XSP, ^SPX), there are 5 groups (1 for each p_value threshold & Wilson alpha combination), with 6 plots per group (one for each day, and one absolute chart).

Delta Heatmap

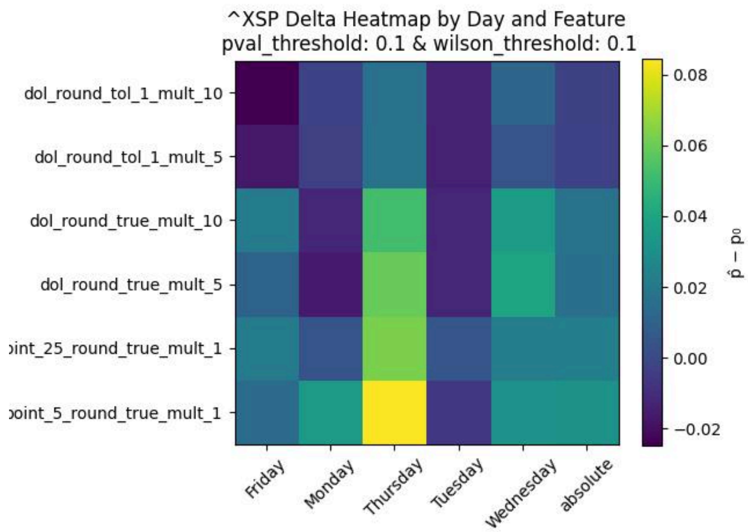

Fig 1B: ^XSP Delta Heatmap by Day and Feature pval_threshold: 0.1 & wilson_threshold: 0.1.jpg

This type of chart aims to showcase the distribution of deviations across days and features, be it random or something statistically significant and worth exploring.This heatmap uses a color gradient scale to represent the difference (delta) of the observed probability against the null hypothesis. Color values that are positive represent instances where the observed is above the null, while a negative color value represents instances where the observed is below the null. This map spans across each day of the week observed, as well as the absolute dataset, against each feature to form the matrix to which the heatmap is applied to.

Here, the objective of the delta heatmap is to locate and identify consistency across days and/or features. As such, the heatmap is exploratory in nature, and is not meant to identify significance through color alone. For instance, consistent colors across rows or columns may indicate structural biases worth exploring, while single color spots may be a factor of noise.

See **Fig 1B** as an example delta heatmap, configured for a p_value threshold of 0.1 and Wilson alpha threshold of 0.1 on the ^XSP index. For each underlying asset (SPY, ^XSP, ^SPX), there are 5 groups (1 for each p_value threshold & Wilson alpha combination), with 1 plot per group.

Effects Size vs Statistical Significance

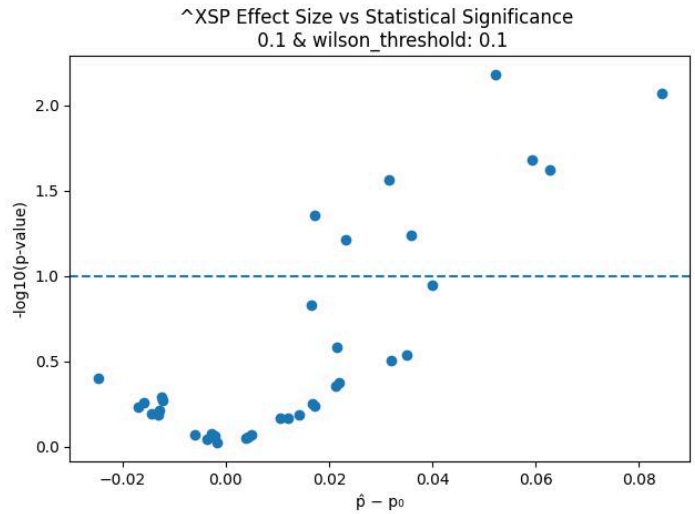

Fig 1C: ^XSP Effect Size vs Statistical Significance 0.1 & wilson_threshold: 0.1.jpg

This type of chart aims to showcase which results are both statistically credible and non-trivial in size. The plot maps the delta between the observed effect size against the null on the X-axis against the statistical logarithmic strength of the p-value on the Y-axis. Additionally, the dashed line represents the statistical significance level, as noted with the p-value threshold on a logarithmic scale.

Here, the objective is to locate points that are on the upper right or left of the plot, as they are both significant in effect and in magnitude. Points in these areas are generally high-confidence results and contain a bias that may be exploited. On the contrary, points below the dashed threshold yet to the edges lack the magnitude that indicate significance in the effect. Thus, points near the edges of the X-axis yet below the dashed threshold may observe large but weak effects. As for points that exist near x=0, the effect, while significant, may lack actionable economic relevance. Ultimately, this plot aims to identify deviations that are significant *and* large enough to explore.

See **Fig 1C** as an example effects plot, configured for a p_value threshold of 0.1 and Wilson alpha threshold of 0.1 on the ^XSP index. For each underlying asset (SPY, ^XSP, ^SPX), there are 5 groups (1 for each p_value threshold & Wilson alpha combination), with 1 plot per group.

Results, Figures, & Observations

To see all the generated charts, check the [GitHub link] and proceed to assets->*{ticker}* -> pval *{pval_threshold}* Wilson *{Wilson CI}* -> *{chart}*.

Point-Estimate vs Null w/ Wilson CI Observations

Applicable Figures:

- ^SPX {**weekday**} … p̂ vs Null with Wilson CI
  - Applicable across all thresholds & weekdays, including absolute
- ^XSP {**weekday**} … p̂ vs Null with Wilson CI
  - Applicable across all thresholds & weekdays, including absolute

Observations & Figs

For the majority of weekday-feature combinations, the observed p-values with their Wilson CIs overlapped the null hypothesis. This means that the data generally implies a failure to statistically reject the null hypothesis that a given day has a superior price pinning effect.

However, there are instances where the observed p-value and their Wilson CIs existed outside of the null hypothesis; These instances were particularly apparent across Thursday in all assets, though weakly at times on Fridays as well. See figures **Fig 2A/B/C** as examples where the observed p-value and its Wilson CI exists outside of the null (Note all images are that of the tightest thresholds, and thus are also apparent in the looser thresholds)

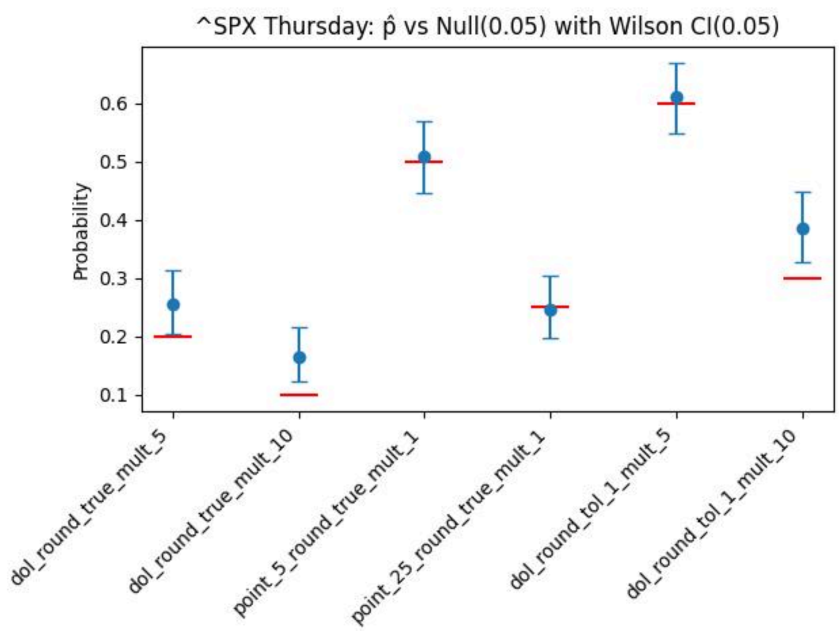

Fig 2A: ^SPX Thursday: p̂ vs Null(0.05) with Wilson CI(0.05).jpg

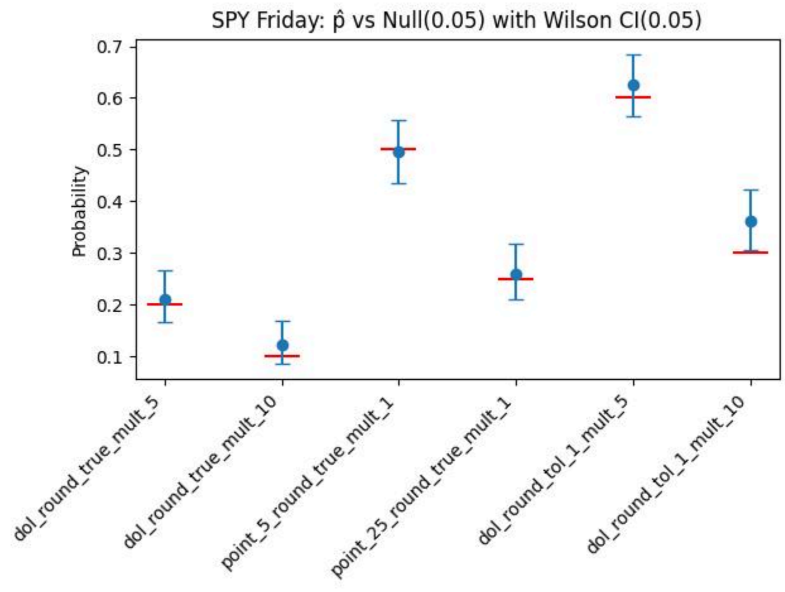

Fig 2B: SPY Friday: p̂ vs Null(0.05) with Wilson CI(0.05).jpg

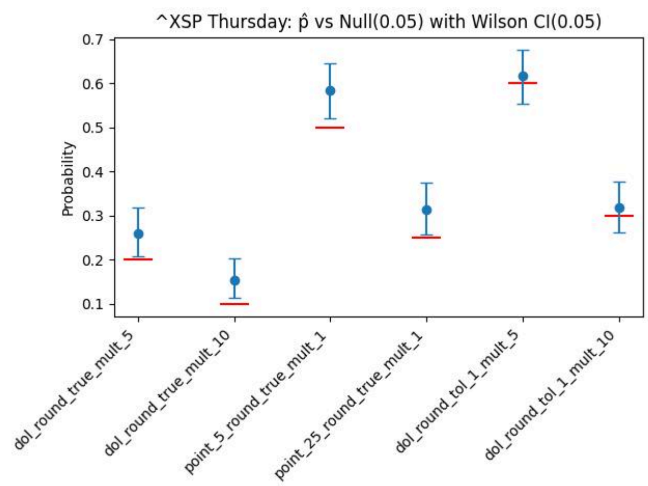

Fig 2C: ^XSP Thursday: p̂ vs Null(0.05) with Wilson CI(0.05).jpg

Another observable distinction is that the feature in question that is outside of the null varies at times depending on the asset, all else being the same. This is particularly apparent with **Fig 2A** and **Fig 2C**, where the only difference is the underlying asset. Despite tracking the same root index (S&P 500), the price pinning bias each observes is different with the former showcasing upwards bias in feature "dol_rounded_tol_1_mult_10" and the later in feature "point_5_round_true_mult_1". This is likely due to liquidity differences and in the usage of each underlying asset, as ^XSP was designed more favorably towards small retail options traders, as opposed to ^SPX which is more appropriate towards institutional options traders.

Delta Heatmap Observations

Applicable Figures:

- ^SPX Delta Heatmap by Day and Feature
    - Applicable across all thresholds observed
- ^XSP Delta Heatmap by Day and Feature
    - Applicable across all thresholds observed
- SPY Delta Heatmap by Day and Feature
    - Applicable across all thresholds observed

Observations & Figs

Between both ^SPX and ^XSP (**Fig 3A** and **Fig 3B** respectively), both delta heatmaps displayed consistently higher positive deltas on Thursday, though the feature in question varied between the two. Further, both ^SPX and ^XSP showcased a small degree of abnormal deltas on Monday, Wednesday, and Friday, though the magnitude of the difference was relatively mute.

Across SPY's columns, features showed a relatively smooth transition from day-to-day, which is a stark contrast to both ^XSP and ^SPX who have a differentiating band on Thursday. This suggests that weekdays may not be a primary condition for price pinning in SPY; That is, features dominate the weekdays, and that SPY is less prone to pinning effects. Some speculative reasoning as to this are:

- Pre-market and after-market trading extends trading times, spreading out large orders and neutralizing price pinning effects
- SPY (unlike ^SPX and ^XSP) does not have 1256 tax benefits, which classify all gains at 60% long-term and 40% short-term regardless of time held . This makes SPY a weaker candidate for options trading, which is strong mechanism that encourages likelihood of the price pinning phenomena

That being said, SPY's heatmap (**Fig 3C**) is noticeably different from **Fig 3A/B**, which represent ^SPX and ^XSP. Specifically, SPY's heatmap has a bias on Fridays and Mondays. However, this is likely not related to a weekday specific price pinning effect, as justified previously through SPY's smooth day-to-day transition and the point-estimate charts discussed earlier. Rather, the reason for this may be:

- Weaker option flows
- Quad witching, Monthly OPEX, and other Friday driven events that create SPY pinning
- Market psychology on covering short positions across weekends, market rebalancing, and options unwinding

As such, SPY appears to showcase a higher than random observed pinning effect for Fridays, though not that attributed to the day of the week.

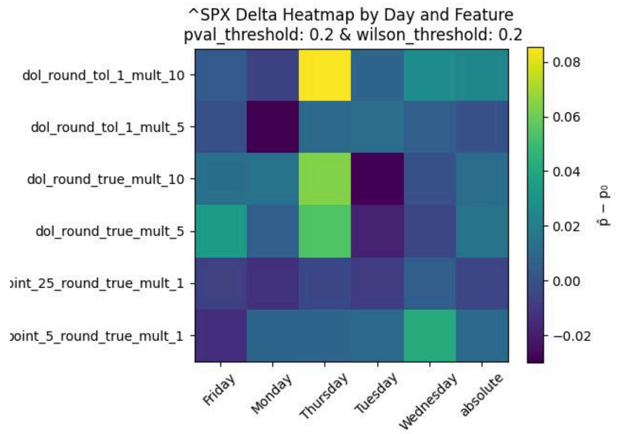

Fig 3A: ^SPX Delta Heatmap by Day and Feature pval_threshold: 0.2 & wilson_threshold: 0.2.jpg

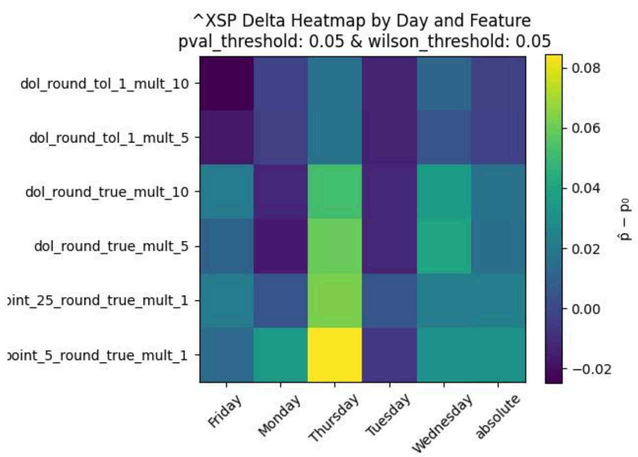

Fig 3B: ^XSP Delta Heatmap by Day and Feature pval_threshold: 0.05 & wilson_threshold: 0.05.jpg

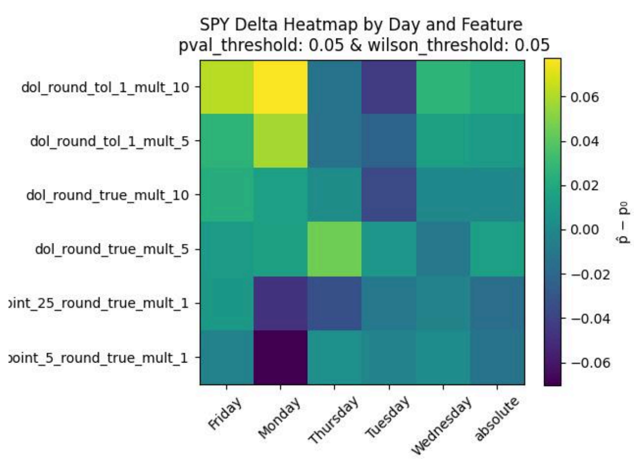

Fig 3C: SPY Delta Heatmap by Day and Feature pval_threshold: 0.05 & wilson_threshold: 0.05.jpg

Effects Size vs Statistical Significance Observations

Applicable Figures:

- ^SPX Effect Size vs Statistical Significance
    - Applicable across all thresholds observed
- ^XSP Effect Size vs Statistical Significance
    - Applicable across all thresholds observed

Observations & Figs

The majority of features for both ^SPX and ^XSP in the effects scatter plots across all thresholds hold a delta with an absolute value near 0 ($|\hat{p} - p_0| \approx 0$–$0.02$) and are below the logarithmic threshold indicated by the dashed line (See examples **Fig 4A/B**). As such, the majority of features lack both the magnitude and significance to display statistically significant deviations from the null.

A second observation is with the overall distribution, with plots for ^SPX and ^XSP both having asymmetry towards positive delta, particularly where p-val & Wilson are 0.2 in ^SPX (See **Fig 4C**); That is, for price pinning features that are statistically significant from the null, they are often biased upward (more often) for ^SPX and ^XSP. Rationally, this makes sense, as both of these assets are favored by option traders who primarily contribute to the price pinning phenomena.

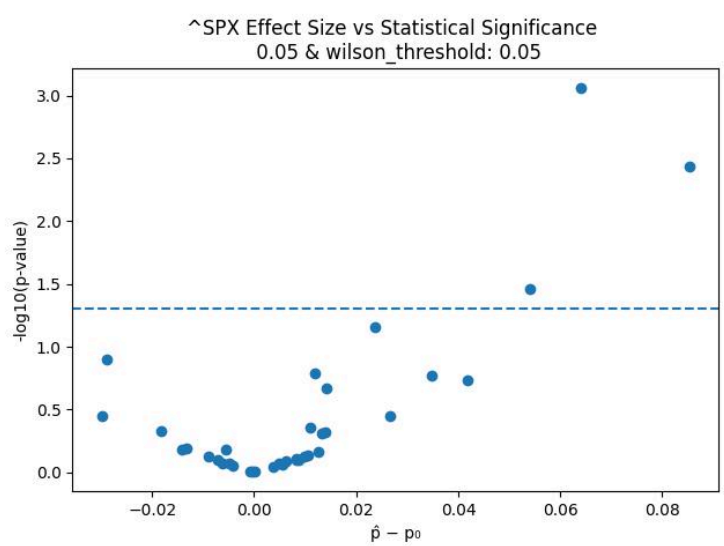

Fig 4A: ^SPX Effect Size vs Statistical Significance 0.05 & wilson_threshold: 0.05.jpg

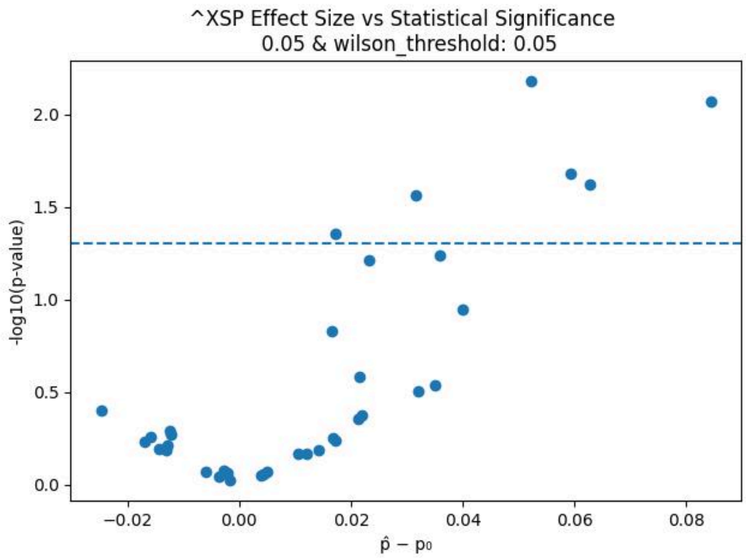

Fig 4B: ^SPX Effect Size vs Statistical Significance 0.05 & wilson_threshold: 0.05.jpg

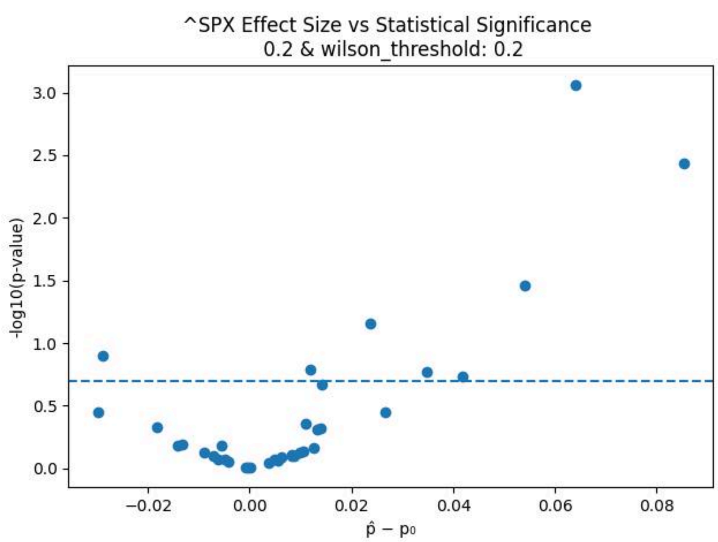

Fig 4B: ^SPX Effect Size vs Statistical Significance 0.2 & wilson_threshold: 0.2.jpg

A third and final observation is that variations in threshold decrease the number of surviving points, but not necessarily in the structure/shape of the scatter plot. This is best observed between **Fig 4A** vs **Fig 4C**, where the difference is in the p-value and Wilson thresholds (0.05 & 0.05 vs 0.2 & 0.2 respectively). This suggests that price pinning is not a product of noise, but instead a natural or mechanical phenomena that is driven by market structure and dynamics. That being said, the intention and significance of price pinning likely varies depending on market regimes; The day of the week likely have, at best, a weak effect on price pinning.

Bottom Line Observations

The data as a whole supports the idea of a weak yet persistent price pinning effect within the S&P 500 assets studied. However, the day of the week does not appear to have any consistent, statistically significant impact on this price pinning phenomena. Rather, contributions to price pinning are likely driven from a combination of several factors, such as dealer hedging and liquidity concentrations.

Critiques

Lack of Intraday Data

The data of this study primarily revolved around daily price data around the S&P 500, and did not investigate the more granular intraday level. However, price pinning is inherently an intraday phenomena and is non-linear near options settlement. This is due to the nature of options pricing according to the Black Scholes model, which displays higher gamma exposure near option contract expiration. In other words, price pinning tends to happen towards the end of a given trading day on an option's expiration date. Thus, analyzing across the granularity of an entire trading day may wash out statistically significant effects.

Limitations in Causal Identification

The conclusions of this study are observational and correlation in nature. While price pinning is supported by the data, the question of what contributes to price pinning and by how much is yet to be answered. A multitude of potential market structures and processes may have an effect on price pinning, including but not limited to:

- Dealer gamma hedging
- Liquidity clustering near round numbers (psychological "safety" zones)
- Algorithmic trading executed based on external metrics and calculations affecting benchmark prices

As such, there are unclear conclusions as to what contributes to price pinning (despite its existence) and how significant those contributors may be.

Small Delta Limits Application

From the data, the observed delta for the observed features are small in absolute probability terms. This limits the number of applications that may be taken, as there is fragility in the bias. In regards to establishing an edge, any that exists from a price pinning effect from this study is subject to failure from regime changes, sampling errors, and model misspecifications. Further, real applications would have to consider tail risks, intraday variance, and other feasibility issues. Thus, price pinning as a whole fails to serve as a stand alone "trading edge", and is simply a measurable and persistent bias in probability

Survivorship and Structural Change Risk

Option markets are a dynamic environment that have been evolving in recent years. With expansion in 0dte volume, shifts in dealer positioning, and advancements in settlement conventions, past data may not accurately reflect the current environment. In other words, past data does not give any indication of future events, especially given the recent advancements in options trading.

Concluding Thoughts

Generally speaking, the phenomena of price pinning is a weak yet persistent effect showcased through S&P 500 data. This pattern is evident throughout a multitude of different asset types (SPY, ^SPX, ^XSP) as well as differing statistical thresholds. However, the effects of price pinning appear to be mechanical and emergent, as opposed to something intentional or manipulative. The day of the week that the asset is traded does not appear to have any overarching, statistically significant influences on price pinning effects. Despite this, weak influences are at times observable when it comes to the strength of a price pinning effect, not direction nor nature of the pin.

Due to the size and strength of the observed effect sizes, no conclusive trading edge can be derived nor interpreted from price pinning alone. As such, results are best utilized in an exploratory manner, not predictive.