

# A Methodology for Measuring Market Inefficiency, and Its Relationship to the Value Spread \*

Aman Rana

24th November, 2024

## Abstract

Extending the work of ?, we propose a method to measure the level of market inefficiency using unbiasedness regressions. We find that the market is not uniformly efficient and goes through periods of higher inefficiency reverting to efficiency. We find that the value spread—the ratio of the average book-to-market of expensive stocks to that of cheap stocks—is cointegrated with market inefficiency. By demonstrating that wider value spreads comove with our measure of market inefficiency, we can empirically support ?’s intuition that high spreads during events such as the dot-com bubble and COVID-19 could reflect mispricing rather than fundamental changes in the notion of value. This evidence strengthens the case that the value premium is still alive but has been masked by market inefficiency. This has implications for risk-premia based strategies and portfolio management.

---

\*Code and data supporting this analysis is available at: [https://github.com/Aman-Rana-02/Irrational\\_Value\\_Spreads](https://github.com/Aman-Rana-02/Irrational_Value_Spreads)

## Contents

# 1 Introduction

## 1.1 Market Efficiency

? defines markets to be informationally efficient if prices at every moment incorporate all available information about future values rationally and instantaneously. This is an intuitive result if we consider that market participants bid prices up once presented with information that an asset is underpriced, and sell when new information shows an asset as overpriced. When market participants are digesting information, they do so with consideration of a firm's current and future states. We would expect that due to competition, firms get more sophisticated in their forecast and procurement of information, and market gets relatively more efficient over time. Market efficiency is important because in an efficient market investors are compensated for taking systematic risk. The expected returns of risky assets must be higher on average because investors require a risk premium to bear the additional uncertainty associated with risky assets.

## 1.2 Value Risk

'Value' in factor investing is an example of one such systematic risk. Its classical definition is the book-to-market ratio; the book value of equity of a firm divided by its market capitalization. According to ?, firms with high book-to-market ratios (value stocks) tend to outperform those with low book-to-market ratios (growth stocks) over time, as this factor captures a systematic risk associated with the financial distress or other underlying characteristics of value firms. Value is a relative measure, there is no threshold that defines a stock as 'cheap' or 'expensive'.

## 1.3 Value Spread

? defines the value spread as the ratio of the average book-to-market of a portfolio of 'expensive' stocks to the average book-to-market of a portfolio of 'cheap' stocks. The value spread can be interpreted as how much more the market is willing to pay for already 'expensive' stocks than 'cheap' stocks. He uses this measure of value spread in a time series, showing that the spread has been increasing steadily for the last decade and proposes that this widening is due to market inefficiency. His claim is that the widening spread has no fundamental backing, the market is irrationally paying more for 'expensive' stocks, and therefore it takes longer for market participants to be appropriately rewarded for holding value risk. His justification lies in

his market experience, and that papers that try to justify that the relative definition of ‘value’ has fundamentally changed, are inconclusive.

## 1.4 Inefficient Value Spreads

In this paper I use a direct approach to investigate the efficiency of the value spread. I propose a measure for market informational inefficiency, and provide a time series of this measure. Asness justifies his conclusion by contradiction, he presents possible explanations for the value spread, and that they don’t work. I am investigating the same phenomena more directly, by constructing a measure of market efficiency and showing its movement with the value spread.

Existing methods for measuring market efficiency typically focus on testing whether stock prices follow a random walk (?) (?). However, these methods do not represent the variation of efficiency over time. A more dynamic approach is needed to investigate a time series of inefficiency.

? uses unbiasedness regressions to examine price informativeness around FOMC meetings. This method visually shows us how prices adjust to information flow, in speed, and under or overreaction, using  $R^2$ s and  $\beta$ s over the unbiasedness regression window. Unbiasedness regressions provide us with a tool to measure the market’s digestion of information, from which I can construct a measure of market efficiency.

Our estimand is market inefficiency, which we define as the relative level of the market’s error in its present value of future prices, when using available information. Building on this, we introduce the *Beta\_SSE* estimate, which measures market efficiency at a point in time. By applying unbiasedness regressions within a rolling window, and constructing the *Beta\_SSE* representation, we create a time series that tracks the relative level of market inefficiency.

Our findings show that:

- The market is not uniformly efficient. Instead, it experiences cycles of inefficiency followed by reversion to efficiency.
- Over the last decade, inefficiency has been increasing, suggesting that the market has become less adept at incorporating available information into prices on longer time-scales.
- This rising inefficiency is correlated with the widening value spread, indicating a link between the weakness of the value premium and market inefficiency.

The rest of the paper is structured as follows: Section ?? describes the data used in this analysis. Section ?? outlines the methodology used to measure market inefficiency. Section ?? presents the results of our analysis. Section ?? concludes and discusses the implications of our findings. Appendix ?? provides additional details on the methodology, simulations to show measure robustness, and illustrative examples of unbiasedness regression results to build intuition.

## 2 Data

We use Python (?) to analyze S&P 500 price data from Yahoo Finance (?) and the Fama-French 3-factor data library (?). Yahoo finance provides daily price data of the S&P 500 which we resample to monthly returns for the period 1950 - 2024. We get monthly log returns by taking the difference of the log of the adjusted close price of the S&P 500 for each month (Figure ??). Our time series starts in 1950 based on the premise that the market prior to this was not as sophisticated, and is not illustrative for the purposes of this analysis of market efficiency and value spreads. (?) justifies this using the instability of value spreads prior to 1950.

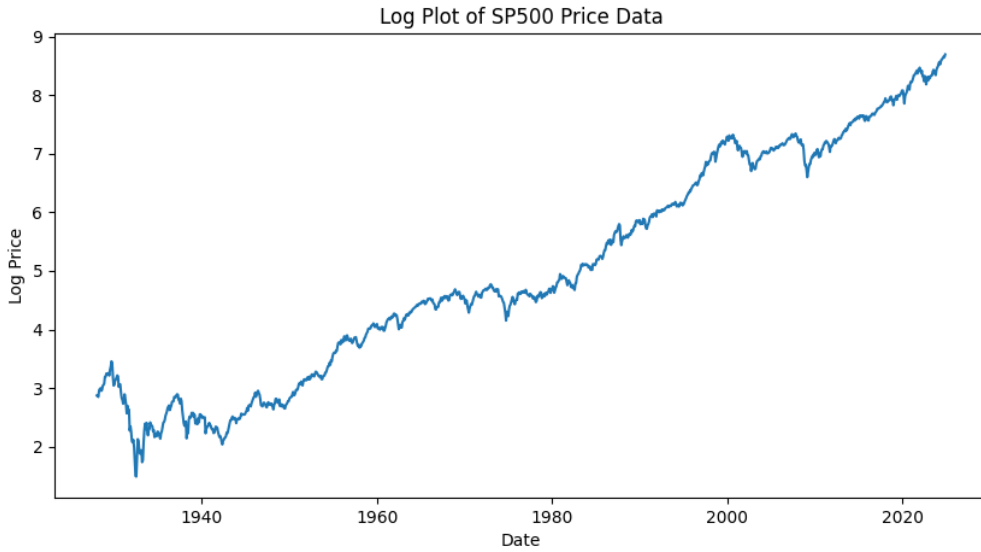


Figure 1: The S&P 500 time series 1950 - 2024.

### 2.1 Cumulative Log Returns

Date	Cum Log Return 1	Cum Log Return 2	Cum Log Return 3	...	Cum Log Return 34	Cum Log Return 35	Cum Log Return 36
1950-01-01	0.0154	0.0253	0.0293	...	0.379	0.424	0.459
1950-02-01	0.0100	0.0140	0.0520	...	0.409	0.444	0.436
1950-03-01	0.00412	0.0421	0.0867	...	0.434	0.427	0.408
1950-04-01	0.0380	0.0827	0.0229	...	0.422	0.404	0.380
1950-05-01	0.0446	-0.0151	-0.00672	...	0.366	0.342	0.315

Table 1: Cumulative log S&P500 return periods for each month. Each column represents the cumulative log return for a given period starting at the row index, and ending  $t$  months ahead.

The unbiasedness regression require windows of cumulative returns. For any month  $t$ , we get every forward month's returns up to  $T$  months. By taking expanding window sums of the columns of log returns, we get a cumulative log returns matrix (Table ??).

## 2.2 The Value Spread

The value spread is a ratio of the average book-to-market of the most expensive 30% portfolio to the average book-to-market of the portfolio of the cheapest 30% of stock portfolio, as per ?. To construct this measure we use Kenneth French's data library (?). Using French's 3x2 sort on book-to-market and market equity, we get the monthly market value weighted average of the book-to-value of the portfolio of the 30% most expensive large-cap stocks and the portfolio of the 30% cheapest large-cap stocks. Dividing the two averages represents how much more expensive the expensive stocks are compared to the cheap stocks (Figure ??), or how much more the market is willing to pay for growth versus value stocks.

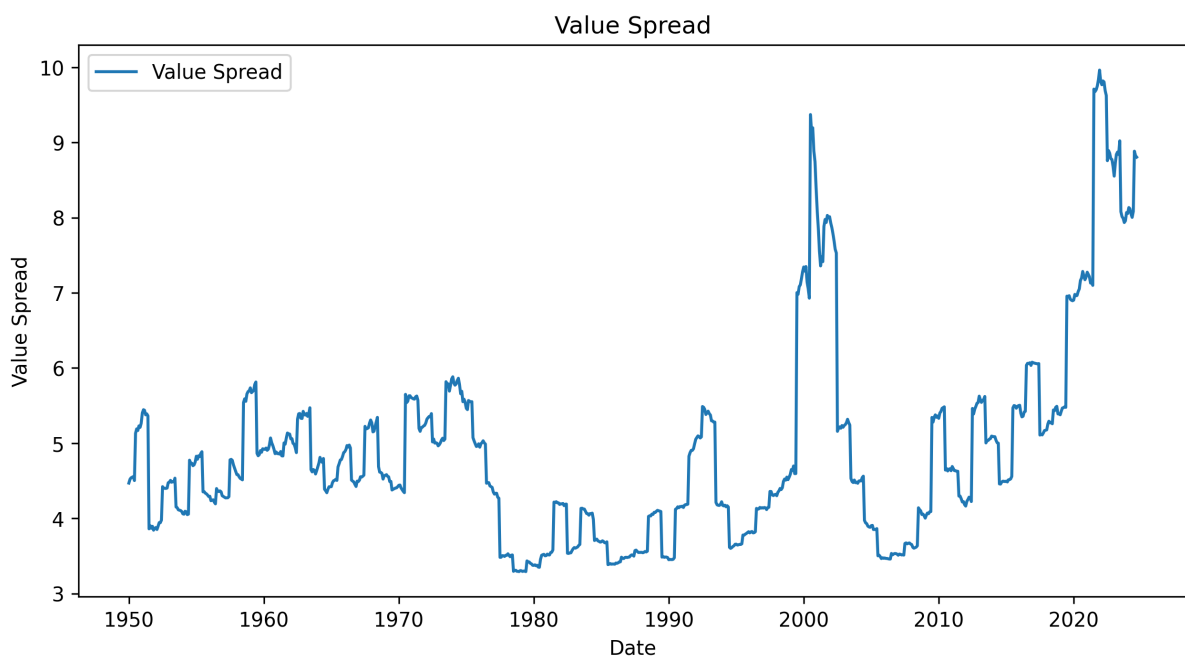


Figure 2: The value spread time series 1950 - 2024. ? argues the widening spread in the last decade is due to increasing market inefficiency.

## 2.3 Measurement

We use cumulative log returns of the S&P 500 because every cumulative return window represents the information content of that time period. Partial cumulative return windows demonstrate the information content up to a point in the entire period. Using the subsets of a cumulative window period together, we get an idea of the information flows within the wider cumulative period window. We use the S&P 500 because its returns are more representative of the average market information ability than the returns for a single firm, which would be subject to idiosyn-

cratic shocks. The S&P 500 also represents larger firms, which are heavily observed by market participants, and have well-regulated information flow, so should represent the most efficiently priced index. I expand on how these components representing information flow lead to a measure of market inefficiency in Section ??.

### 3 Modelling Market Efficiency

In this section, we will introduce the  $Beta\_SSE$  estimate, which is our proposed measure of market inefficiency based on the unbiasedness regressions framework.

#### 3.1 Unbiasedness Regressions

The unbiasedness regression is a simple OLS regression of the cumulative logarithmic return of the asset over a normalized window of time  $SP500_{[0,T]}$ , on a subset of the logarithmic cumulative returns,  $SP500_{[0,t]}$ , where  $t \leq T$ , and  $t$  is the number of months from the beginning of the window.

<sup>1</sup>

$$SP500_{[0,T]} = \alpha_t + \beta_t SP500_{[0,t]} + \epsilon_t, \text{ where } 0 \leq t \leq T. \quad (1)$$

For an illustrative example, let  $T = 36$ , so that the window  $[0, T]$  represents a 3-year period, and let us look at the returns of the SP500.

For  $t = 1$ , the regression is of the form:

$$SP500_{[0,36]} = \alpha + \beta SP500_{[0,1]} + \epsilon \quad (2)$$

From this regression we extract the coefficient  $\beta$  and the  $R^2$  value. The regression is repeated for each partial return window  $t = 2, t = 3, \dots, t = 36$ , and the coefficients  $\beta$  are plotted against  $t$ , the depth of the cumulative return window. We plot the  $\beta$  coefficients against  $t$  to get a sense of the market's ability to process information over time (Figure ??). The  $\alpha_t$  intercept accounts for any time-varying risk premia.

$\beta_t = 1$  for all  $t$  when log prices move efficiently, following a random walk. When  $\beta_t = 1$  the partial returns from 0 to  $t$  provide a forecast of the return from  $t$  to  $T$  that is as efficient

---

<sup>1</sup>SP500<sub>[0,1]</sub> is the matrix of returns of the SP500 from January 1950 to February 1950, February 1950 to March 1950, ..., August 2024 to September 2024, September 2024 to October 2024. SP500<sub>[0,2]</sub> would be the matrix of cumulative returns from January 1950 to March 1950, February 1950 to April 1950, ..., August 2024 to October 2024, and so on for the entire dataset.



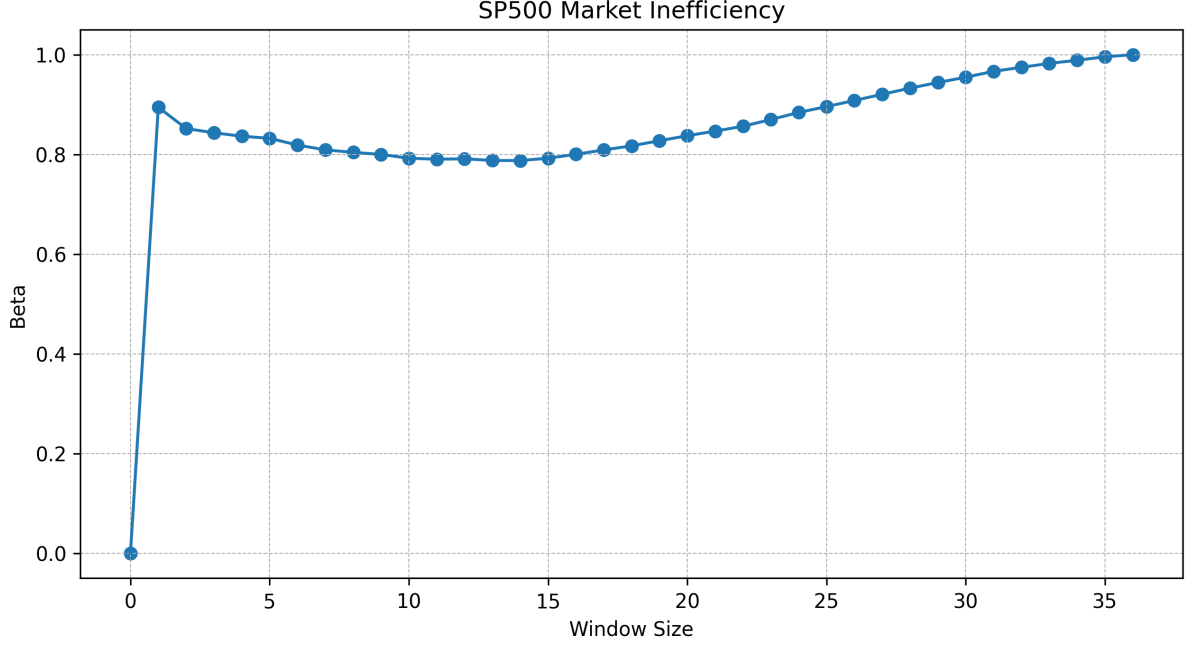


Figure 3: An unbiasedness regression on the SP500 (1950 - 2024) showing the market’s tendency to overreact.

as possible and does not need to be amplified or dampened (?). If  $\beta_t < 1$ , the partial return is dampened in the total returns, therefore there was a temporary component in prices which decays representing an “overreaction” (?). Conversely, if  $\beta_t > 1$ , the partial return is amplified in the total return, suggesting the market had improperly processed the information relevant in pricing the T sized window, and has underreacted.

In this paper we will be focusing on the  $\beta_t$  coefficient. We now have the foundation to construct a score for market efficiency.

### 3.2 *Beta\_SSE Score*

We have established the unbiasedness regression infrastructure, and that a  $\beta_t < 1$  indicates overreaction, a  $\beta_t > 1$  indicates underreaction, and  $\beta_t = 1$  indicates efficiency. To construct a score that measures the level of market inefficiency at a point in time, we have to create a spot representation of the  $\beta_t$  graph for a time period.

We define the *Beta\_SSE* score as the sum of the squared differences between  $\beta_t$  and the horizontal line at  $\beta = 1$ , for all  $t$  in the window (Equation ??). Since a  $\beta_t = 1$  indicates efficiency, the *Beta\_SSE* score will be lower when the market is more efficient and larger when the market is less efficient, whether by under- or over-reaction. We provide a visual representation of the *Beta\_SSE* score for the S&P 500 over our entire data window in Figure ??.

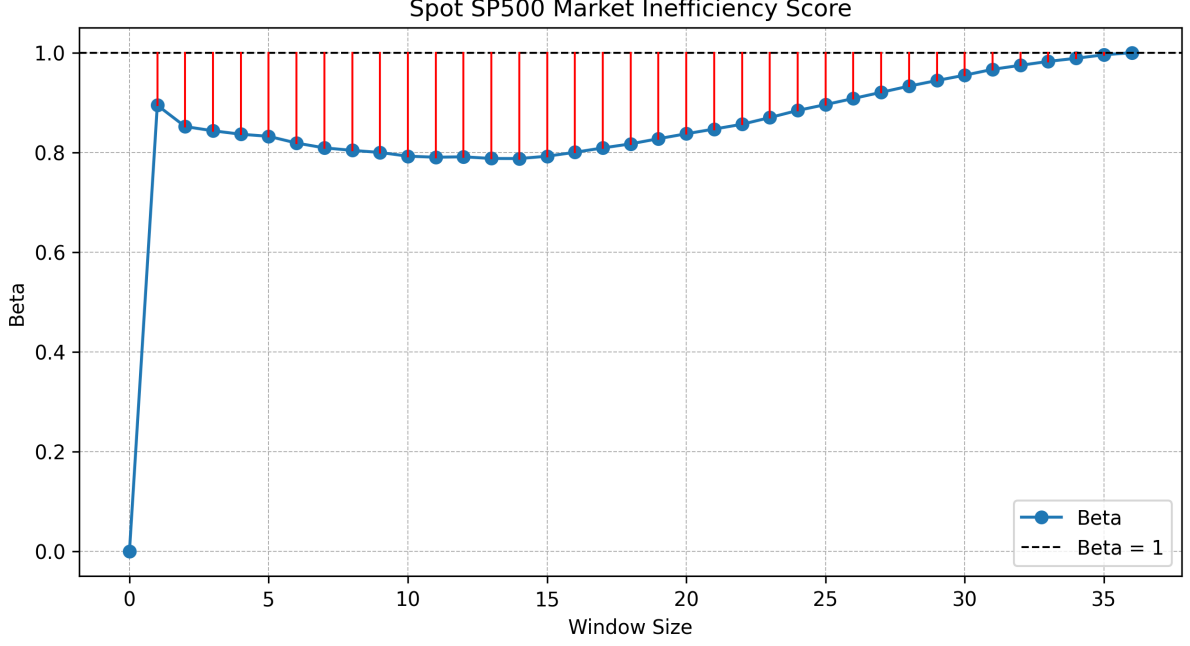


Figure 4: The  $Beta\_SSE$  score visualization for the SP500 (1950 - 2024). The  $Beta\_SSE$  score is a measure of market efficiency at a point in time, and is the sum of the squares of the red distances. To construct a timeseries of the  $Beta\_SSE$  score, we run this regression for shingled five-year windows and plot their scores.

To construct a time series of these scores, we split our entire period of S&P 500 returns in shingled 5-year windows. Each window is 5 years long, and the window is moved forward by 1 month at a time. We run the same sequence of unbiasedness regressions as we did for the entire period in Figure ??, however on each window independently, and calculate the  $Beta\_SSE$  score for each window. So  $Beta\_SSE_t$  is constructed from the unbiasedness regressions on windows of the SP500 returns from  $t - 60$  to  $t$ , where  $t$  represents a month-year. Note that our unbiasedness regressions extend out to  $t + 36$ , so  $Beta\_SSE_t$  score isn't tradable at time  $t$  but at time  $t + 36$ .

<sup>2</sup> This is by design, a measure of market inefficiency at a time  $t$  is to be representative of the market's ability to price the  $T$  months ahead.

$Beta\_SSE_t$  is calculated as follows:

$$Beta\_SSE_t = \sum_{i=0}^{36} (\beta_{t,i} - 1)^2 \quad (3)$$

Where  $\beta_{t,i}$  is the  $\beta$  coefficient from the unbiasedness regression from a window starting on the date  $t$ , at the  $i$ th depth of the unbiasedness window.

<sup>2</sup>Consider the rolling window [January 2000, January 2005], the observation starting at January 2005 ( $t = 0$  on January 2005) requires the window of returns from January 2005 to January 2008 to make  $SP500_{[0,36]}$ .

### 3.3 Justification of Model

Consider ?'s definition of informational market efficiency. The market is efficient if prices at every moment incorporate all available information about future values rationally and instantaneously. If we have a window of cumulative returns, and the market has efficiently incorporated information, and is sophisticated enough to appropriately discount future returns, we would expect cumulative returns to grow linearly as information comes in and aligns to the market expectation. This is the case presented when  $\beta_t = 1$ . If the market receives information during a window that does not align with expectations, we would expect the partial returns up to that point to be dampened or amplified in the total returns. Events that would cause this are shocks such as earnings surprise, or other information that undermines the information consumed by the market by that point. ?'s example of potential inefficiency, that prices typically rise or fall on the release of insider information, is picked up by this framework. It is an inefficiency, because information that was known by some participants, but not all, was not properly priced in. The market will adjust to this new information but, the price up to that point was not efficient. These are the cases when  $\beta_t \neq 1$ , which often spikes as these inefficiencies are shocks. Since we are summing the squared differences, we are penalizing the market for not being efficient by under- or over-reaction, and the *Beta\_SSE* score is an appropriate measure of relative market inefficiency. It does not differentiate between known possible shocks such as earnings seasons, and unprecedented shocks or black swan events, so we do see spikes of inefficiency. But, by definition of market efficiency we would expect it to revert to a low mean, and drift downwards over time as the market becomes increasingly better at pricing in information and less prone to surprise. An in-depth exploration of the *Beta\_SSE* score's properties and its relationship to market efficiency can be found in the Appendix ??, where we provide idealized results from simulations of efficient and inefficient markets.

## 4 Results

Using the *Beta\_SSE* score, we construct a time series of the level of market efficiency as seen in Figure ??.

We see the market is inefficient around the dot-com bubble, the 2008 financial crisis, and the COVID-19 pandemic, which is in line with our expectations. These are times the market had to contend with inefficient pricing due to uncertainty and information asymmetry. We

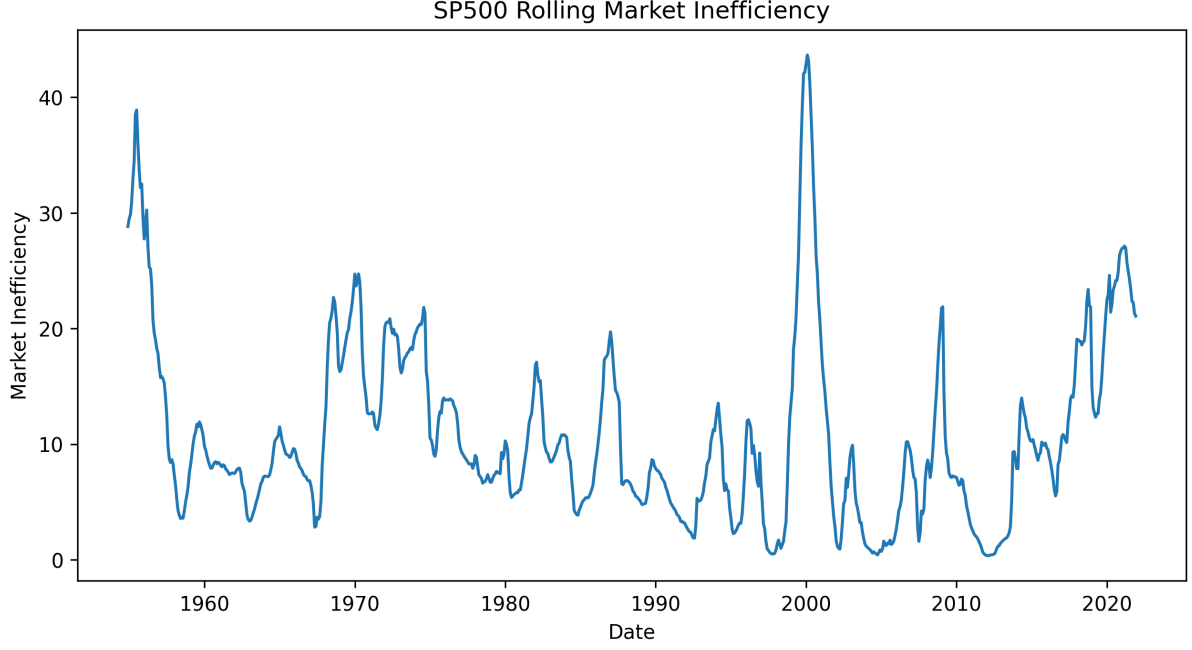


Figure 5: The  $Beta\_SSE$  score time series of the SP500 (1950 - 2024) showing mean reversion and periods of inefficiency.

can visually assess that the market is not uniformly efficient. Instead, it goes through periods of higher inefficiency followed by reversion to efficiency. These are the potential sources of inefficiency shocks (Earnings surprise, insider information becoming public, and information asymmetry) that ? discusses in his paper, and are expected. We also see that the market has been increasingly inefficient in the last decade, which is in line with ?. We expect inefficiency to drift downward with occasional shocks, and we see that in the timeseries from 1950 - 2014, but from 2014 - 2024 we see an upward drift.

Empirically we can test the mean reversion through an Augmented Dickey-Fuller test (?) on the  $Beta\_SSE$  time series, which is significant at the 0.1% level.

Figure ?? shows the value spread and the  $Beta\_SSE$  score time series. We see visually that the two have been correlated for the last decade, which we test with a Spearman rank correlation on the rolling mean's of both values which we find to be 0.94 for the period 2014 - 2024. We use the Spearman rank correlation to account for any non-linear relationship between the two variables, and take a 5-year rolling mean to smooth out temporary inefficiencies and capture the trend of the two variables.

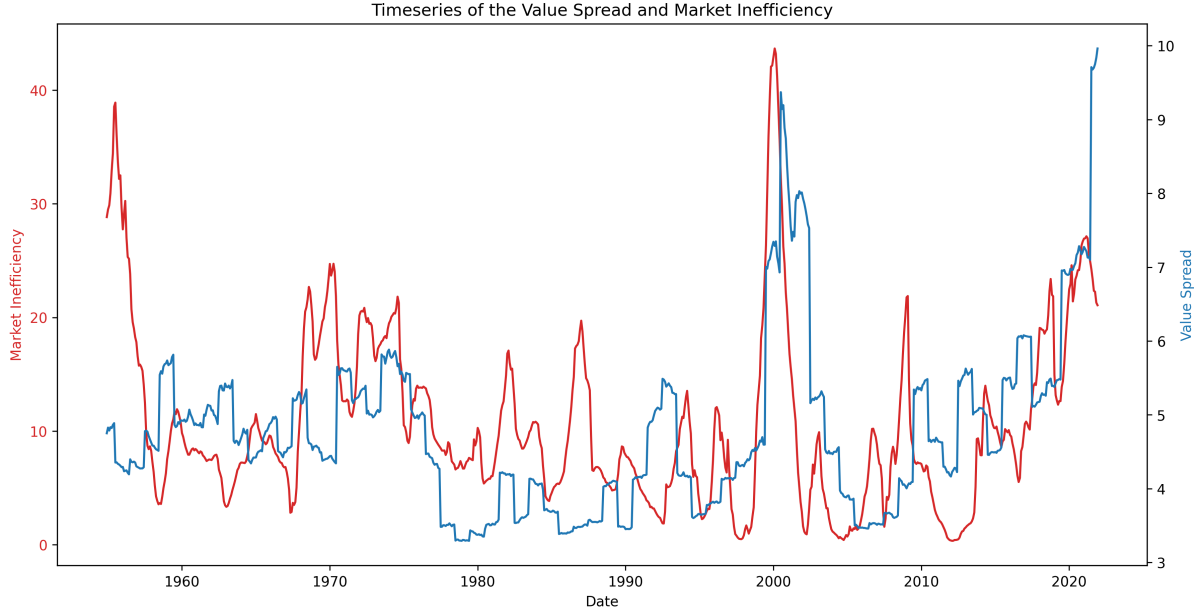


Figure 6: The value spread, and market inefficiency move together, and have been increasing in tandem for the last decade.

## 5 Discussion

### 5.1 Summary

In this paper we construct a measure of market efficiency using  $\beta$ 's definition of informational efficiency. Our measure is based on unbiasedness regressions, and captures where the market has inefficient expectations. We build on  $\beta$ 's case on inefficient value spreads by showing a high correlation of market inefficiency and value spreads in the last decade.

### 5.2 Drifting Inefficiency

We learn the market's ability to present value future prices has weakened over the last decade. The cause of which is worth investigating in future work. Either the market has been fundamentally more inefficient due to some microstructure, or the world could be in a regime of such innovation and technological change that the market has been systematically forced to revise its present values of future expectations with accelerating error. This may not be entirely far-fetched, the last decade has seen the rise of the internet, the proliferation of smartphones, and the rise of machine learning. These are all innovations that have changed the way we live and work, learn and build. It is obvious the rate of innovation is accelerating, but is the market correctly pricing the higher moment, the rate of innovation acceleration? Perhaps the markets haven't reached the level of sophistication to correctly price the current accelerating rate of innovation,

indicating that the rate of innovation is a systematic surprise and the cause of  $Beta_{SSE}$  drift.

### 5.3 Limitations and Future Work

A limitation of this paper is that we have not established causality between market inefficiency and value spreads. This is a topic for future work. They are clearly correlated in the last decade, but we cannot say that one causes the other. It is possible that growth stocks' high expectations are reasonable, value spreads are reflecting this new regime of expectations, and our co-moving inefficiency is because of the unprecedented (Therefore difficult to price) nature of these innovations. ? argues otherwise, showing possible explanations and dismissing them, but he caveats that this is not a complete proof. An implication of their co-movement is on portfolio management, if value spreads do widen during periods of inefficiency, portfolio managers with risk-premia mandates may choose to underweigh their value exposure during inefficient regimes, and lever back up once markets tend back to efficiency. Conversely, anomalies such as momentum should outperform during periods of informational inefficiency as the market comes to grips with unexpected information flows.

In future work we hope to investigate the decomposition of market inefficiency into under-reaction and over-reaction components. There is an existing literature on the market's tendency to overreact and the robustness of our methodology would be strengthened if it aligns with the current state of the literature. Additionally, since our measure has a forward-looking bias, attempting to forecast  $Beta_{SSE}$  would make it more practical for its use in portfolio and risk management.

# Appendix

## A Appendix

### A.1 Unbiasedness Regressions

In this Appendix we build on the intuition behind the unbiasedness regressions and the measurement of market inefficiency. We provide plots from simulated efficient and inefficient markets, and show how the unbiasedness regressions can be used to measure the degree of inefficiency in the market.

#### A.1.1 Simulated Efficient Market

In this section we simulate an efficient market, where returns are normally distributed with a mean of 0 and a standard deviation of 0.1. They are independent random events.

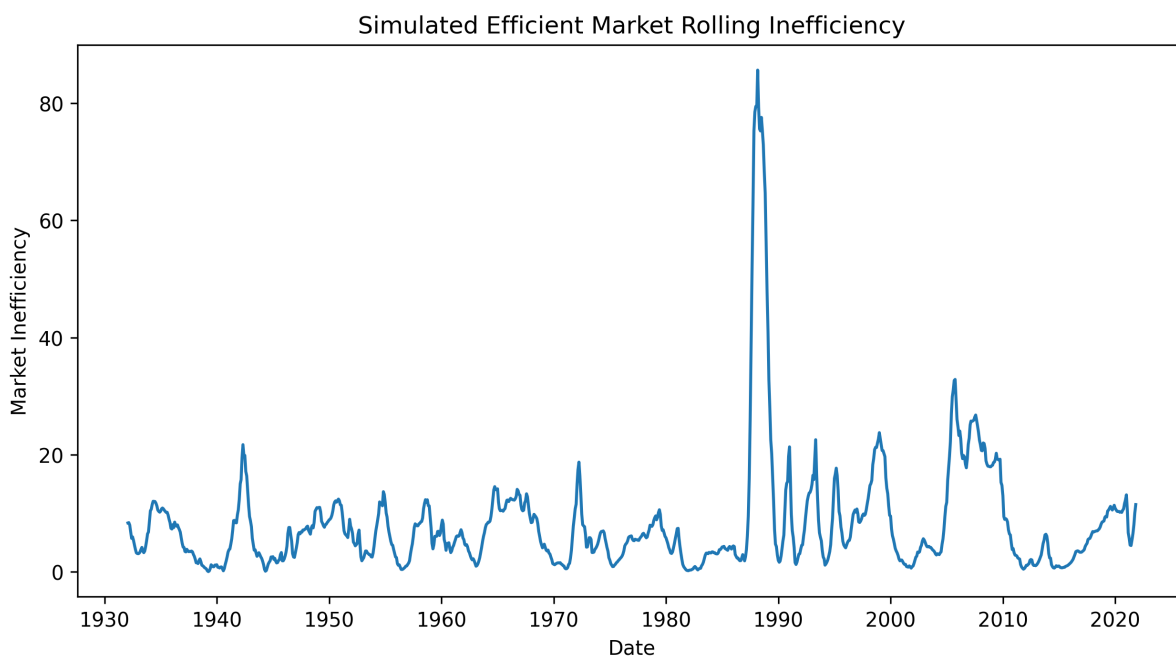


Figure 7: Simulated Efficient Market

In Figure ?? we see the rolling inefficiency of the simulated efficient market. Notice that the level remains low, with only small spikes in inefficiency.

### A.1.2 Simulated Inefficient Market

In this section we simulate an inefficient market, where returns are an AR(1) process:

$$r_t = \phi r_{t-1} + \epsilon_t \quad (4)$$

where  $\phi = 0.5$  and  $\epsilon_t \sim N(0, 0.1)$ . The resulting series of market inefficiency is shown in Figure ???. We see that there are much more frequent spikes of inefficiency, and the level of inefficiency is much higher than in the efficient market.

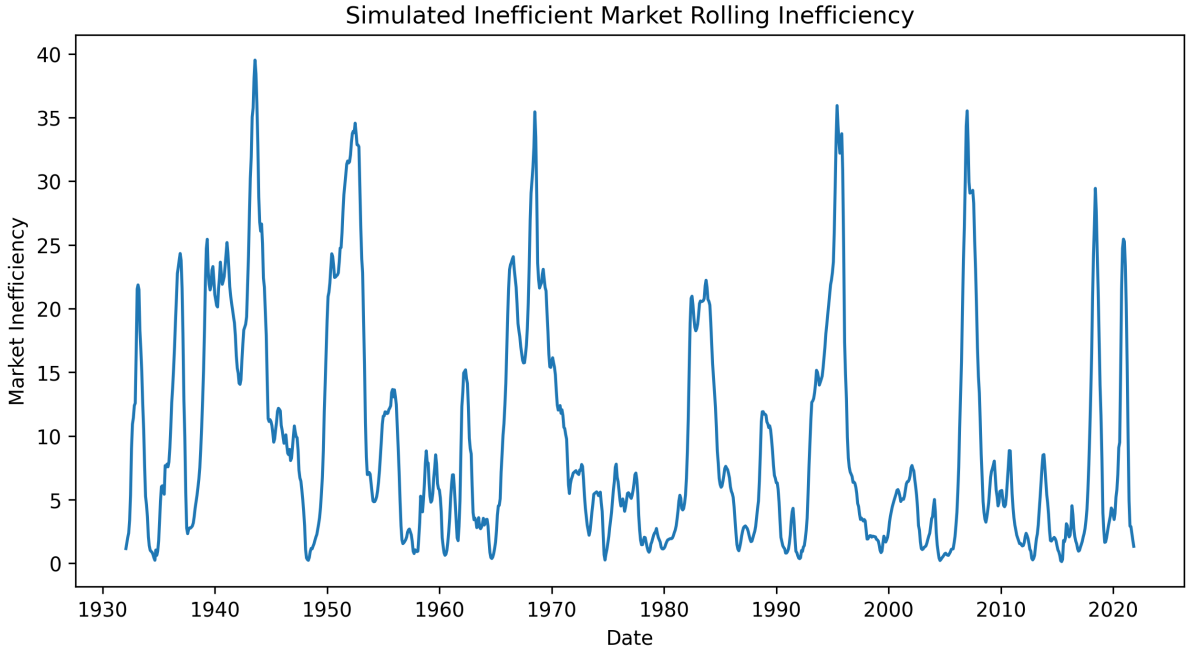


Figure 8: Simulated Inefficient Market

### A.1.3 $R^2$ and Intuition Building

The  $R^2$  value of the unbiasedness regression is a measure of how well the partial returns from 0 to  $t$  predict the total return from  $t$  to  $T$ . While  $\beta_t$  measures the efficiency with which the market processes information,  $R^2$  measures the information content of the partial returns.

I provide an example of the  $R^2$  and  $\beta_t$  values around 'Earnings Months' and 'Non-Earnings Months' in the SP500. I define 'Earnings Months' as the months where the SP500 companies typically report their earnings (January, April, July, October), and 'Non-Earnings Months' as the remaining months.