

# White Noise

Tristan Farrow

## Introduction

Time series analysis relies on understanding the statistical properties of data over time, particularly whether a series contains predictable patterns or consists of pure randomness. Discrete white noise, a sequence with zero mean, constant variance, and no autocorrelation, serves as a benchmark for determining whether patterns have been successfully removed from data. While theoretical white noise is easy to construct and verify through simulation, real-world financial data rarely exhibits these properties in its raw form. Stock prices, for instance, typically display strong trends and autocorrelation that violate white noise assumptions. However, through appropriate statistical transformations such as differencing, we can convert non-stationary price data into white noise, revealing the underlying random fluctuations and enabling reliable modeling and forecasting. This analysis explores these concepts by first establishing white noise principles using simulated data, then applying those principles to examine actual stock price data from SoFi, demonstrating both how real-world series fail to meet white noise criteria and how differencing can transform them into white noise processes suitable for further analysis.

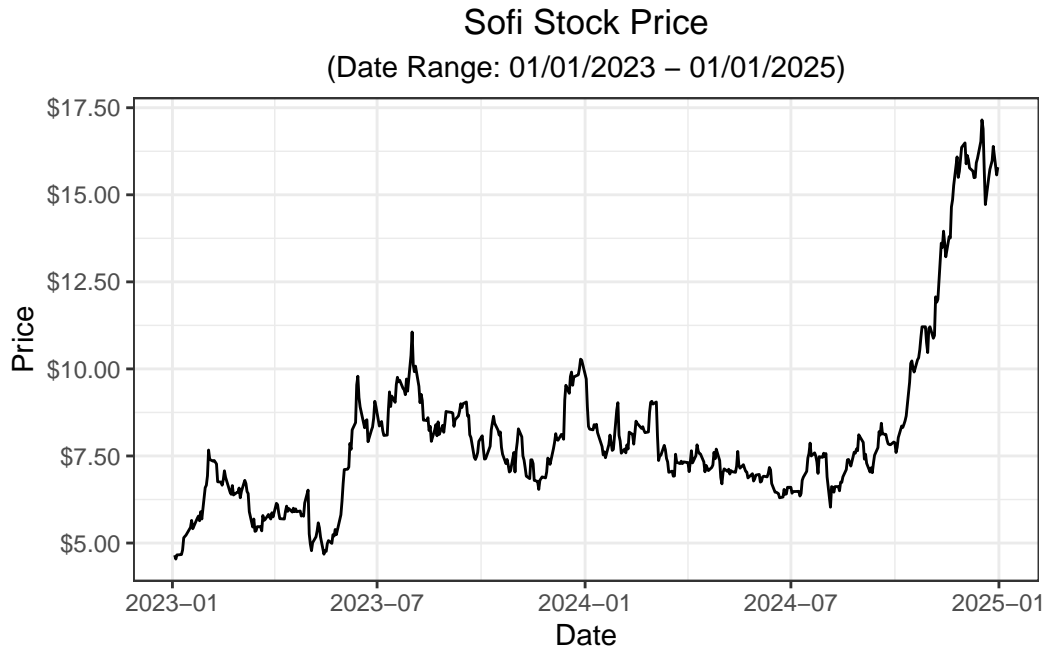
## Important Concepts

### Stationary Data

A stationary time series does not have a trend, seasonal patterns, or other systemic changes so that the mean stays the same, variability is relatively the same, and no recurring patterns. Time series models and forecasts often are best with stationary data or are based on the data being stationary. Transformations can be made to convert a non-stationary series to a stationary one. In a stationary series, the behavior you observe today is statistically similar to what you would observe next month or next year. Many models rely on the idea that the relationship between current values and past values stays consistent over time; without stationary, the model's parameters would change as the series evolves, making estimation unreliable and forecasts inaccurate.

## Data

The analysis uses historical daily opening prices for Sofi Technologies stock, sourced from the tidyquant R package, which provides convenient access to financial market data.



## Company Profile

[SoFi](#) (Social Finance, Inc.) is a financial technology company that offers a range of digital banking and lending products. Here's what they do:

**Company Overview:** SoFi Technologies is a financial services provider based in San Francisco, California with services available across the United States, Latin America, Canada, and Hong Kong. The company offers lending, financial services, and other products that help its members borrow, save, spend, invest, and protect their financial futures.

**Core Services:** SoFi provides personal loans, student loan refinancing, mortgage lending, and investment products through an online platform. They're known for offering competitive rates and streamlined digital processes that aim to make financial services more accessible.

**Banking:** They operate as an online bank, offering checking and savings accounts, along with cash management features. Their accounts typically come with perks like no monthly fees and competitive interest rates.

**Investing:** SoFi provides brokerage services, allowing customers to invest in stocks, ETFs, and cryptocurrency. They also offer robo-advisor services for automated portfolio management.

**Business Model:** They make money primarily through interest income on loans and deposits, investment fees, and other financial service charges. They're a publicly traded company (ticker: SOFI) and have obtained a national bank charter, which allows them to operate more like a traditional bank while maintaining their fintech-forward approach.

In essence, SoFi tries to be a one-stop financial shop for customers who prefer managing their money digitally rather than through traditional brick-and-mortar banks.

## Discrete White Noise

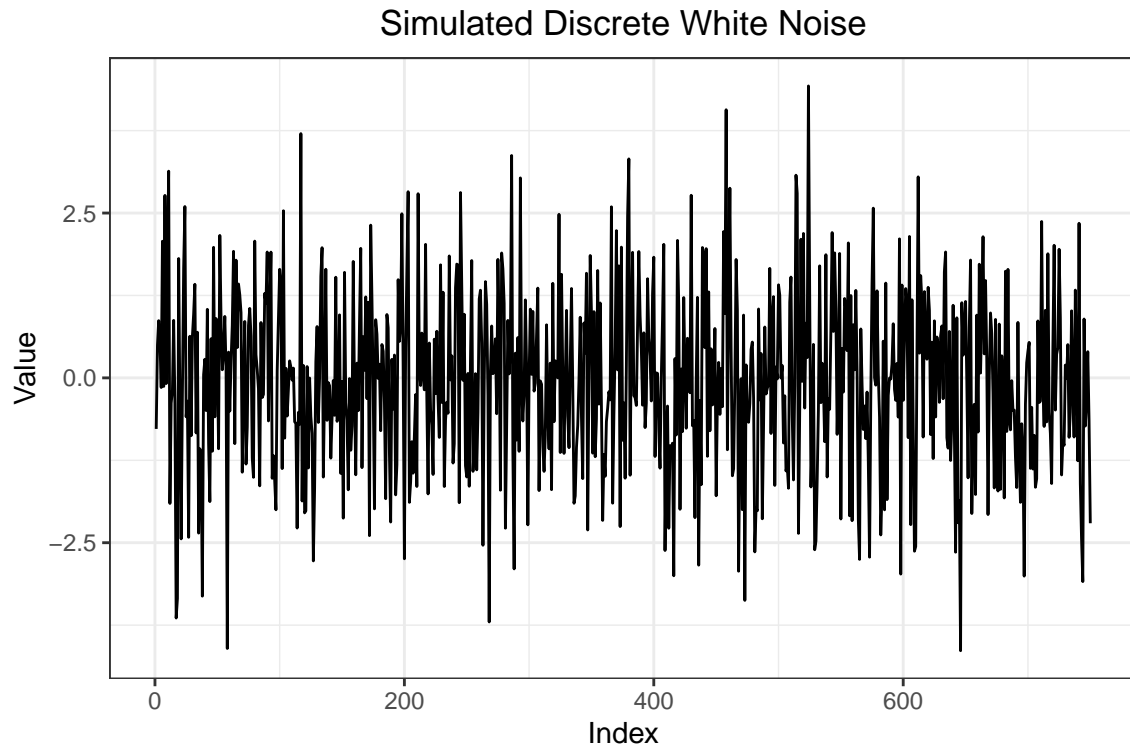
**Definition:** Discrete white noise is a sequence of random variables that has three key properties:

1. A mean of 0
  - The data is centered around 0 with no upward or downward drift
2. Constant Variance
  - The variance does not change over time and the level of randomness is stable
3. No Autocorrelation
  - Each noise term is independent from past and future noise and has not predictable pattern

White noise is essentially pure randomness. If these assumptions are not true then the data is not white noise. Stationarity and white noise are closely related concepts, with white noise representing the ideal endpoint of a stationary time series. A stationary process maintains constant statistical properties, like mean, variance, and no autocorrelation, over time, and white noise satisfies all these conditions perfectly, making it the most extreme form of stationarity.

## Simulated Data

Simulated data offers a controlled environment for demonstrating theoretical concepts while preserving the essential statistical properties of real-world phenomena. By using simulation, we can introduce the principles of discrete white noise with complete transparency and accuracy before applying these concepts to actual data.



#### **Constant Variance**

Examining the chart visually, the variance remains consistent throughout the entire series, with fluctuations maintaining a stable amplitude. Any observable changes in the magnitude or scale of these fluctuations would indicate non-constant variance, which would violate the white noise requirement.

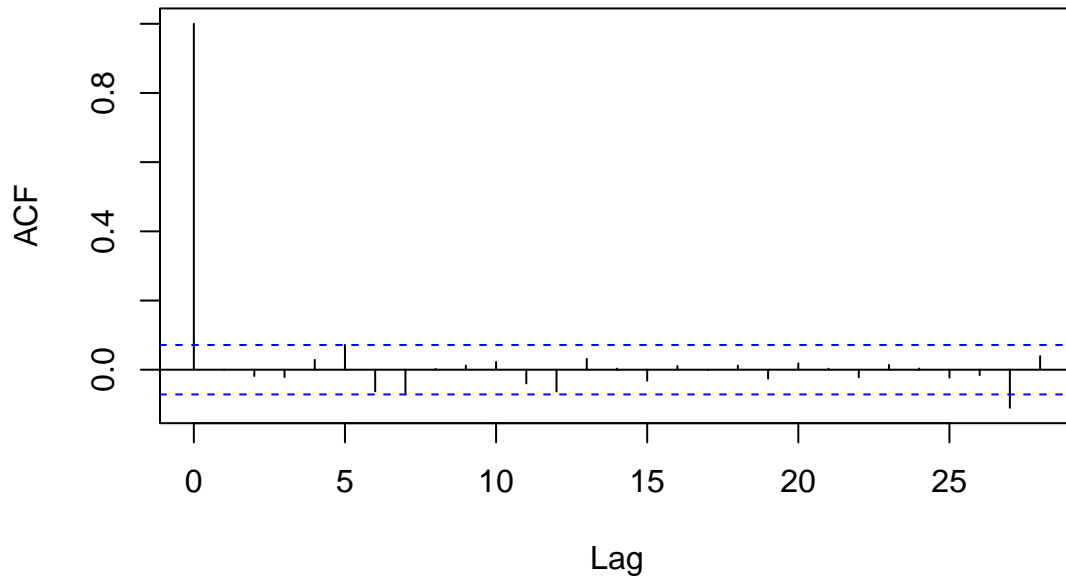
#### **A Mean of 0**

```
[1] "Simulated Data Mean:  -0.0604"
```

The mean is approximately zero following the white noise assumptions. This being not exactly zero but rather close is close enough practically to say that it does not violate the assumption of a mean of 0.

#### **Autocorrelation**

## Simulated Data Correlogram



### Constant Variance

The chart above is called a correlogram and this is how to read it. Each line represent a lag in time compared to a point in time and the length of the bar is the degree of correlation between the two. The taller the line the greater the correlation and the blue lines represent the 95% confidence interval. The bar that go outside the blue lines can be considered statistically practical but should be looked at in the context of the data if it is practically significant. We can expect for 5% of the data to fall outside of these lines. The bar at lag 0 will always be 1 as the series is perfectly correlated with itself.

This shows that the there is no autocorrelation in the simulated data across time as would be expected from a white noise.

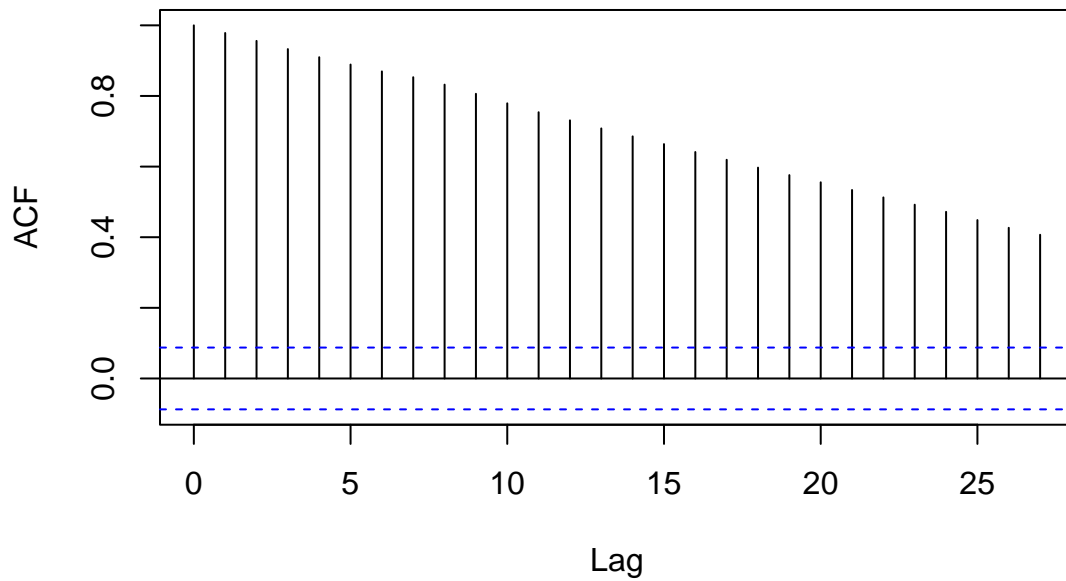
Discrete white noise represents the purest form of randomness in time series data, defined by three fundamental properties: a mean centered at zero, constant variance that remains stable over time, and complete independence between observations with no autocorrelation. Using simulated data generated from a normal distribution, we can clearly demonstrate these principles before applying them to real-world scenarios. The simulation confirms a mean very close to zero, displays visually consistent variance throughout the series, and produces a correlogram showing no significant autocorrelation. This simulated example serves as a controlled baseline, illustrating what pure white noise looks like when all three assumptions are satisfied, thereby establishing the criteria against which actual time series data can be

evaluated to determine whether it qualifies as white noise or exhibits patterns that suggest underlying structure.

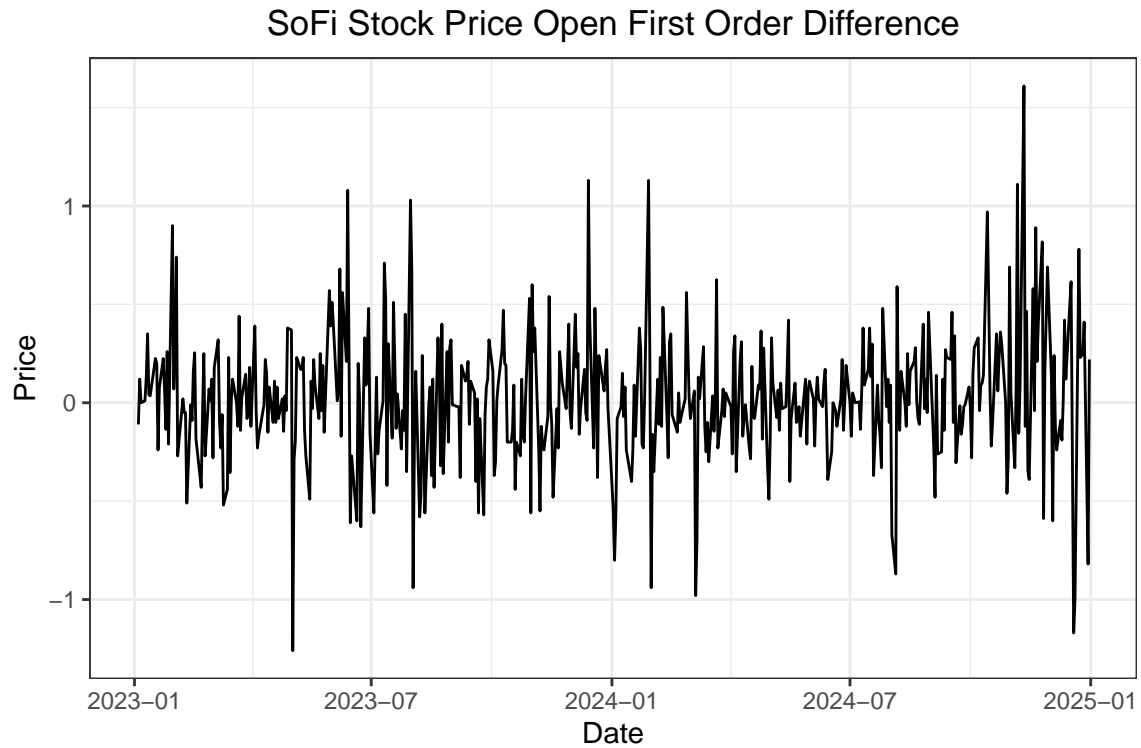
## Stock Data

### SoFi

#### SoFi Opening Price Correlogram



The correlogram of SoFi opening prices reveals high autocorrelation across all lags, with each lag exhibiting statistical significance. This indicates the series is not white noise. However, applying a first-order difference, which measures the change between consecutive observations, transforms the series into white noise.

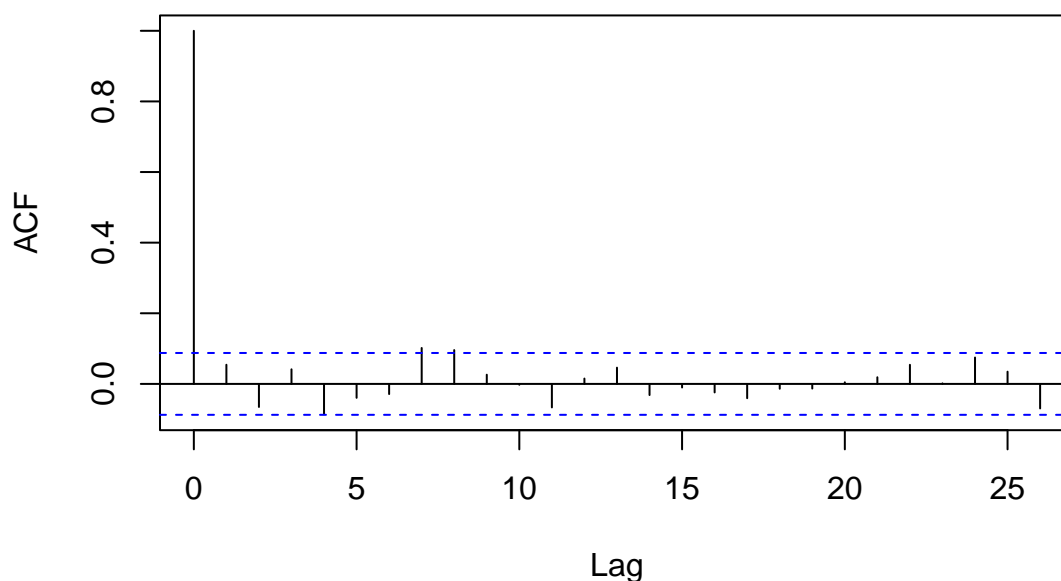


#### **Constant Variance**

The rolling variance displays a pattern far more consistent with simulated white noise than the original stock price series, with fluctuations remaining relatively stable throughout. However, the final 10% of the data shows some deviation from this overall pattern, though additional statistical testing would be required to determine whether this represents a meaningful violation of the constant variance assumption.

#### **Autocorrelation**

## SoFi First Difference Correlogram



The correlogram displays the expected pattern for white noise, with the lag-0 bar at perfect correlation (1.0) and nearly all subsequent lags falling well within the blue dashed confidence interval lines, indicating no significant autocorrelation across the series.

### A Mean of 0

```
[1] "First Order Open Price Differnce:  0.0222"
```

The mean is approximately zero and appears to follow the assumptions of white noise.

### Conclusion

The analysis of SoFi stock data demonstrates a critical principle in time series analysis: while raw price data exhibits strong autocorrelation and fails to qualify as white noise, transforming the series through first-order differencing, calculating the change between consecutive observations, reveals an underlying white noise process. This transformation illustrates why differencing is a fundamental technique in time series modeling. It removes the trending behavior that characterizes non-stationary price data and exposes the random fluctuations beneath, allowing analysts to work with stationary white noise that can be reliably modeled and forecasted.