

# VIX and Metaculus Correlation

Varun Varanasi

September 23, 2022

This document outlines a series of statistical tests and methods used to evaluate the correlation between the returns of metaculus user data and the returns of financial timeseries. All code can be found in the `return_corr.py` script. An outline of each statistical method and its corresponding results can be found below. All tests were run using data between 1/1/2021 and 4/7/2022.

## Correlation Methods

### Pearson Correlation

Pearson correlation is a measure of linear correlation between two datasets.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$$

Pearson correlation coefficients range between -1 and 1 indicating the strength and direction of linear correlation.

### Spearman Correlation

Spearman (rank) correlation is a measure of rank correlation or how well a monotonic function can approximate the relationship between the two datasets. Spearman correlation is simply calculated as the pearson correlation of the ranked data. With a coefficient range between -1 and 1 Spearman correlation coefficients provide insight into the general strength and direction of correlation between two datasets (linear or otherwise).

|          |  | VIX   | SPY   |
|----------|--|-------|-------|
| Pearson  |  | 0.095 | -0.04 |
| Spearman |  | 0.098 | 0.009 |

### Cross Correlation

Cross correlation is a timeseries correlation method in which the timelags of one timeseries are pearson correlated with the other. A summary of the 5 strongest time-lag correlations can be found below.

|   |  | VIX            | SPY            |
|---|--|----------------|----------------|
| 1 |  | Lag 314: 0.61  | Lag 315: -0.96 |
| 2 |  | Lag 315: 0.49  | Lag 314: -0.72 |
| 3 |  | Lag 293: 0.33  | Lag 311 -0.62  |
| 4 |  | Lag 295: -0.31 | Lag 295: 0.53  |
| 5 |  | Lag 313: -0.30 | Lag 313: 0.51  |

## Time Lagged Regression

Time lagged regression is the process of estimating a timeseries by using another timeseries and its timelags as predictors.

$$\hat{Y} = \beta_1 X + \beta_2 X_{t-1} + \beta_3 X_{t-2} \dots$$

A summary of OLS regression with the metaculus data and its timelags as predictors can be found below. Timelags were restricted to 45 days for model simplicity.

|                              |  | <b>VIX</b> | <b>SPY</b> |
|------------------------------|--|------------|------------|
| Adjusted R-squared           |  | 0.016      | 0.121      |
| Statistical Significant Lags |  | 3          | 8          |

## LASSO Regression

LASSO regression is a sparse regression method capable of removing predictors from the model. For this reason, it is often used for feature selection. In the context of this project, we used LASSO regression on the 45 day lag predictor set to select the most informative ones. A summary of our results can be found below.

|                              |  | <b>VIX</b> | <b>SPY</b> |
|------------------------------|--|------------|------------|
| Adjusted R-squared           |  | 0.053      | 0.120      |
| Remaining Predictors         |  | 5          | 6          |
| Statistical Significant Lags |  | 2          | 4          |

## Granger Causality

Granger Causality is a statistical method used to determine whether a given timeseries provides explanatory power on another timeseries. The statistical test specifically measures whether adding the timelags of the series in question improves an autoregressive model of the second timeseries. Once again, tests were limited to 45 lags for model simplicity. A summary of the most significant time lags and their corresponding p-values can be found below.

|   |  | <b>VIX</b>   | <b>SPY</b>   |
|---|--|--------------|--------------|
| 1 |  | Lag 1: 0.11  | Lag 1: 0.03  |
| 2 |  | Lag 2: 0.14  | Lag 3: 0.07  |
| 3 |  | Lag 3: 0.16  | Lag 2: 0.10  |
| 4 |  | Lag 17: 0.04 | Lag 16: 0.01 |
| 5 |  | Lag 18: 0.05 | Lag 15: 0.02 |

## Dynamic Time Wrapping

Dynamic Time Wrapping is a process commonly used in signal processing to assess the similarity of two timeseries. At a high level, the method works to find minimize the euclidian distance between two timeseries by mapping points from one timeseries to another. A summary of the average distances between the timeseries can be found below.

|                  |  | <b>VIX</b> | <b>SPY</b> |
|------------------|--|------------|------------|
| Distance         |  | 309.23     | 344.07     |
| Average Distance |  | 0.67       | 0.74       |