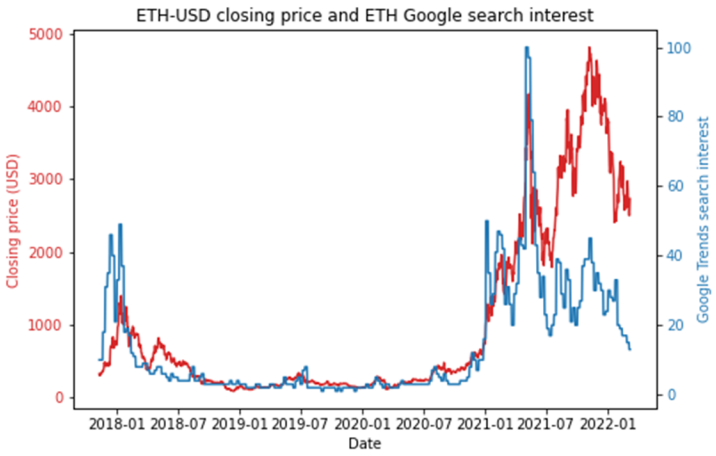
Predicting Ethereum's Price from Google Search Data (An attempt)

Kyle Weng - Ge/Ay 117, Winter 2022

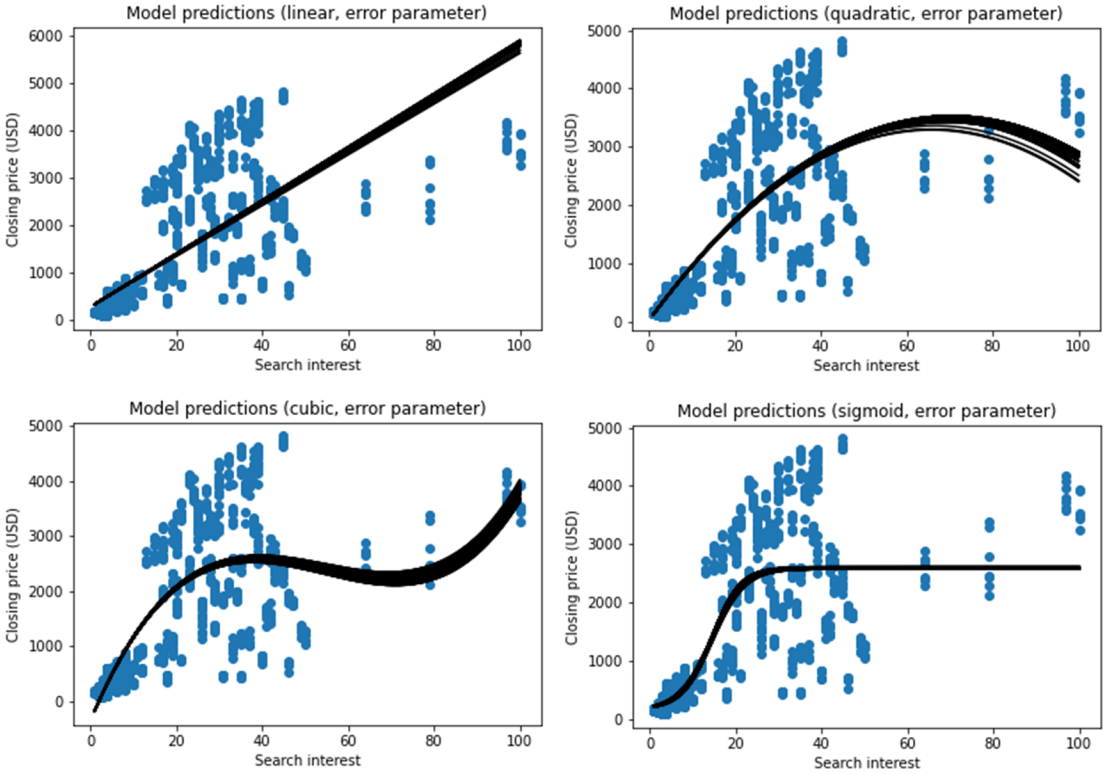
Parameterizing error makes a difference in model fits!

Error term: Quadratic model has best fit
No error term: Sigmoid model has best fit



Abstract

We fit several models to data of Ethereum's closing price versus its Google search interest to find how well simple models could perform when applied to complex, noisy data. Both our initial parameters and our MCMC results were largely unsuccessful. However, besides the models themselves, introducing a proper measure of error and interpolating search data more intelligently would improve the results. This analysis lays the groundwork for future research in this area.



Data and Methods

We used Google's search trend data and Yahoo Finance's Ethereum price data. First, we made adjustments to align the frequency of our datasets (originally weekly and daily, respectively). Then, we obtained initial parameter estimates via optimal estimation. We used those values to run MCMC chains, from which we obtained median model parameters, 68% confidence intervals and BIC (Bayesian information criterion) values.

We ran at least two versions of each of our models: one that did not parameterize error and one that did. We found that the latter's MCMC chains did not converge, as their error parameters could be indefinitely increased to decrease their chi-squared values. Besides the actual parameters, this also affects the models' BIC.

From the models that didn't parameterize error, the sigmoid model has the lowest BIC and is thus the best fit; otherwise, the quadratic error-parameterizing model has the lowest BIC.

Takeaways + Future Work

First, we plan to use more sophisticated models. We will consider the reverse possibility (that Ethereum's price affects Google search interest). Second, we plan to define an error metric incorporating volatility and the fact that prices do not update instantaneously across all exchanges.