Kalman and Recursive Bayesian Filtering for Realizing Stock Market Processes

Omar Riaz
Northeastern University

2021 April

1 Introduction

Eugene Fama, the Father of the efficient market theory, described the total accumulation of noise onto stock prices as independent, random, and unrelated to real world events, in order to justify a random walk model. The random walk hypothesis has since been disproven by autocorrelation functions and spectral analysis, [1]. Market prices have non-stationary means, and correlation between consecutive prices. A mathematical model that harnesses these correlations while considering the uncertainty in noisy stock data, relative to some steady instrinsic market value, would be suited to the application of realizing this Brownian stock market motion [2]. In this project a Kalman filter will be used to estimate a smooth trend-line within the data which represents instrinsic value before perturbation by market noise. [3]

Kalman filtering is an algorithm usually performed on time series data with statistical noise, which produces estimates of a an unknown true state. We assume a time series' process can be modeled recursively:

$$x_{k+1} = \Phi x_k + w_k$$

$$z_k = Hx_k + v_k$$

where z_k is our measurement vector (market price), and x_k is our state vector (intrinsic value). Both w_k and v_k are Gaussian white noise models with covariance matrices called Q_k and R_k, respectively. Given parameters for the process and an a priori Gaussian estimate of the state's mean μ_k and covariance P'_k, a prediction of the initial state x'_k is formed. Then, when the measurement z_k is observed, both this prediction and the state covariance are updated:

$$x_k = x'_k + K_k(z_k - Hx'_k)$$

$$P_k = (I - K_k H)P'_k$$

where K_k is the Kalman gain: a minimum mean-square error estimator. The updated state estimation is then projected into the prediction of x'_{k+1} using the recursive process model.

2 Methods

To define the filter, the following matrices must be specified: Φ, the state-transition model, H, the observation model, Q, the covariance of the process noise, R, the covariance of the observation noise

Because we assume the market noise to be completely random, the observation model is assumed as unity. The state-transition model correlates to market momentum which is calculated as the previous day's closing price subtracted from the current closing price.

Bassett [4] suggests initalizing Q_k to the empircal covariance from the previous day. While using the covariance between two days was be sufficient in estimating the uncertainty in consecutive states, this covariance was also inflated by including varying numbers of prior market prices. R_k was based on the bid ask spread of 1/8 tick size, as the discrepancy between ask price and bid price indicates tight liquidity and thus uncertainty [5].

The Gaussian distribution representing the initial intrinsic value state x'_0 may be determined using a variety of valuation techniques. For the purposes of this project, the initial state x_-1 will be chosen as the last closing price. Likewise, fundamental analysis techniques would traditionally be employed to manage risk and estimate an initial uncertainty, but we initialize $P'_0 = Q$.

An Expectation-Maximization algorithm is also implemented to optimize model parameters. However, this optimization problem is non-convex, meaning it may converge off the global maximum, so sensible initial parameters were prepared. The EM algorithm seeks to maximize the likelihood of observation given the current state mean and covariance.

The Kalman filter is implemented in Python using the `pykalman` library[6]. A Kalman smoother function is also tested, which iterates through and updates the entire history of states every time a measurement is observed. Because the smoother estimate considers future data, it has no advantage in prediction.

3 Results

The result for applying the Kalman filter to the AAPL and GME stock data from the duration of January 1st, 2020 through June 1st, 2020, is shown in the Results Appendix A. Each figure contains a plot of the stock data, the Kalman estimate of the data, and the Kalman smoother estimate. Two figures for each ticker are shown. The first with initial covariance calculated from prior bar data, the second with initial parameters determined with EM.

References

[1] James Martin Rankin. *Kalman filtering approach to market price forcasting*. PhD thesis, Iowa State University, Ames, Iowa, 1986.

[2] Rick Martinelli & Neil Rhoads. Predicting market data with a kalman filter. `https://www.haikulabs.com/pmdwkf26.htm`, 2009. Accessed: 2021-4-19.

[3] Tom Arnold, Mark J. Bertus, and Jonathan Godbey. A simplified approach to understanding the kalman filter technique. *The Engineering Journal*, 53:140–155, 2008.

[4] Jr Gilbert W. Bassett, Virginia G. France, and Stanley R. Pliska. Kalman filter estimation for valuing nontrading securities, with applications to the mmi cash-future spread on october 19 and 20, 1987. *Review of Quantitative Finance and Accounting*, 1:135–151, 1991.

[5] Roger R. Labbe. Kalman and bayesian filters in python. `https://nbviewer.jupyter.org/github/rlabbe/Kalman-and-Bayesian-Filters-in-Python/tree/master/`, 2015. Accessed: 2021-4-20.

[6] Daniel Duckworth. pykalman. `https://github.com/pykalman/pykalman`, 2012.

A Results Appendix

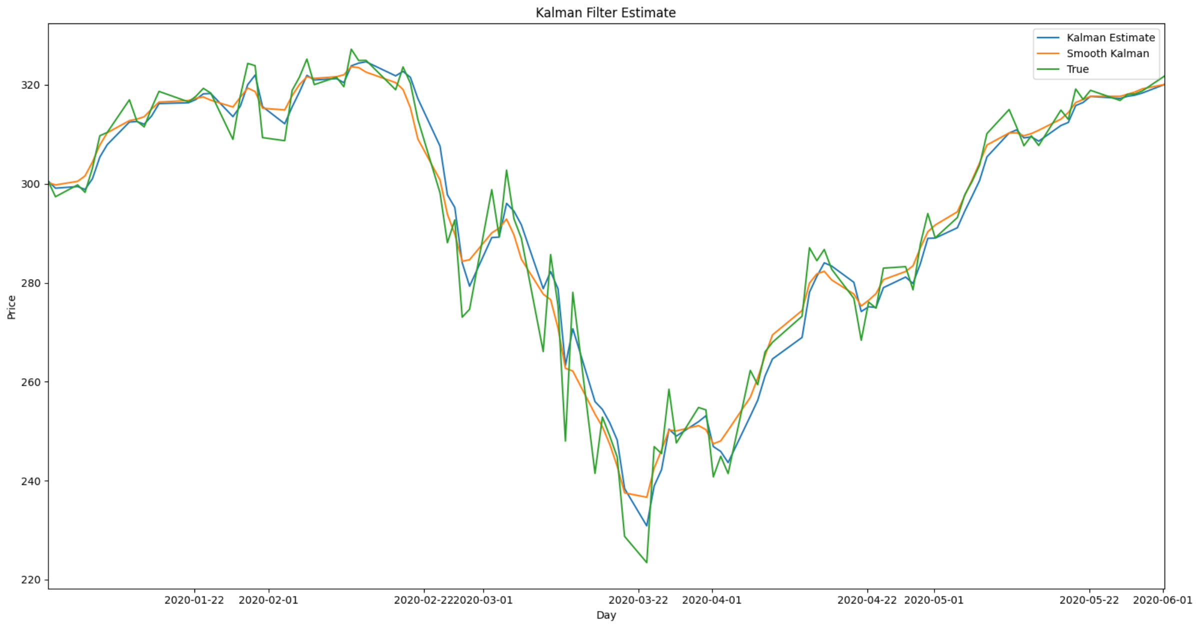

Figure 1: AAPL Kalman Estimation

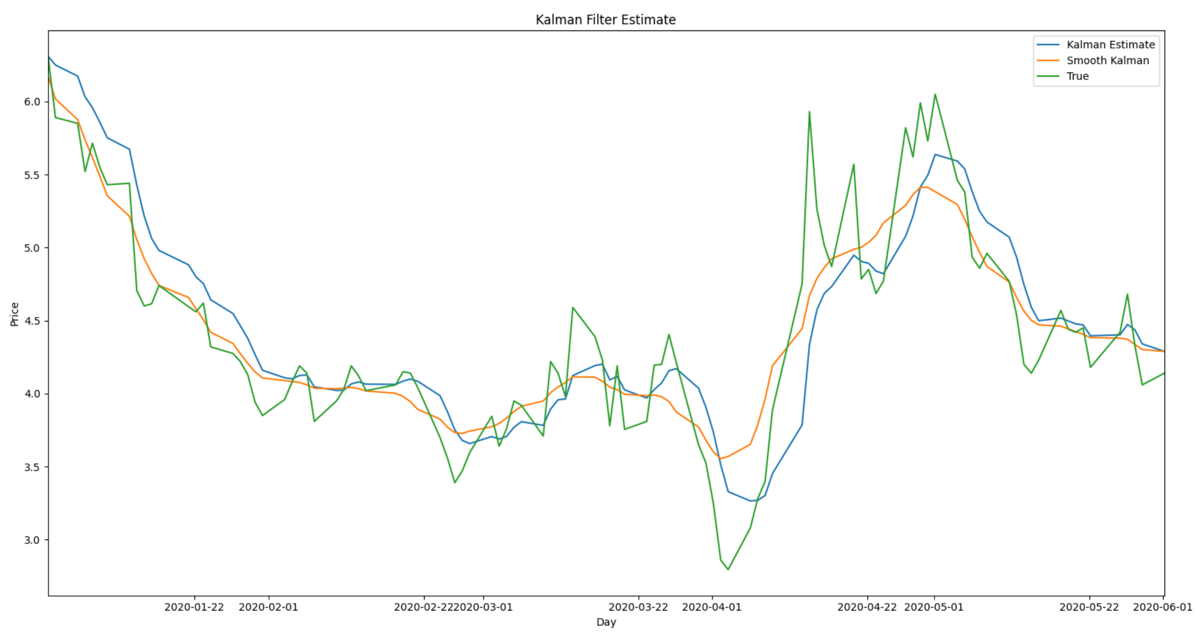

Figure 2: GME Kalman Estimation

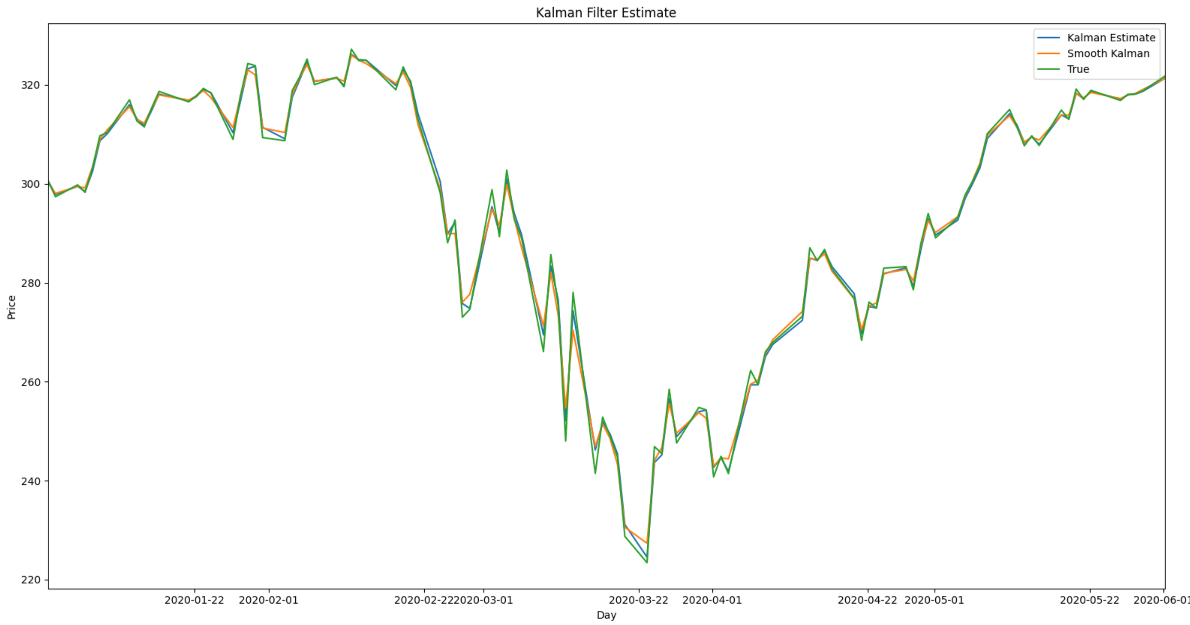

Figure 3: AAPL Kalman Estimation with EM optimized parameters

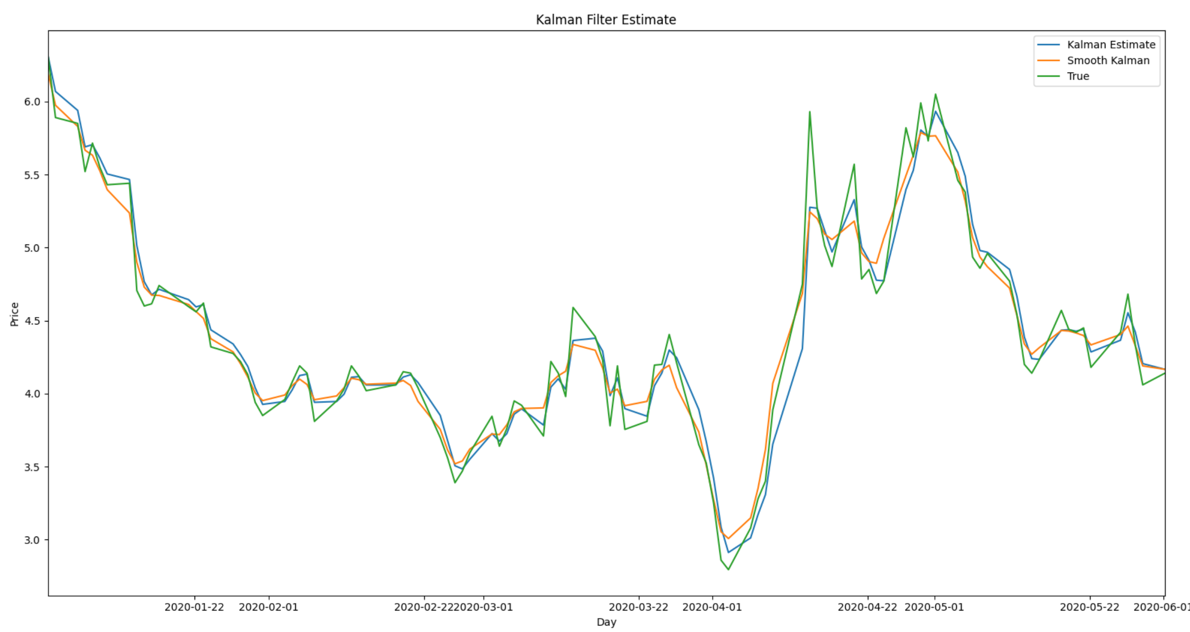

Figure 4: GME Kalman Estimation with EM optimized parameters