Kalman Filter vs Alternative Modeling Techniques and Applied Investment Strategies

Submitted by

Heather E. Dempsey

Doctor of Business Administration in Finance Program

In partial fulfillment of the requirements

For the degree of Doctor of Business Administration in Finance

Sacred Heart University, Jack Welch College of Business and Technology

Fairfield, Connecticut

Date: December 17, 2021

Dissertation Supervisor: Dr. W. Keener Hughen   Signature:

Committee Member: Dr. Lior Menzly   Signature:

Committee Member: Dr. Michael Gorman   Signature:

Kalman Filter vs Alternative Modeling Techniques and
Applied Investment Strategies

By
Heather E. Dempsey
Doctor of Business Administration in Finance Program

Thesis
Submitted as partial fulfillment of the requirements for the degree of Doctor of Business
Administration in Finance in the Jack Welch College of Business
Sacred Heart University 2021

Fairfield, Connecticut

Defense Committee:

Dissertation Supervisor: Dr. W. Keener Hughen

Committee Member: Dr. Michael Gorman

Committee Member: Dr. Lior Menzly

Dedicated to my father, Jack Dempsey.

Acknowledgements

Dr. W. Keener Hughen, my mentor, for his limitless help, Dr. Lior Menzly for the idea of creative destruction, and to both of them for making the time to answer any questions I had.

HD

Abstract

This thesis examines the efficacy of alternative modeling techniques to predict stock market returns modeled with time-varying coefficients with the goal of developing and implementing a trading strategy that yields excess returns. First, we determine the modeling technique with the smallest forecast error using historical predictors: the differenced dividend-price ratio, lagged S&P 500 returns, and the change in implied volatility. The candidate modeling techniques include both constant and recursive ordinary least squares (OLS) regression methods and diverges from previous return forecast literature with the comparison of a state-space model (SSM) cast as a VAR(1) process to each OLS technique. The state-space model is found to be the superior modeling technique with the smallest RMSE 3.76% and greatest out-of-sample R^2 of 2.62% using delta VIX as the forecasting variable. Second, we demonstrate economic significance, using 1) monthly stock return forecasts in a market timing strategy, and 2) daily price forecasts in a simulated live pairs trading strategy taking into account implementation shortfall. In both trading strategies, the state-space model Kalman filter significantly outperforms the alternative OLS modeling techniques with an annualized total return of 21.64% in the market timing strategy and an annualized total return of 13.21% unlevered in the pairs trading strategy.

Contents

List of Tables

List of Figures

Used abbreviations and symbols

~	distributed as
ADF	Augmented Dickey-Fuller
D/P	dividend-price ratio
DW	Durbin–Watson statistic
EGT	Engle-Granger test
EWOLS	expanding window OLS
HMM	hidden Markov model
i.i.d.	identically and independently distributed
KF	Kalman filter
LL	loglikelihood of data under Gaussian innovations
MAE	mean absolute error
MSE	mean-squared error
OLS	ordinary least squares
RMSE	root mean-squared error
RWOLS	rolling window OLS
R^2	coefficient of determination
SSE	sum of square errors
SSM	state-space model
SSR	regression sum of squares

SST	total sum of squares
TVP	time-varying parameters

Greek Letters

β_t	Latent-state variable beta vector at time t
ε	observation disturbances
ϵ	state-transition disturbances
Ω	variance/covariance matrix of state errors
σ^2	measurement variance
δ	standard deviation
φ	root
Δ	differenced one period

Other letters

a_t	level term of regression model; the 1st element of beta vector
b_t	slope coefficient on independent variable; 2nd element of beta vector
I_n	n-dimensional identity matrix
Q_{sq}	Ljung–Box test statistic
R	observation error variance

1. Introduction

This thesis examines the efficacy of alternative modeling techniques to predict stock market returns modeled with time-varying coefficients. We identify the modeling technique that best forecasts monthly returns—the model with the smallest forecast error from alternative modeling techniques. We aim to answer three empirical questions: 1) are market returns forecastable? 2) Can forecasts be improved using a state-space model (SSM) approach instead of recursive and constant OLS regression? 3) If so, can that knowledge be exploited to develop investment strategies that provide excess returns in a realistic setting?

The body of literature seeking to forecast stock market returns has remained inconclusive. Goyal and Welch (2008) using dividend-price ratio amongst others found poor predictive performance in-sample and out-of-sample. Cochrane (2008) similarly finds poor out-of-sample R^2 and weak return forecasting power using the dividend yield and regression (Pesaran & Timmermann, 1995) with no updating. Menzly, Santos, & Veronesi (2004) discuss the instability of the dividend yield as a predictor of returns finding a non-linear relationship with market beta and price-dividend ratio amongst other market observed variables. The interaction between state or unobservable variables that are time-varying are shown to influence return predictability. Rapach et al. (2009) found combinations of linear forecasts have greater stock return predictability using the constant OLS model. Leitan and Ludvigson (2001) find broad market index returns are predictable using the wealth ratio (Kandel & Stambaugh, 1996) and Chiang and Hughen (2017) find strong predictive performance between the curvature factor of the oil future's curve and stock market returns.

Each of the aforementioned authors relied on OLS methods to estimate coefficients, which when constant produce the best linear unbiased estimates (BLUE). However, there are reasons to believe coefficients may vary with time and that OLS estimates are no longer optimal. For example, consider investors' risk preferences, confidence in future investment stability, or expected dividend growth (Menzly et al., 2004). There are also exogenous shocks affecting the

parameters of financial time-series, i.e., the COVID-19 pandemic which changed parameter values for variables such as economic output, interest rates, and unemployment rates.

Faff et al. (2000) and Mergner & Bulla (2005) compare different time-varying parameter (TVP) modeling techniques when estimating time-dependent systematic risk in the contemporaneous CAPM (Adrian & Franzoni, 2009) but do not address the problem of forecasting returns. The current study expands on this objective by comparing modeling techniques to forecast monthly S&P500 returns using three independent variables: change in dividend-price ratio (ΔD/P), the change in the implied volatility index (ΔVIX), and the lagged S&P 500 return, each in separate forecasts, not as a multivariate model. This study further extends the literature by allowing both coefficients in the forecasting relationship, the intercept, and the loading on the independent variable, to vary with time.

The forecasting relationship explored in the current study between the broad market and an independent variable is similar to the well-known single factor capital asset pricing model (CAPM) but differs in two important aspects. First, CAPM seeks to *explain* the relationship between the dependent and explanatory variables, whereas the current study attempts to *forecast* the value of the dependent variable. Second, CAPM assumes time-invariant coefficients, whereas we model parameters that evolve with time.

There are several approaches to estimating TVP, however, no research until now has compared existing OLS based methods to a state-space model (SSM) to forecast returns on a broad market index. This thesis addresses this gap by employing a SSM approach to compare the forecast performance to three OLS based methods: constant OLS, expanding window OLS (EWOLS), and rolling window OLS (RWOLS). The goal of the comparison is to identify the superior method—the technique generating the minimum forecast error.

The need for empirical identification of optimal TVP forecasting techniques to forecast returns on the market, is in part, what this thesis aims to satisfy. In addition, we demonstrate the

potential gains in wealth resulting from applying the best technique to 1) a market-timing investment strategy and 2) a pairs trading strategy that explicitly accounts for implementation shortfall.

This thesis finds in support of stock return predictability, the ΔVIX and Kaman filter specification had the highest out-of-sample R^2 of 2.62%. Lagged S&P returns using EWOLS exhibited some explanatory power with the second highest out-of-sample R^2 of 2.35%. However, the magnitude of the size effect is insignificant for each variable suggesting a weak relationship with returns for all examined predictors.

We find the Kalman filter using ΔVIX as the forecasting variable outperforms all other methods in terms of RMSE, in-sample and out-of-sample R^2 and had the largest annualized total return from the market timing strategy of 21.64%. The pairs trading strategy using the Kalman filter, produced an annualized total return of 13.21 % unleveraged. The results are scalable and the potential for higher returns exists.

2. Methodology

The general one-step-ahead linear forecast of y, using x as the forecasting variable can be expressed as

$$y_{t+1} = a + bx_t + \varepsilon_{t+1} \tag{1}$$

Eq. (1) can be rewritten in matrix notation as

$$y_{t+1} = X_t \beta + \varepsilon_{t+1} \tag{2}$$

where $\beta = \begin{bmatrix} a \\ b \end{bmatrix}$ is a $(2x1)$ vector

and X t is a $(1x2)$ vector whose first component is the constant 1, and second component is the forecasting variable x_t.

The question now is how to choose the best values of a and b, components of the vector β? "Best" here means the values of a and b for which the forecast yields the smallest RMSE among all linear forecasts that use X t as the independent variable. The coefficients a and b are either time-invariant or time-variant. If a and b are time-invariant, they can be best estimated by constant OLS. In this case the OLS estimates yield the best linear unbiased estimates (BLUE), defined by

$$\beta^{OLS} = (X'X)^{-1}X'y \tag{3}$$

where the following model assumptions are satisfied:

- linear in its parameters
- expected value of the errors is zero
- no multicollinearity
- no serial correlation
- constant error variance (homoscedasticity)

However, if the coefficients evolve with time, OLS is no longer appropriate. In this case, the most commonly used techniques to estimate the elements of β_t, that is the intercept a_t and the loading b_t, are RWOLS and EWOLS. These are hybrid OLS approaches that address the time-varying parameter problem by simulating the changes in β_t using a moving/expanding estimation window. However, both these models assume that the parameters are constant in each estimation window, which is not strictly consistent with the notion that the parameters vary, in our case, monthly.

2.1. Modeling techniques

To calculate parameters a_t and b_t which are used to calculate next month's return forecast, we examine the following three modeling techniques comparing each to the benchmark constant OLS model.

- Rolling window ordinary least squares (RWOLS)
- Expanding window ordinary least squares (EWOLS)
- Kalman filter (KF) parameterized as a bivariate first-order autoregressive AR(1) process

2.1.1. Constant OLS

The constant OLS model can be expressed as

$$y_{t+1} = X_t\beta + \varepsilon_{t+1} \qquad \sim N(0, \sigma_\varepsilon^2) \tag{4}$$

where

scalar y_t denotes the market return in period t.

X_t is the $(1x2)$ forecasting variable observed at the beginning of period t.

β is the $(2x1)$ vector containing the variables of interest: intercept a and weight b, on x_t.

$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ is Gaussian white noise with a constant variance σ_ε^2.

This method uses an in-sample period of 106 months of the 421 months in the entire data sample because it is a quarter of the dataset. The method then uses those estimates formed in the data sample, fixes them, then forecasts one-month ahead for each of the remaining 315 months. OLS is the go-to technique for most econometricians wishing to forecast some variable y. However, this technique makes simplifying assumptions which do not capture the dynamic nature of stock market conditions. The notion that economic relationships are constant is unrealistic—lending to the phrase "past performance is no guarantee of future results".

2.1.2. Time-varying OLS methods

Rolling window OLS (RWOLS) and Expanding window OLS (EWOLS) are defined by

$$y_{t+1} = X_t\beta_t + \varepsilon_{t+1} \qquad \sim N(0, \sigma_\varepsilon^2) \tag{5}$$

where

y_{t+1} is the return forecast.

$X \ t$ is a $(1x2)$ row vector of the forecasting variables.

β_t is a $(2x1)$ vector that contains the variables of interest, the intercept a_t, and weight b_t on the forecasting variable $X \ t$.

RWOLS and EWOLS alternative techniques calculate the changing parameters by making successive approximations. Both models use an initial window size of 106 months. When the in-sample estimates are calculated, they are fixed and used to forecast one month forward. The window is either rolled forward one month, remaining 106 months long, or it is expanded by one month. However, the estimated parameters in RWOLS and EWOLS are fixed, held constant, in each window. Although an improvement over constant OLS, the models remain inconsistent.

2.1.3.   State-space model (SSM)

Due to the highly dynamic nature of financial time series, there is an ever increasing interest in the use of the SSM Kalman filter in econometrics. The KF is a Markov process meaning its next state, one time-step ahead, depends only on its current state. In probability theory, Markov chains model stochastic systems (Zhang, 2004) where the state at time $t + 1$ is determined by its parameters, or state, at time t. Hidden Markov models (HMM) are simply Markov chains observed in noise (Lauri, 2014). HMMs describe those stochastic processes that have unobservable states, for example, the vector β_t, our latent-state variable. The realizations of y_t are observable, but the latent-state variable β_t, which drives the change in y_t, is 'hidden' because of the introduction of noise. The KF relates these two quantities through a system of linear equations. With each prediction, update, measurement, and correction the estimates improve, converging to the correct value. This reduces the variance of the new estimate with each pass by placing greater weight on the value with the least variance going into to the updated estimate. The general model, following Kim (2018), can be expressed as

$$z_k = Hx_k + v_k \qquad\qquad \sim N(\mu, R) \qquad\qquad (6)$$

$$x_{k+1} = Ax_k + w_k \qquad\qquad \sim N(\mu, Q) \qquad\qquad (7)$$

where

x_k is the latent-state variable, an $(nx1)$ column vector; z_k is the measurement or observation vector, an $(mx1)$ column vector; A is the state-transition matrix, an (nxn) matrix; H is the state-to-measurement mapping matrix, an (mxn) matrix; w_k is the state transition or process noise, an $(nx1)$ column vector; and v_k is the measurement noise, an $(mx1)$ vector. Appendix A presents a mapping of the general notation to the KF specifications, and Appendix B presents an in-depth discussion of the KF algorithm.

2.1.4. Bivariate AR (1) Kalman filter

To address the issue of modeling time-varying parameters I propose the following method, the SSM KF parameterized as a bivariate AR(1) process. We chose this specification to reduce the possibility of overfitting by restricting the loading on the latent state variable Eq. (8) to a diagonal matrix we reduce the number of free parameters (Watson, 1983). Additionally, we want to allow for both coefficients to vary with time but not influence the other. We made the simplifying assumption that coefficients are stationary and will mean revert in the long run. This made sense with the lengthier forecast using monthly data rather than the daily data used in a later application where we specify a Random Walk. We tried the Random Walk with the model performance and market timing strategy but the results were not as good as the bivariate AR(1) specification in general. This points to coefficients following a mean-reverting process. In practice it may not always be feasible to have a large data set to train the model and that is the benefit to using the Random Walk in a simulated live environment. The KF is designed to update using only the previous estimate and is computationally light making it easier to implement with no free parameters to calibrate. However, the bivariate AR(1) performed better when we had a quarter of the data set to train. In this approach the EM algorithm is used to calibrate the parameters and estimate the state variables of β_t that provide the best fit for each predictor.

Let β_t denote the 2-vector whose first component is a_t and whose second component is b_t. In this scenario, β_t is an unobserved $(2x1)$ state vector specified as a bivariate AR(1). There are eight free parameters in the bivariate AR(1) approach: two in the constant vector A a $(2x1)$, two in the diagonal constant matrix B, $(2x2)$, three free parameters in the symmetric variance/covariance matrix, Ω, including rho, ρ, and finally, the last free parameter is R, the variance in the observations, Eq. (9). The next state of β_t is estimated by the state-transition equation, Eq. (8):

$$\beta_{t+1} = A + B\beta_t + \epsilon_{t+1} \qquad \sim iid(0, \Omega) \qquad (8)$$

The second equation is the observation equation, Eq. (9), which forecasts scalar y $(1x1)$, the monthly return on the S&P 500, given x at time t:

$$y_{t+1} = X\ t\beta_t + \varepsilon_{t+1} \qquad \sim iid(0, R) \qquad (9)$$

where $X\ t$ is a row vector $(1, x\ t)$ and the scalar variable x_t is one of three independent variables: ΔD/P, ΔVIX, or lagged return. Each independent variable is used to generate a separate forecast for each modeling technique; not as a multivariate model.

2.1.5. Expectation-maximization (EM)

The advantage of the bivariate AR(1) approach is that the expectation-maximization (EM) algorithm can be used to find the maximum log-likelihood (LL) estimates of the hyperparameters and latent-state vector β_t. Given an initial guess for the parameters, which in a SSM can be almost arbitrary (Moon, 1996), the EM involves two steps: 1) the expectation (E-step), where the KF forms estimates of the latent-state variables given the initial values of the model parameters. Once the latent-state estimates are calculated, they are then passed to the maximum-log-likelihood estimator (MLE) as complete data. And 2) the maximization (M-step) where the LL that the estimated parameters are the true population parameters is maximized. These steps are repeated until the estimates converge. The LL function for the model can be expressed as

$$\ln \mathbf{L}\,(\beta_t | y, x) = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - (X_t\beta_t))^2 \qquad (10)$$

16

2.2. Out-of-sample procedure

The constant OLS uses the first 106 months—March 31st, 1986, to December 31st, 1994—as the in-sample period. Once the in-sample estimates are formed, they are fixed and used to estimate each out-of-sample forecast one-month ahead. The remaining 315 months are used to form one-step ahead out-of-sample forecasts for period—January 31st, 1995, to March 31st, 2021. The advantage to using OLS is that the estimators are the best linear unbiased estimators when certain assumptions hold such as:

1) Linear in parameters
2) Expected error term zero
3) Constant parameters
4) Normally distributed
5) Homoscedastic errors
6) Spherical disturbances

However, in the current study we assume coefficients are time-variant rendering OLS inappropriate. Another drawback to OLS is there is only one period the parameters are estimated on and this may not give an accurate representation if coefficients change.

Rolling Window OLS has the benefit of estimating parameters over a sliding window so that we now have a series of betas rather than a single forecast however, the structure of the data may not be captured because the lookback period remains fixed. RWOLS similar to OLS begins with an initial window size of 106 months, a quarter of the data set, to estimate the in-sample parameters. We chose a small window in order to have more forecasts and updating. The window is then rolled forward by one month while dropping the last month to maintain a fixed window size and are re-estimated. This method produces a time-series for both a_t and b_t, the two elements of beta vector β_t. However, the coefficients are held constant in each window, which is not truly consistent with TVP.

Expanding Window OLS EWOLS is also trained on the initial 106 months. The benefit to this method is also having a series of beta estimates by allowing the in-sample to expand to include new observations. However, this method does not drop the last observation and may have events in the distant past dampening the estimates that should represent the fluctuation in betas. In general, OLS methods are computationally expensive requiring memory of the in-sample period to forecast the next month-ahead. Once the initial forecast parameters are fixed and the one-month-ahead forecast is calculated, the window "expands" by one month. The same issue persists with this approach, that the estimated parameters are held constant in each window.

The KF AR(1) parameterization, Eq. (8) uses the first quarter of the sample— March 31st, 1986, to December 31st, 1994—to form the initial estimates using the EM, and then the remaining three-quarters—January 31st, 1995, to March 31st, 2021—to evaluate the model.

In the Kalman filtering procedure, the latent state variable is re-estimated each month and a new forecast is formed one-month-ahead Eq. (9). This means that at each time t, the KF algorithm is run to obtain forecasts using only the information available at time t. This yields the forecast of y_{t+1}. Then, at time $t + 1$, the KF algorithm is re-run to obtain the forecast using only the information available at time $t + 1$, which then yields the forecast of y_{t+2}; this procedure is continued until reaching the end of the sample (March 31st, 2021). The KF has the added benefit of a computationally light footprint. In the bivariate AR(1) specification the KF updates it's beta estimates with each iteration allowing for the capture of variation in the estimates especially if there is a sudden shift in parameters.

2.3. Measuring predictive performance

The following goodness-of-fit measures are evaluated for each modeling technique and specification.

2.3.1. Root mean square error

The root mean square error (RMSE) is the standard deviation of the squared residuals, which measure how far the estimate is from the observation. The RMSE is one of the best

indicators for measuring the how well the model predicts returns. A well-fit model will have small errors, indicating that the observed values are very close to the estimates. The RMSE will then also be small.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_t - \hat{y}_t)^2}{n}}$$

2.3.2. Coefficient of determination

R^2 is a measure of predictive accuracy. It reflects how much of the variation in the response variable is explained by the forecasting variable; it is calculated as the sum of squared residuals divided by the total sum of squares.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{i=1}^{n}(y_t - \bar{y}_t)^2}$$

2.3.3. Historical forecast average

$$\bar{y}_{t+1} = \frac{1}{t} \sum_{i=1}^{t} y_i$$

2.3.4. Out-of-sample R^2

R_{OS}^2 tests the null hypothesis that the forecasting variable has no predictive power. Where the historical average is $\bar{y}_t$, a positive value of R_{OS}^2 indicates that the forecast outperforms the historical average.

$$R_{OS}^2 = 1 - \frac{\sum_{t=T_1+1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=T_1+1}^{T}(y_t - \bar{y}_t)^2}$$

where $T_1 = $ initial window size.

2.3.5. Mean absolute error

The mean absolute error (MAE) is a common measure of forecast error in the time-series analysis. It represents the absolute value of the difference between the predicted and observed values.

$$MAE = \frac{\sum_{i=1}^{n}|y_t - \hat{y}_t|}{n} = \frac{\sum_{i=1}^{n}|e_t|}{n}$$

2.3.6. Diebold–Mariano test statistic

The Diebold-Mariano (dm) tests against the null hypothesis of equal predictive accuracy between models (Diebold & Mariano, 1995). and is the ratio of the sample mean to the standard error of the loss differential—the difference between the squared forecast errors generated by one model and the squared forecast errors generated by an alternative model.

Loss differential

$$d_t = L\left(e_{t+h|t}^1\right) - L\left(e_{t+h|t}^2\right)$$

Diebold-Mariano test statistic is

$$dm = \frac{\bar{d}}{\sqrt{\sigma_{\bar{d}}^2/n}}$$

3. Data

I obtain the monthly S&P 500 prices for the period, as reported on Yahoo Finance, March 31st, 1986, to March 31st, 2021, and use the adjusted closing prices which are already adjusted for dividends and splits, on the last business day of each month to calculate the return. A separate return time series is generated using one month-lagged S&P 500 returns as an independent variable. The first-differenced implied volatility index (ΔVIX) is calculated using end-of-month values as reported on the Chicago Board Options Exchange. We take the dividend-price ratio sourced from the Quandl database and difference the series once to correct for the

autocorrelation. The 3-month Treasury bill (^IRX) is the risk-free rate obtained as reported on Yahoo Finance, using the closing prices on the last business day of the month. The series is differenced for autocorrelation. Table 1 (Appendix A) provides a description of the variables and Table 2 describes the summary statistics of each independent and response variable.

4.   Results

Table 3 summarizes regression results. The modeling techniques' performance and statistics are partitioned into 3 sections, one for each independent variable. Similar to Mergner & Bulla (2005) & Faff, Hillier & Hillier (2000) the Kalman filter outperforms, on average, the alternative TVP modeling techniques in terms of RMSE, MAE and both the in-sample and out-of-sample R^2.

The top four performing specifications listed in order of rank are 1) KF bivariate AR(1) using ΔVIX, 2) OLS using lagged S&P returns, 3) EWOLS using lagged S&P returns, and 4) KF bivariate AR(1) using lagged S&P returns. Lagged S&P returns was the best predictor in terms of the in-sample and out-of-sample R^2. Surprisingly, OLS using lagged returns as the forecasting variable outperformed the KF when also specified using lagged returns. This suggests the KF is not always the best model. The outperformance of OLS to ROLS and EWOLS points to beta being close to stationary where OLS estimates are BLUE.

The point estimate of each technique's slope coefficient was calculated by taking the arithmetic average of each beta series. The first five figures in each panel reports out-of-sample goodness-of-fit measures. The RMSE is the decision criterion which determines the model that best fit the data. The following rows of statistics measure significance and test the null hypothesis that beta is not significantly different from zero and the results are not replicable.

The modeling technique with the smallest RMSE and superior performance measures is the KF bivariate AR(1) at 3.76% using the ΔVIX as the forecasting variable. The worst performing technique using the same predictor was ROLS, with a high 3.88%.

The KF technique boasts the highest R^2 for all three independent variables. The highest OSR^2 2.62% the KF ΔVIX, OLS using lagged S&P 500 returns was second.

The lagged S&P 500 return exhibited greater explanatory power than using the ΔVIX or D/P. All betas were significant at the 1% level for lagged S&P. The KF estimates using each independent variable were significant at the 1% level in support of the decision to reject the null hypotheses.

We use the Diebold-Mariano (1995) (dm) test statistic to reject the null hypothesis of equal predictive accuracy. dm for OLS LagSP and KF AR(1) ΔVIX dm = -2.7033. EWOLS using lagged S&P and KF AR(1) lagged S&P dm = -0.1897 p-Value= 0.424.

5.   Market timing strategy

A market timing strategy is governed by the ability to forecast returns with some degree of certainty. The benchmark strategy Buy and Hold holds only the market portfolio for the duration of the period January 31rst, 1995, to March 31rst, 2021 using monthly returns to forecast one-month ahead. The reason we use monthly data is to capture complete market cycles for the 35-year period and determine the best model. Later, in a pairs trading strategy we move to daily data to investigate the superior model in a different problem domain where we can have more trades than once a month and because the strategy relies on cointegrating relationships which may change, and we wanted to capitalize on as many trades as possible while the statistical properties held. We define the market portfolio as the S&P 500 index. The trading rules for the market timing strategy are straightforward following Campbell and Thompson (2008) and Rapach, et al.(2016). We consider a mean-variance investor who invests fully in the market portfolio if the forecasted return is larger than the 13-week Treasury bill (^IRX) and will invest fully in the risk-free rate when the forecasted return is less than the risk-free rate. Table 4 reports the market timing strategy results for each modeling technique using the Buy and Hold strategy as the benchmark. The KF dominated again outperforming each alternative technique using ΔVIX as the forecasting variable. The annualized time-weighted return is calculated as the

cumulative product of $(1 + r_i)^{\frac{12}{n}}$ where r_i is the monthly return for each of the 315 months out-of-sample. The KF has a steep annualized time-weighted return of 21.64%, nearly twice as large as the alternative techniques.

The volatility was the lowest 2.47% using the KF and ΔVIX set up and the second lowest was OLS using lagged S&P. The annualized Sharpe ratio (SR) for KF was exceptionally high at 2.08 with the second-best annualized SR only 0.90. Maximum drawdown (MDD) measures the size of the largest loss the portfolio experienced in the entire 26 years and 3 months period. The KF bivariate AR(1) performed best with the lowest maximum drawdown (MDD) of 15.80%. MDD months (MDDM) describes how long the portfolio was in decline from peak to trough. The KF using lagged S&P had the fastest recovery time of 34 months. The longest drawdown was 257 months for both ROLS specifications ΔVIX and D/P. The lowest dip occurred in month 165, observed September 31rst, 2009, the Great Recession. February March 2020, the beginning of the pandemic is also captured by a steep decline in each portfolio. It is interesting the Great Recession (see Figures 14 and 15) had a larger impact on market returns than the pandemic recession. This may be an area of interest for future research.

6.   Pairs trading strategy

Pairs trading is a form of statistical arbitrage that is a dollar neutral, where the value of each position equals one another, allowing traders to profit in more than one type of market due to a low positive correlation with the market (Nobrega & Oliveira, 2014). Pairs trading takes advantage of the mean-reverting relationship that exists between cointegrated securities. Exceptions to the usual relationship of a finite spread between the pair present arbitrage opportunities during the periods in which the series sufficiently depart from equilibrium. When the prices of the pair deviate sufficiently, the prices tend to revert to the historically observed mean price. By monitoring the standardized forecast errors an actionable trade signal is generated when the series deviates more than ± .5 sd. from the mean. The signal triggers an

IBALGO set to urgency, in which opposing dollar neutral positions are taken, i.e., go short the overperforming security and long the underperforming security. Financial instruments in pairs trading can include broad market indices, individual securities and/or baskets of securities[1].

The strategy is developed using historical daily closing prices courtesy of ActiveTick and implemented using the simulated live trading environment provided by Interactive Brokers Traders Workstation (IB/TWS). This approach allows us to simulate the impacts of liquidity, order types, margin requirements, transaction costs, regulatory constraints, (NBBO, REG SHO, Uptick Rule), slippage, SEC fees—collectively known as implementation shortfall.

A total of 676 securities yielding 228,150 distinct pairs of securities from ActiveTick, LLC trading on the New York Stock Exchange (NYSE) meeting the following trading criteria are evaluated using the daily OHLC bars.

- trading since 2018-01-01

- minimum tick size of .01 cents

- minimum close > $5.00

- average 20-day volume >= 1,500,000 shares per day.

Once the universe of potential instruments is identified the process of choosing which pair of securities to use involves five basic steps described below.

Step 1 - Engle-Granger test (EGT)

It is generally accepted that most economic time series variables are non-stationary, i.e. integrated to the order of one, I(1). That is to say that such series' properties tend change over time and trend away from any mean in the long term. However, there exists linear combinations of integrated series when differenced yield stationary I(0) series. The financial instrument pair

---

[1] Baskets of securities use the Johansen method as opposed to the Engle Granger method to find cointegrating relationships, I use the Engle-Granger method for a pair of individual securities.

we choose, series $y_{1,t}$ and $y_{2,t}$ when differenced should yield Δz_t (notation following MATLAB):

$y_{1,t} = X_{1,t}\beta_{2,t} + \xi_t$ is I(1)

$y_{2,t} = X_{2,t}\beta_{2,t} + \nu_t$ is I(1)

$\Delta z_t = y_{1,t} - y_{2,t}$ is I(0)

Because the results can vary based on which series is chosen to be the dependent series, the EGT is run twice alternating each series as the dependent series. The series configuration with the most significant Augmented Dickey-Fuller statistic is chosen.

Step 2 - Augmented Dicky-Fuller (ADF)

We reject the null hypothesis $\varphi = 1$ of the existence of a unit root in favor of the alternative hypothesis $\varphi < 1$ the series is covariance stationary.

$\Delta z_t = \alpha + \varphi \, \Delta z_{t-1} + \mu_t$

Step 3 – Hurst exponent[2]

The Hurst Exponent (H) reveals the extent to which a time series mean reverts (or not).

When

H > 0.5	the series is trending
H $= 0.5$	the series is white noise
H < 0.5	the series is mean reverting

[2] Tomaso Aste (2021). Generalized Hurst exponent (https://www.mathworks.com/matlabcentral/fileexchange/30076-generalized-hurst-exponent), MATLAB Central File Exchange. Retrieved September 12, 2021.

Step 4 – Residual analysis

A well-behaved residual series has these three characteristics:

1) Distributed normally
2) No serial-correlation
3) Homoscedastic

We use the Kolmogorov-Smirnov test for normality; the Ljung-Box Q-test for residual autocorrelation; and the Engle's ARCH test for residual heteroscedasticity. Table 5 describes the test results.

Step 5 – Sector/Industry

Although two financial instruments can pass the above four steps it is often reassuring when the pair are related. When the pair are in the same sector or industry external market forces are more likely to affect each the similarly.

After evaluation of the results of the above five steps candidate pairs can be selected for further evaluation.

Using daily closing prices in the pairs trading application in order to have more trades than monthly data would allow. Fidelity National Information Service, Inc. (NYSE:FIS), a finance and technology (fintech) company and Visa (NYSE:V), a financial services company, both listed on the New York Stock Exchange, are chosen for further analysis.

Since pair trading involves two financial instruments and positions in the pair involve both a long position in one instrument and a short position in the other[3]; we will refer to one instrument as the base instrument and the other as the hedge instrument. The position of the pair is based on the position of the base instrument. The base security is defined to be Visa and the hedge security is Fidelity National Information Service, Inc. based on the results of the ADF.

---

[3] Unless it is an inverse Exchange Traded Fund (ETF) for example.

We use a KF random walk (RW) parameterization described by Commandeur and Koopman (2007), Durbin and Koopman (2012) and initialize the system model hyperparameters following Kinlay (2020) we initialize matrix Q to 0.0001 and set R to 1. to develop a pairs trading strategy.

$$y_t = X_{t-1}\beta_{t-1} + \epsilon_t \qquad \sim i.i.d.\,(0, R) \qquad\qquad \text{Measurement equation (12)}$$

$$\beta_t = \beta_{t-1} + \omega_t \qquad \sim i.i.d.\,(0, Q) \qquad\qquad \text{State-transition equation (13)}$$

where the observed (1×1) scalar y_t represents the price of security VISA (V) at time t and the X_t (1×2) row vector represents the constant 1 and the price level for Fidelity National Information Services, Inc. (FIS) at time t. β_t is the (2×1) stochastic latent-state vector whose elements are the intercept, a_t and slope, b_t (hedge ratio) at time t. The hedge ratio represents the number of shares of the hedge security to trade for each share of the base security.

The pairs trading model was developed and tested on 515 days of daily data for the period January 2nd, 2020, to May 31rst, 2021. The trading rules are:

If Z(t) >= δ;   Close any open positions; Short base security; Buy hedge security

If Z(t) <= -δ;   Close any open positions; Buy base security; Short hedge security

If Z(t) > -δ and Z < δ; Maintain current positions – no trades

where:

δ= .5 standard deviations

Z = standardized residual from KF.

Although short positions involve borrowing shares we do not have, we must still have enough equity in the account to cover the shares. We needed a margin account in order to short sell but never borrowed on credit, that is to say we put up a dollar of equity per dollar of share

borrowed. This was so that the reader could easily translate their results based on their leverage needs. For example, Reg T requires the trader to put up a 150% for short sales whereas we maintained 200% of the value of shares that were short sold. That is to say we have unleveraged returns. We ran one pair only to keep the strategy simple and the small sample period of 82 days is due to the strategy being set up to communicate with traders workstation late beginning June 17th and each observation was taken one-day at a time whereas in the development of the strategy we were able to run a large sample of historical data in one day using simulated trades. The results are promising with an annualized unleveraged absolute return of 30.72% and an impressive Sharpe Ratio of 2.5. The percentage wining trades for the full period was 57%. Maximum drawdown was 4.97% for the period and the decline lasted 63 days.

One of the objectives of this thesis is to demonstrate using a state-space model to implement a trading strategy, that meaningful excess returns are achievable when accounting for trading inefficiencies. To run the strategy using Interactive Brokers (IB), the MATLAB code is converted to a Windows™ dynamic link library (dll). That strategy dll is hosted in a .Net c# console application which subscribes to real-time price updates from Active Tick™ and communicates with the Traders Workstation API to send orders, receive executions, maintain positions, and track profit and losses. The account configuration consists of

- o Initial account equity[4] - $49,425.32

- o Reg T Margin

- o TWS API enabled

- o Order Type – Urgent IBALGO

The model was implemented for 107 trading days (excluding weekends and holidays) during the period of June 17th, 2021, to November 16th, 2021. Table 6 describes the total profit

---

[4]Account minimum of $25,000 is required to avoid restrictions of a FINRA Pattern Day Trader (PDT) designation which restricts the number of day trades that can be made within five business days using the same account.

and loss for trades executed for entire 107-day period. Shown are the last two executed trades for the period. Table 7 Mark-to-market performance summary reports the profit and loss using the current market value of the asset. Table 8 reports the realized and unrealized performance of both opened and closed positions. Table 9 Net asset value (NAV) lists both asset classes; stocks and cash. The first total column is the beginning value of $49,425.32 and the long and short columns describe the total NAV of each position at the close of the previous business day for the period reported. The last total column is the portfolio's ending NAV of $53,933.60 for a realized profit of $4,508.28.

7.   Conclusion

In this thesis I compared the ability of alternative modeling techniques to forecast stock returns. I find the results are sensitive to the choice in modeling technique when using the same predictors. The Kalman filter estimate gives the smallest RMSE indicating that it is the superior technique compared with the least squares approaches in forecasting returns (Table 4). The independent variables, ΔVIX, lagged S&P 500 returns and D/P were not significant predictors of returns.

The Kalman filter's performance translated to real economic gains when applied to a pairs trading application in a robust simulated trading environment. The pairs trading strategy using the Kalman filter resulted in an annualized total return of 20.55% unleveraged. These results are scalable and explicitly account for implementation shortfall.

The KF is also better suited for real-time high frequency intraday trading than OLS methods because the KF has a very light computational footprint when compared to OLS.

Whereas I used daily time bars, future research may explore López de Prado's suggestion of using alternative data with the Kalman Filter to use information from tick, volume, or dollar bars.

The take-away is the Kalman Filter does improve forecasting ability and significant economic gains can be achieved as a result.

REFERENCES

Adrian, T., & Franzoni, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance*, *16*(4), 537-556. https://doi.org/10.1016/j.jempfin.2009.02.003

Campbell, J.Y., & Thompson, S.B. (2008). *Predicting excess stock returns out of sample: can anything beat the historical average?* Cambridge: Harvard https://citeseerx.ist.psu.edu/viewdoc/citations?doi=10.1.1.67.6604

Chiang, I. E., & Hughen, W. K. (2017). Do oil futures prices predict stock returns? *Journal of Banking & Finance*, *79*, 129-141. https://doi.org/10.1016/j.jbankfin.2017.02.012

Cochrane, J. H. (2008). The Dog That Did Not Bark: A Defense of Return Predictability. *Review of Financial Studies*, *21*(4), 1533–1575. https://doi.org/10.1093/rfs/hhm046

Commandeur, J.J., & Koopman, S.J. (2007). An introduction to state space time series analysis. OUP Oxford.

Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253. https://doi.org/10.2307/1392185

Durbin, J., & Koopman, S.J. (2012). Time series analysis by state space methods (2nd ed.). Oxford University Press

Faff, R. W., Hillier, D., & Hillier, J. (2000). Time varying beta risk: An analysis of alternative modelling techniques. *Journal of Business Finance Accounting*, *27*(5&6), 523-554. https://doi.org/10.1111/1468-5957.00324

Kandel, S., & Stambaugh, R. F. (1996). On the Predictability of Stock Returns: An Asset-Allocation Perspective. *The Journal of Finance*, *51*(2), 385–424. https://doi.org/10.1111/j.1540-6261.1996.tb02689.x

Lauri, L. (2014). *Algorithmic evaluation of Parameter Estimation for Hidden Markov Models in Finance*. (n.d.). Retrieved February 8, 2021, from https://www.math.kth.se/matstat/seminarier/reports/M-exjobb14/140214.pdf

Lopez de Prado, M. (2018). Advances in Financial Machine Learning. Hoboken: John Wiley & Sons, Inc.

Lettau M. & Ludvigson, S. (2001). "Consumption, Aggregate Wealth, and Expected Stock Returns," *Journal of Finance, American Finance Association*, vol. 56(3), pages 815-849, June. DOI: 10.1111/0022-1082.00347

Martinelli, R., & Rhoads, N. (2010). Predicting Market Data. *Journal of Stocks & Commodities*, V. 28, pp.44-47.

Menzly, L., Santos, T. and Veronesi, P (2004). Understanding Predictability. *Journal of Political Economy*, 112(1), pp.1-47.

Mergner, S., & Bulla, J. (2008). Time-varying beta risk of Pan-European industry portfolios: A comparison of alternative modeling techniques. *The European Journal of Finance*, *14*(8), 771-802. https://doi.org/10.1080/13518470802173396

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*(6), 47–60. https://doi.org/10.1109/79.543975

Nobrega, J. P., & Oliveira, A. L. (2014). A combination forecasting model using machine learning and Kalman filter for statistical arbitrage. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. https://doi.org/10.1109/smc.2014.6974093

Pesaran, M. H., & Timmermann, A. (1995). Predictability of stock returns: robustness and economic significance. The Journal of Finance. 1201-1228.

Rapach, D.E., Strauss, J.K., & Zhou, G. (2009). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies, 23(2), 821–862. http://www.jstor.org/stable/40468327*

Rapach, D.E., Ringgenberg, M.C., & Zhou, G. (2016) Short Interest and Aggregate Stock Returns. *Journal of Financial Economics (JFE). https://ssrn.com/abstract=2474930 or http://dx.doi.org/10.2139/ssrn.2474930*

Watson, P. K. (1983). Kalman filtering as an alternative to Ordinary Least Squares — Some theoretical considerations and empirical results. *Empirical Economics*, *8*(2), 71–85. https://doi.org/10.1007/bf01973191

Welch, I., & Goyal, A. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, *21*(4), 1455–1508. http://www.jstor.org/stable/40056859

Appendix A
Tables

Table 1
Variable description
This table reports monthly data provided from Yahoo Finance, Chicago Board Options Exchange, and Quandl for period March 31st, 1986, to March 31st, 2021. Securities VISA and Fidelity National Information Services, Inc. daily prices are sourced from ActiveTick for period January 2nd, 2020, to October 7th, 2021.

Symbol	Variable
S&P 500	Return on the S&P 500 index
ΔD/P	First differenced dividend-price ratio, lagged one-period.
ΔVIX	First differenced volatility index, lagged one-period.
Lag Return	Lagged return on the S&P 500 index, one-period.
FIS	Security price of Fidelity National Information Services, Inc.
V	Security price of Visa, Inc.

Table 2
Descriptive Statistics.

	Mean	Standard deviation	Skewness	Kurtosis	Min	Max
Lag return	0.01%	0.04%	-0.8931	4.0449	-0.20%	0.13%
ΔVIX	0.00%	0.05%	1.3455	9.7045	-0.19%	0.36%
Dividend/Price	-0.01%	0.09%	1.0838	9.4815	-0.45%	0.59%
FIS	140.54	9.4	-1.1925	1.8206	102.45	157.44
VISA	205.6	20.82	-0.2884	0.2202	135.74	250.93

Table 3

Comparison of OLS beta and point estimates of various beta series.

This table summarizes model performance measures for each technique and independent variable for the out-of-sample period January 31st, 1995, to March 31st, 2021. The mean slope coefficient is reported for each beta series for each modeling technique. Standard errors are reported in parenthesis. *** means that the slope coefficient is significant at the 1% level (**: 5%, *: 10%). In-sample period March 31st, 1986, to December 31st, 1994.

		OLS	RWOLS	EWOLS	KF
Delta VIX	RMSE	3.80%	3.88%	3.83%	3.76%
	MAE	2.76%	2.83%	2.78%	2.73%
	MSE	0.14%	0.15%	0.15%	0.14%
	$OOSR^2$	0.54%	-3.55%	-0.61%	2.62%
	SE	(0.214)	(0.219)	(0.222)	(0.247)
	$\bar{b}$	-0.028	0.004	-1.96	0.058
	t-stat	-12.91***	-0.02	-9.07***	27.40***
	p-value	0.E+00	7.E-02	0.E+00	0.E+00
Lagged S&P	RMSE	3.77%	3.79%	3.77%	3.78%
	MAE	2.71%	2.77%	2.72%	2.72%
	MSE	0.14%	0.14%	0.14%	0.14%
	$OOSR^2$	2.12%	1.50%	2.35%	1.65%
	SE	(0.213)	(0.213)	(0.212)	(0.212)
	$\bar{b}$	0.320	0.202	0.261	0.216
	t-stat	150.60***	96.34***	124.19***	102.33***
	p-value	0.E+00	0.E+00	0.E+00	0.E+00
Dividend/Price	RMSE	3.81%	3.79%	3.82%	3.81%
	MAE	2.76%	2.77%	2.78%	2.76%
	MSE	0.15%	0.14%	0.15%	0.15%
	$OOSR^2$	0.42%	1.50%	-0.41%	-0.11%
	SE	(0.214)	(0.213)	(0.215)	(0.215)
	$\bar{b}$	0.000	0.010	-0.003	0.017
	t-stat	-1.E-03	4.74***	0.0137	8.06***
	p-value	9.E-01	3.E-05	2.E-01	0.E+00
	No. Obs.	315	315	315	315

Table 4

Market timing strategy performance summary.

This table reports the market timing strategy results for each modeling technique for period January 31, 1995, to March 31, 2021. Maximum drawdown (MaxDD) is defined as the maximum observed loss (peak to trough) the investment experiences during the 26.25 year period. Maximum drawdown months (MaxDDM) is the maximum number of months the investment has seen between peaks. This table additionally restates each modeling technique's goodness-of-fit measures.

Modeling Technique	Variable	Strategy performance								
		Anlzd. TWR	Vol. σ	Sharpe Ratio	Anlzd. SR	Max DD	Max DDM	RMSE	MAE	$OOSR^2$
KF^{AR}	ΔVIX	21.64%	2.47%	0.60	2.08	9.46%	3	3.76%	2.73%	2.62%
	Lag S&P	10.20%	3.13%	0.22	0.76	22.62%	38	3.78%	2.72%	1.65%
	D/P	8.58%	3.67%	0.16	0.55	11.41%	17	3.79%	2.77%	0.87%
$EWOLS$	ΔVIX	8.50%	3.81%	0.15	0.52	46.04%	17	3.83%	2.78%	-0.61%
	Lag S&P	10.13%	2.71%	0.24	0.83	23.34%	18	3.77%	2.72%	2.35%
	D/P	8.50%	3.81%	0.15	0.52	46.04%	17	3.82%	2.78%	-0.41%
$ROLS$	ΔVIX	7.63%	3.26%	0.15	0.52	24.94%	30	3.88%	2.83%	-3.55%
	Lag S&P	8.51%	3.12%	0.18	0.62	31.38%	30	3.79%	2.77%	1.50%
	D/P	8.14%	3.40%	0.16	0.55	34.59%	52	3.79%	2.77%	1.50%
OLS	ΔVIX	8.35%	3.81%	0.15	0.52	42.22%	17	3.80%	2.76%	0.54%
	Lag S&P	10.37%	2.63%	0.26	0.90	20.44%	8	3.77%	2.71%	2.12%
	D/P	8.47%	3.81%	0.15	0.52	46.55%	17	3.81%	2.76%	0.42%
Buy & Hold Strategy		7.94%	3.81%	0.15	0.52	38.72%	110	-	-	-
No. obs.		315	315	315	315	315	315	315	315	315

Table 5

Diagnostics.

This table reports residual diagnostic tests for pairs data period June 17th, 2021, to October 7th, 2021.

Test	H_0	Exponent	Test Stat	CV	p-Val	Decision
Kolmogorov–Smirnov	Normality	-	0.07	0.072	0.05	Accept
Ljung-Box Q-Test	No serial correlation	-	37.97	31.41	0.0089	Accept
ADF	Non-stationarity	-	-21.36	-1.94	0.0001	Reject
Engle ARCH	Homoscedastic	-	0.165	3.84	0.685	Accept
Hurst Exponent	Mean-reverting	-0.05	-	-	-	Accept

Table 6

Trade summary.

This table reports the total realized profit and loss for all closed trades during the period June 17th, 2021, to October 7th, 2021 and displays the last two trades of reporting period.

SYMBOL	DATE/TIME	QUANTITY	T. PRICE	C. PRICE	PROCEEDS	COMM/FEE	BASIS	REALIZED P/L	MTM P/L
Stocks									
USD									
FIS	2021-10-04, 15:59:50	-372.00	121.05	121.03	45,030.60	-2.13	-35,750.77	-405.98	7.44
FIS	2021-10-07, 15:59:50	374.00	121.30	121.30	-45,366.20	-1.87	31,369.89	-48.39	0.00
Total FIS		**189.00**			**-18,667.12**	**-55.94**	**-45,557.81**	**4,196.33**	**-332.93**
V	2021-10-04, 15:59:50	200.00	224.78	224.73	-44,956.00	-1.00	45,510.87	553.87	-10.00
V	2021-10-07, 15:59:50	-200.00	230.43	230.55	46,086.00	-1.26	-45,520.87	563.87	-24.00
Total V		**-100.00**			**21,166.00**	**-40.67**	**-116,139.00**	**-1,923.07**	**73.00**
Total					**2,498.88**	**-96.61**	**-161,696.81**	**2,273.25**	**-259.93**

Table 7

Mark-to-Market Performance Summary.

This table reports Mark-to-Market (MTM) profit and loss which shows how much profit or loss was realized for the statement period June 17th, 2021, to October 7th, 2021 regardless of whether positions are opened or closed.

	QUANTITY		PRICE		MARK-TO-MARKET P/L				
SYMBOL	PRIOR	CURRENT	PRIOR	CURRENT	POSITION	TRANSACTION	COMMISSIONS	OTHER	TOTAL
Stocks									
FIS	0	189	--	121.3	4,591.51	-332.93	-62.91	-68.64	4,127.03
V	0	-100	--	230.55	-1,962.00	73	-46.2	32	-1,903.20
Total Stocks					2,629.51	-259.93	-109.11	-36.64	2,223.83
Total (All Assets)					2,629.51	-259.93	-109.11	-36.64	2,223.83
Broker Interest Paid and Received									-2.38
Total P/L for Statement Period									2,221.45

Table 8
Realized & Unrealized Performance Summary.
This table reports profit and loss by asset class and underlying.

SYMBOL	COST ADJ.	REALIZED					UNREALIZED					TOTAL
		S/T PROFIT	S/T LOSS	L/T PROFIT	L/T LOSS	TOTAL	S/T PROFIT	S/T LOSS	L/T PROFIT	L/T LOSS	TOTAL	
Stocks												
FIS	0	8,293.53	-4,097.20	0	0	4,196.33	0	-0.66	0	0	-0.66	4,195.67
V	0	3,872.02	-5,795.10	0	0	-1,923.07	0	-12.13	0	0	-12.13	-1,935.20
Total Stocks	0	12,165.55	-9,892.30	0	0	2,273.25	0	-12.79	0	0	-12.79	2,260.47

Table 9
Net Asset Value (NAV) summary.
This table reports the starting and ending net asset value
for the period June 17th, 2021, to November 16th, 2021.

CHANGE IN NAV	TOTAL
Starting Value	49,425.32
Mark-to-Market	4,664.96
Dividends	0.86
Interest	-2.92
Commissions	-154.62
Ending Value	53,933.60
Annualized Total Return	20.55%

Table 10

General State-Space Model Notation.

Name	Description	Dimension	Traditional	KF evaluation & pairs
Observation (measurement)	Measure of true state	(m x 1) column vector	z_t	y_t
True state	True value of latent-state variable	(n x 1) column vector	x_t	β_t
Observation matrix	Maps current state into observed space	(m x n) matrix	H	x_t
Observation noise	Innovations	(m x 1) column vector	v_t	ϵ_t
State-transition matrix	Transition measurements applied to previous state	(n x n) matrix	A_t	I
Previous estimate	The estimate preceding the current	(1 x 1) matrix	x_{t-1}	β_{t-1}
Control-input matrix	Up dates measurements in control vector	(m x n) matrix	B_t	Intentionally left out
Control vector	Used in control engineering	(m x 1) column vector	u_t	Intentionally left out
Process noise	Variance/covariance matrix	(n x 1) column	w_t	ω_t

Appendix B

Kalman Filter Algorithm

The KF was designed by Rudolf Kalman (1960s) to track a moving target. Since its introduction, the KF has found widespread applicability to other domains as well, such as econometrics. It is well equipped to handle multiple dimensions in both the state and observation matrices and is computationally efficient because of its predictor-corrector nature (recursive calculation) scheme and its light computational footprint. The KF uses a set of equations to iteratively measure successive observations with increasing accuracy by using only the previous and current estimates of the latent-state variable(s) and variances. One of the benefits of its calculations being recursive is the requirement of lesser memory and fewer floating-point operations, which ultimately increases the speed, which is crucial for today's high-frequency trading environment. The KF quickly "filters out the noise" (Martinelli & Rhoads, 2010) and converges to the better value under investigation by reducing the errors in the estimate and those in the measurement and then uses KG to weight the errors based on their marginal contribution to the estimated error (Van Biezen, 2015). Generally, the application of OLS linear regression to such a problem requires a larger dataset in combination with a moving window. Thus, the KF is ideal for high-frequency applications.

Algorithm

With the KF's first data input (measurement) introduced into the system, the following three steps are performed iteratively as each new observation arrives following Van Biezen (2015).

Step 1. $$KG = \frac{Error\ in\ the\ estimate}{Error\ in\ the\ estimate + error\ in\ the\ measurement}$$

Calculate the KG. The KG places more or less certainty on both the measurement and the previously calculated estimate (in the first iteration, this is the original estimate value) by acting as a weight. As the scaling factor, KG is a number ranging between 0 and 1 and decides how much of each value to take into the next estimation. Both the error in the estimate (original for the first iteration) and that in the measurement are used for computing the KG. The KG is the

weight that determines which value, either the current measurement or previous estimate, has greater weight when calculating the next step.

Step 2.
$$EST_t = EST_{t-1} + KG[MEA - EST_{t-1}]$$

The main step, updating the current estimate, takes the previous estimate (in the first iteration, this is the original estimated value) and summing adds with the weighted difference between new measurement and the previous estimate. It is in this step that the KG plays a leading role. According to step one's calculation, if the errors in the estimates are large, this means that there is considerable uncertainty in the previous estimate and the KG will be large. Ideally, the value containing the larger error has a smaller impact moving to the fresh estimate. In this scenario, more weight is placed on the observation as its errors are relatively small. The KG when multiplied by the difference between the measurement and previous estimate retains a larger portion of the measured value when feeding into the current estimate's calculation (Commandeur & Koopman, 2007).

Step 3.
$$E_{EST_t} = \frac{(E_{MEA}) \cdot (E_{MEA_{t-1}})}{(E_{MEA}) + (E_{MEA_{t-1}})}$$

$$E_{est_t} = [1 - KG]\left(E_{est_{t-1}}\right)$$

Calculate the new error in the estimate. The error in the estimate is updated by multiplying the previous error in the estimate with the difference between one and the KG. This means that if the previous error in the estimate exceeds that in the data, the KG will be large. Subtracting from one will yield a small value. Multiplication by the previous error in the estimate quickly reduces the error in the estimate for the next iteration. If the KG is small, this indicates that the measurement errors are large. In this case, we do not want the new estimates to be heavily influenced by the incoming observations. Therefore, the error in the estimate decreases at a slower rate.
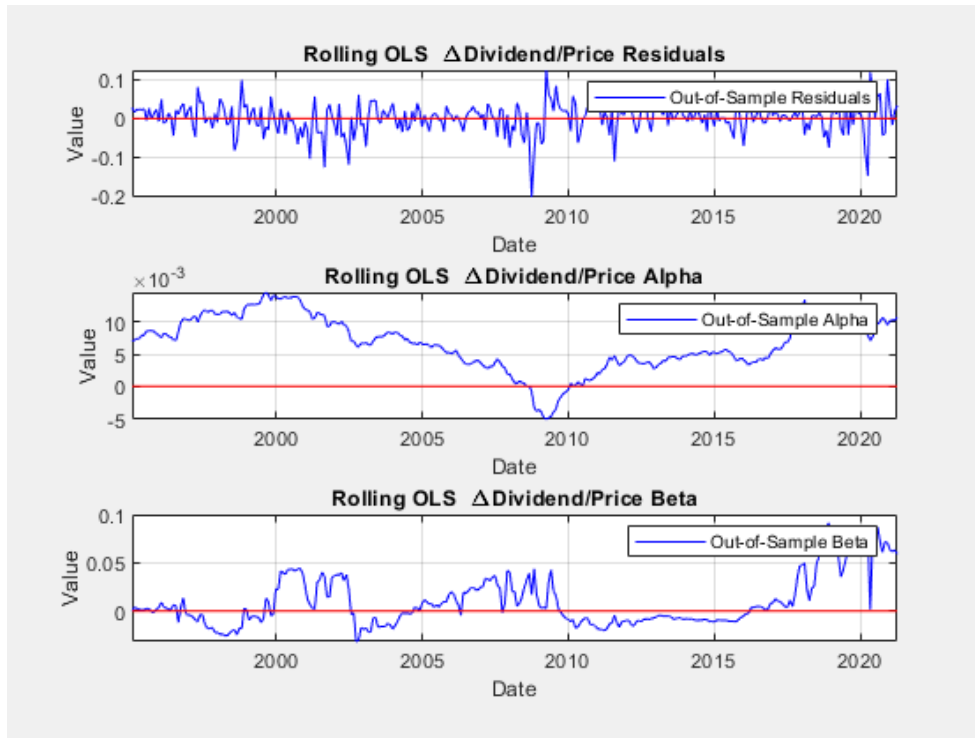
Appendix C
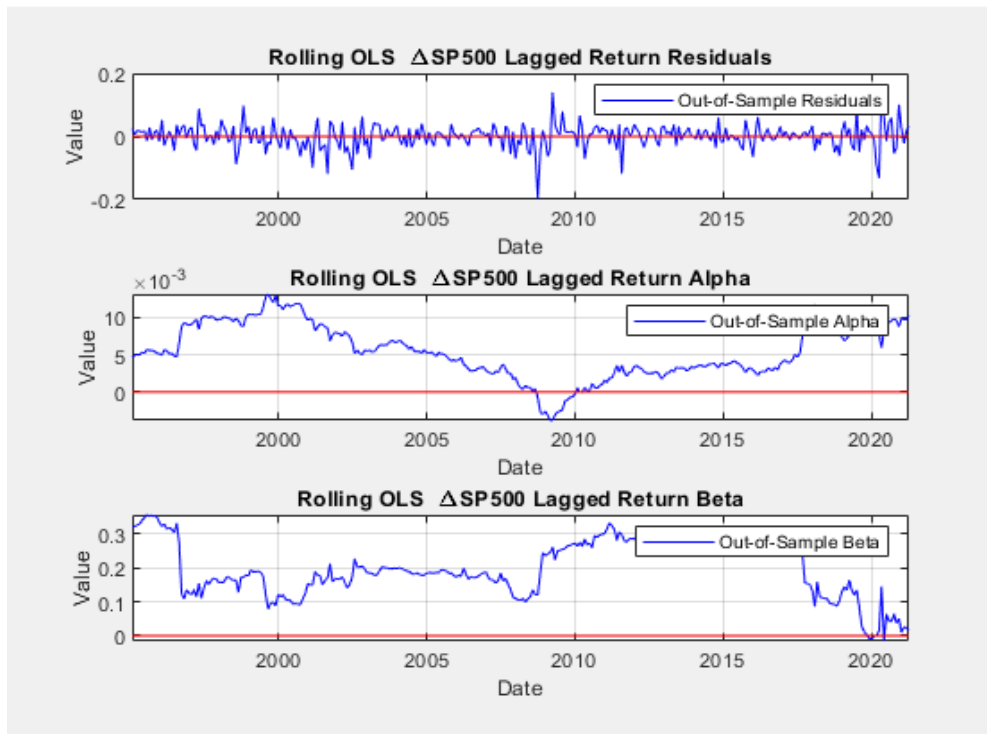
Figures

Figure 1

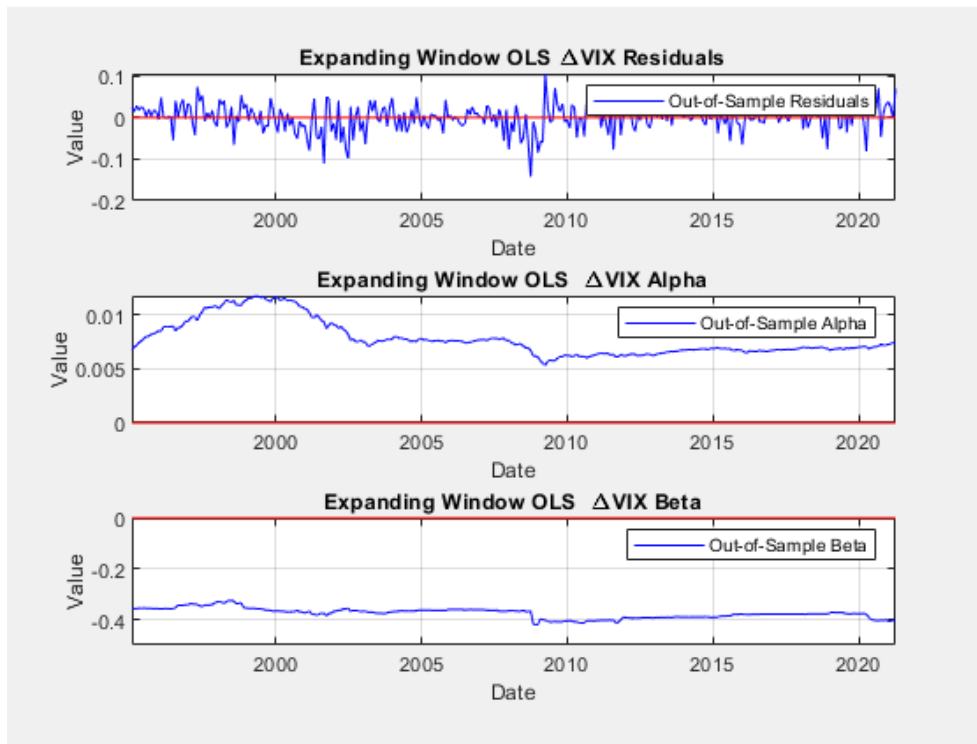Rolling Window OLS Delta VIX Estimates and Residuals
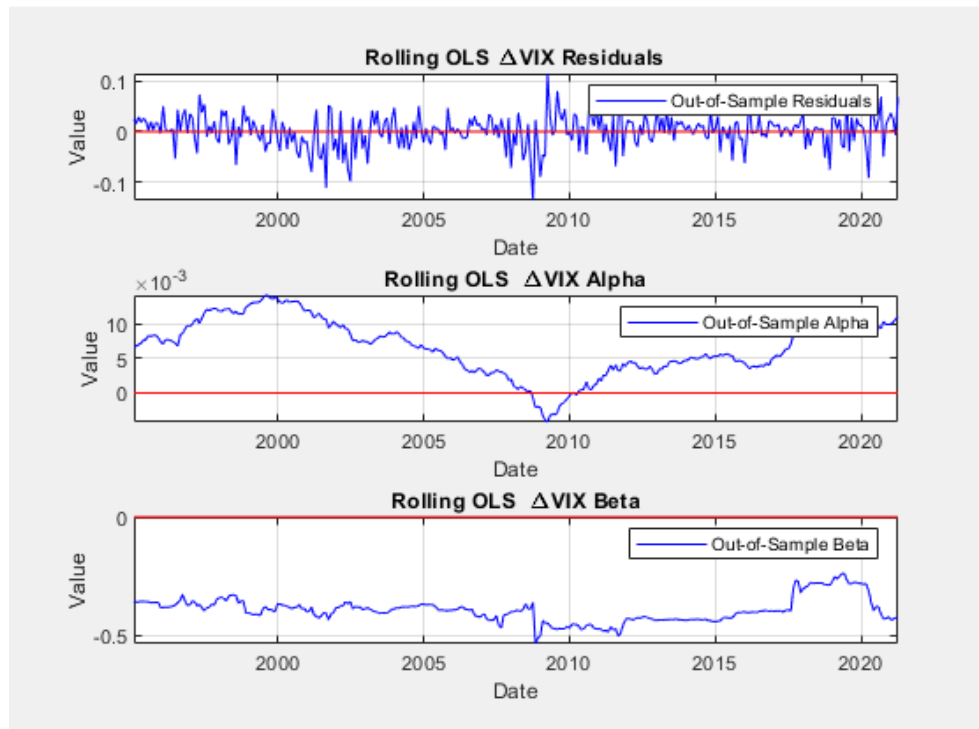
Figure 2
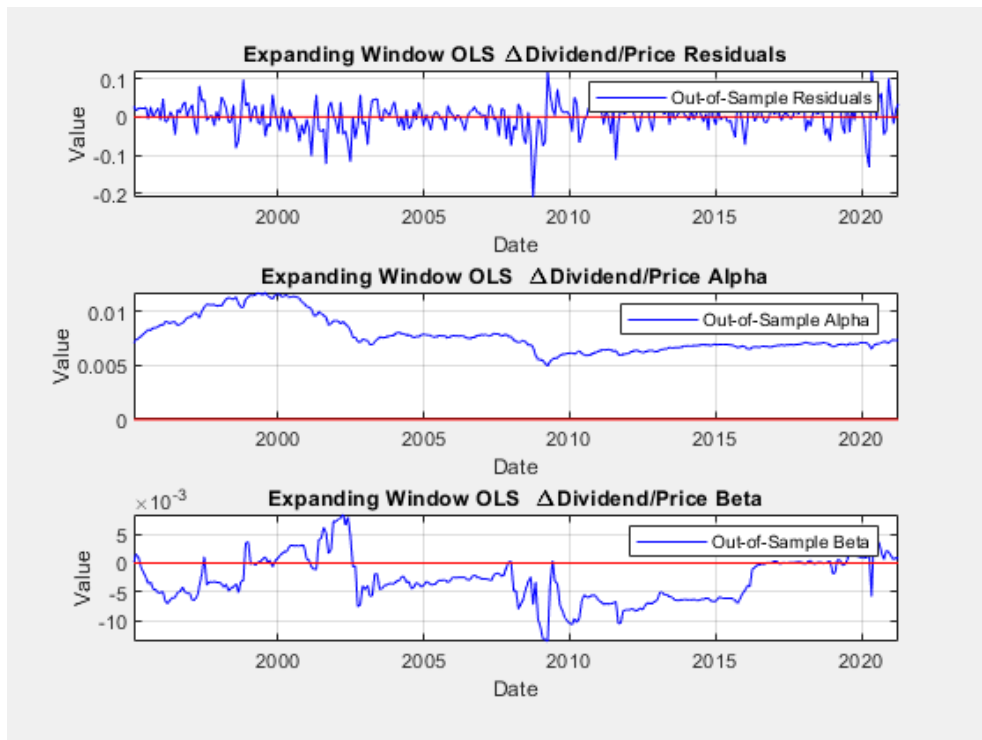Rolling Window OLS D/P Estimates and Residuals

Figure 3
Rolling Window OLS Lagged S&P Returns Estimates and Residuals

Figure 4
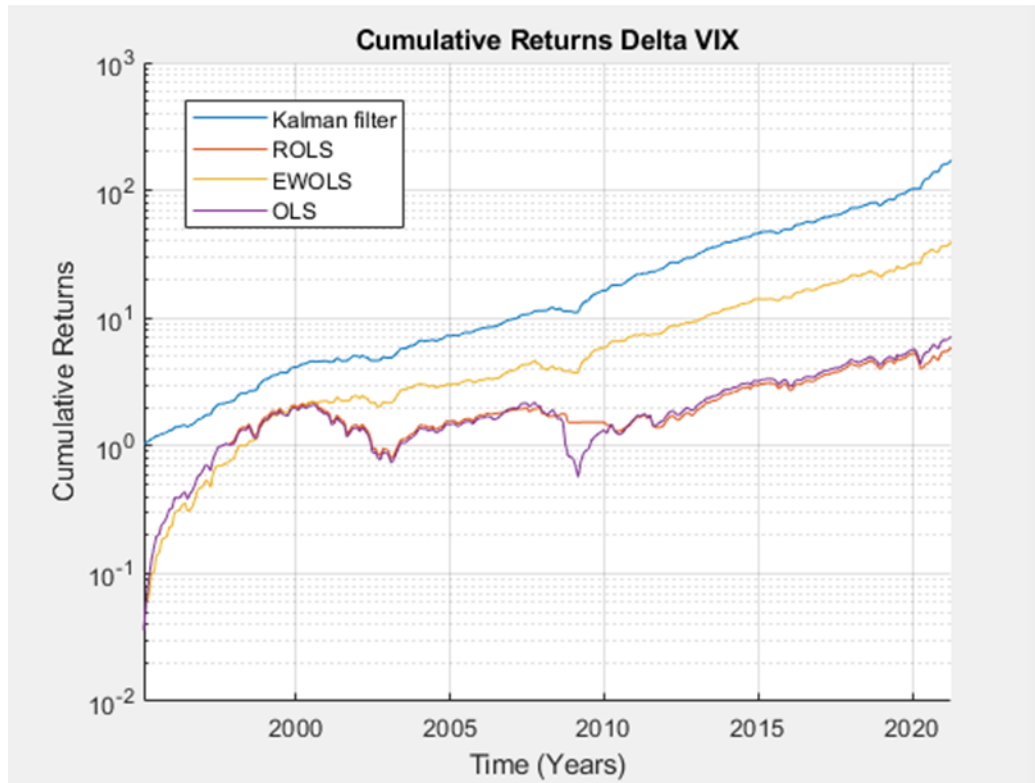Expanding Window OLS Delta VIX Estimates and Residuals

Figure 5
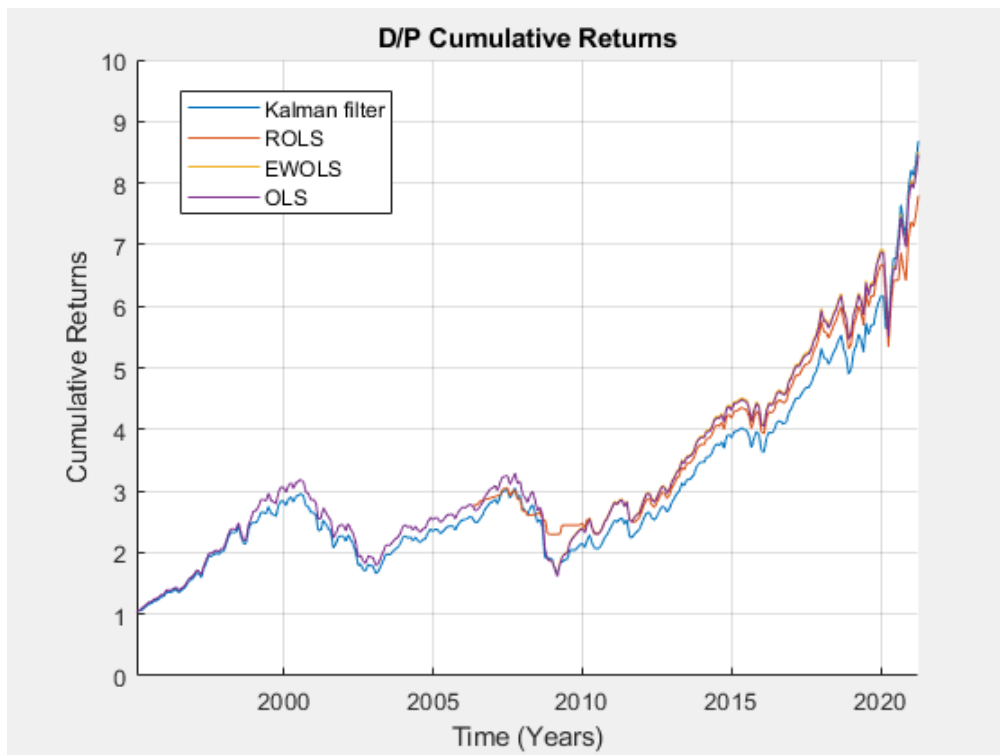Expanding Window OLS D/P Estimates and Residuals

Figure 6
EWOLS using Lagged S&P
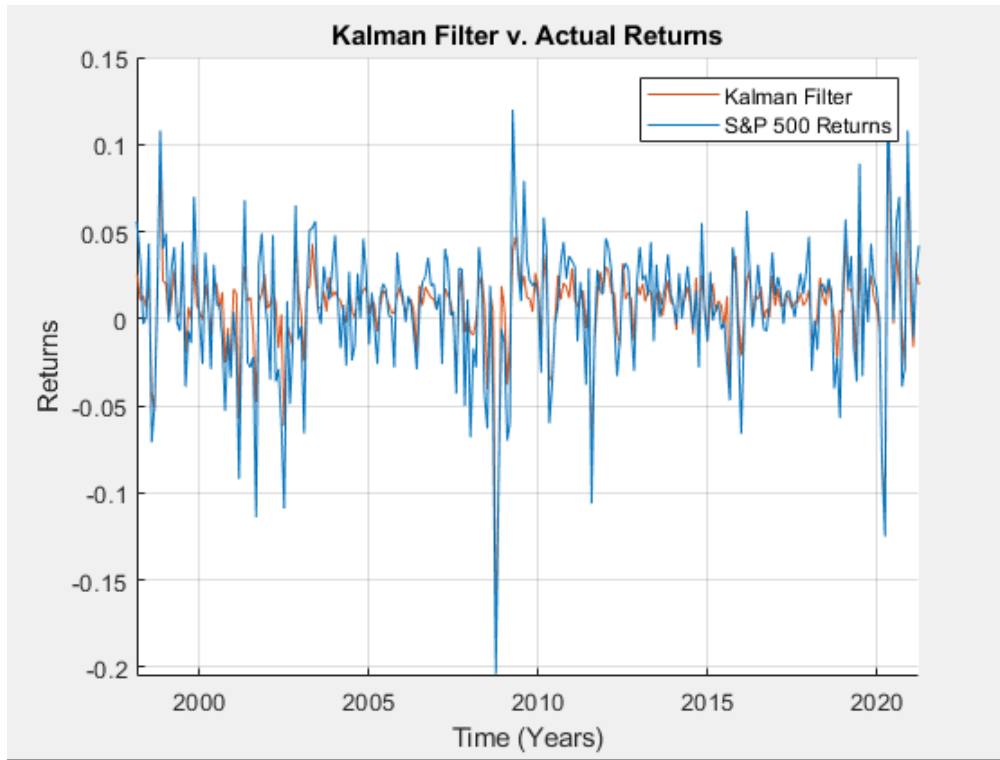

Figure 7
Cumulative returns for market timing strategy; each technique using Delta VIX

Figure 8
Cumulative returns for market timing strategy; each technique using D/P

Figure 9
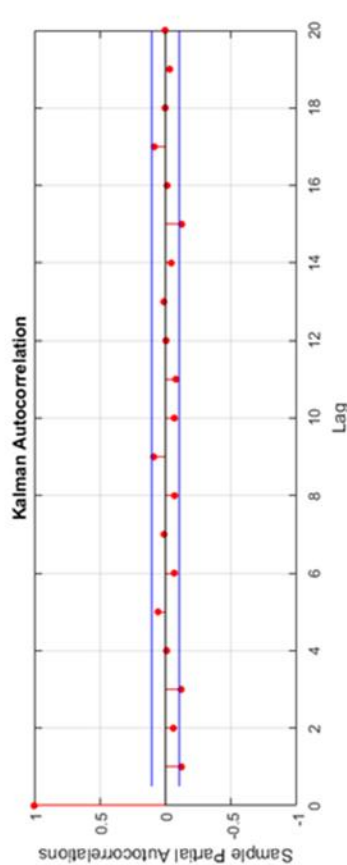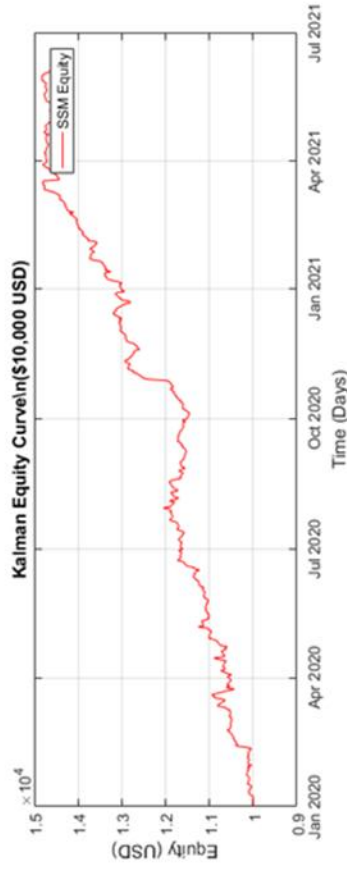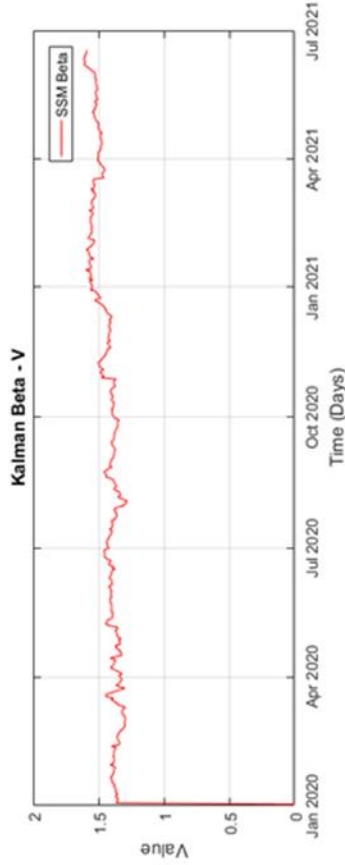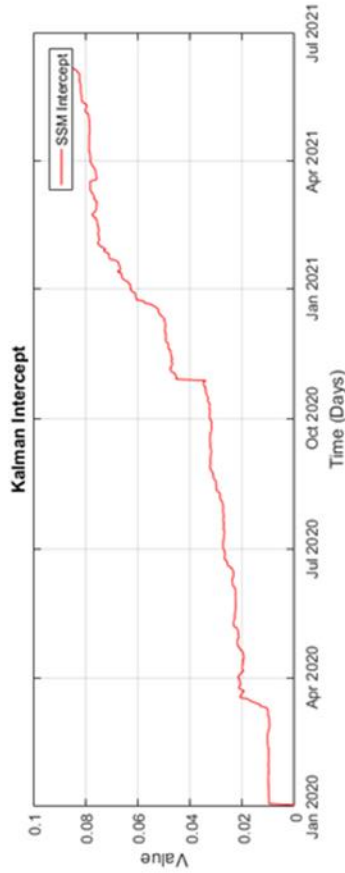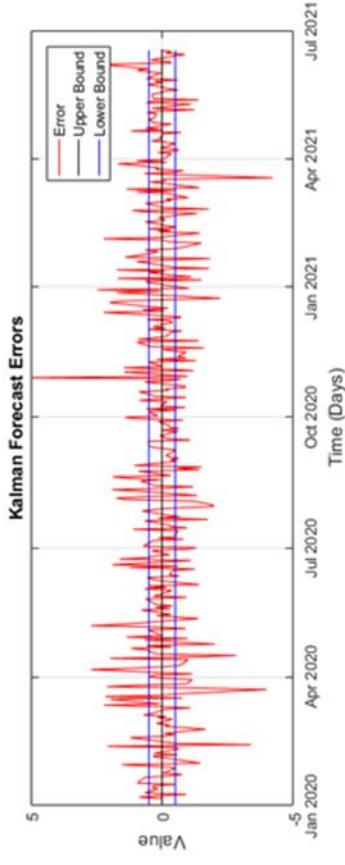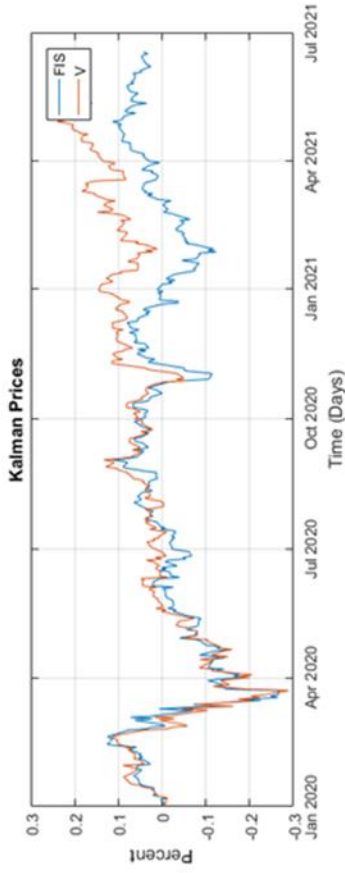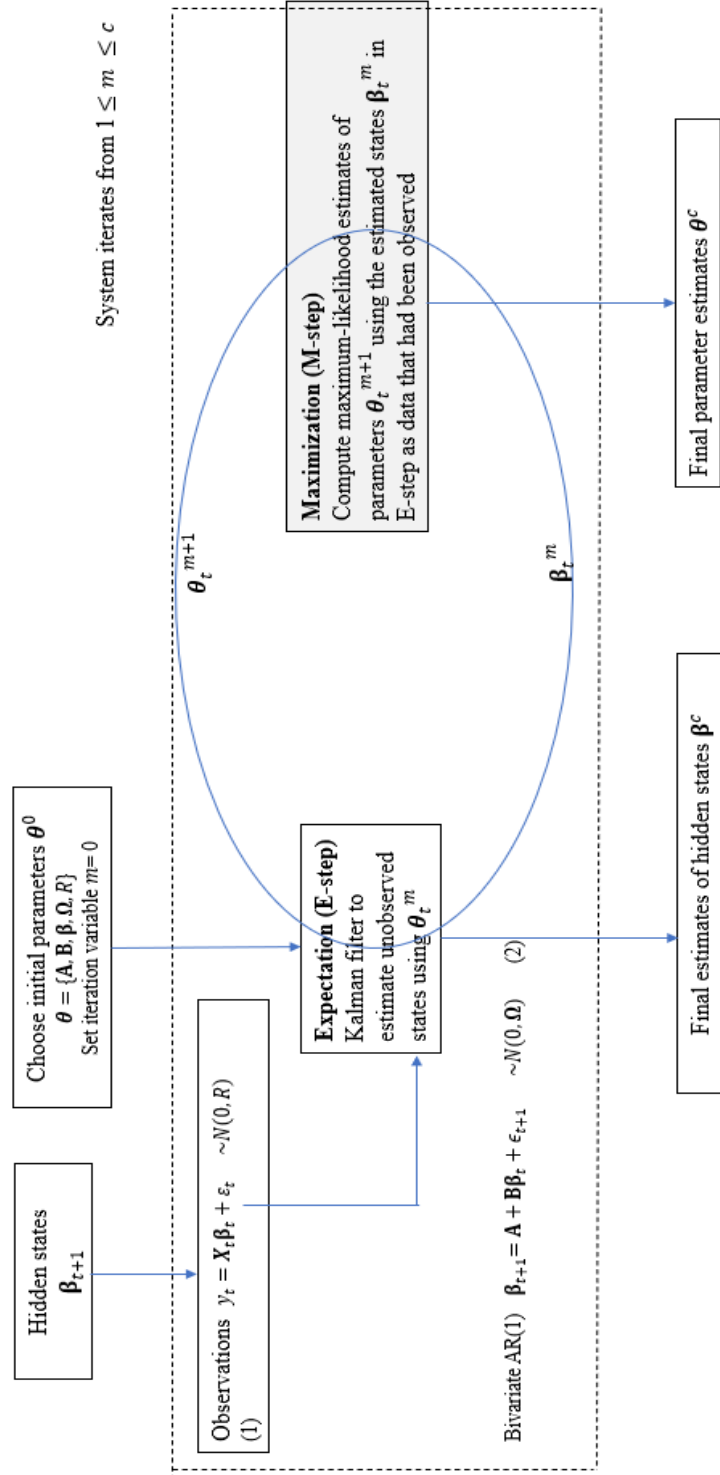KF forecasted returns series plotted against actual S&P 500 returns.

Figure 10
EWOLS and ROLS forecasted return series plotted against actual market returns.

Expectation-Maximization (EM) algorithm

Hidden states β_{t+1}

Choose initial parameters θ^0
$\theta = \{A, B, \beta, \Omega, R\}$
Set iteration variable $m=0$

Observations $y_t = X_t \beta_t + \varepsilon_t$ $\quad \sim N(0,R)$ (1)

Expectation (E-step)
Kalman filter to estimate unobserved states using θ_t^m

Maximization (M-step)
Compute maximum-likelihood estimates of parameters θ_t^{m+1} using the estimated states β_t^m in E-step as data that had been observed

Bivariate AR(1) $\quad \beta_{t+1} = A + B\beta_t + \varepsilon_{t+1}$ $\quad \sim N(0,\Omega)$ (2)

θ_t^{m+1}

β_t^m

System iterates from $1 \leq m \leq c$

Final parameter estimates θ^c

Final estimates of hidden states β^c

Following (Mader et al., 2011) and (Moon, 1995)