

Market Inefficiency: Pairs Trading with the Kalman Filter

Heather E. Dempsey, Sacred Heart University

December 8th, 2017

www.hedempsey.com

Abstract

Pairs trading is a form of statistical arbitrage that seeks to profit by exploiting the mean reverting relationship between pairs of securities. Pairs trading came to the forefront in the mid-80s and over the years has achieved widespread acceptance. As with any trading strategy, the more ubiquitous it becomes the less likely it is to remain profitable. The motivation for this paper is to explore whether there exists a niche where smaller, individual investors can profit from trading pairs of securities, specifically Exchange Traded Funds/Notes (ETF/ETNs)

This paper examines the efficacy of 1) using a Kalman Filter (KF) algorithm to estimate the optimal hedge ratio for Traded Funds (ETF) and Exchange Traded; 2) use the iterative results of the algorithm to provide trading signals; and 3) use the trading signals to develop a pairs trading strategy to answer the question as to whether small, individual investors can still profit from such a scheme.

Keywords: Pairs trading, Kalman Filter, Statistical arbitrage, Arbitrage Pricing Theory (APT)

JEL codes: C6 C15 C32 C88 G17

I. Introduction

Pairs trading is a form of statistical arbitrage that seeks to profit by exploiting the mean reverting relationship between pairs of securities. The pairs trading strategy has been around since the mid-80s and over the years has achieved widespread acceptance. As with any trading strategy, the more ubiquitous it becomes the less likely it is to remain profitable. The motivation for this paper is to explore whether smaller, individual investors¹ can profit from trading pairs of securities such as liquid Exchange Traded Funds (ETF) and Exchange Traded Notes (ETN).

This paper examines the efficacy of 1) using a Kalman Filter (KF) algorithm to estimate the optimal hedge ratio; 2) use the iterative results of the algorithm to provide trading signals; and 3) use the trading signals to develop a pairs trading strategy to answer the question as to whether small, individual investors can still profit from such a scheme.

¹ Small, individual investors for purpose herein will be investors with investment portfolios less than \$100,000.

II. Literature review

There is an increasing interest in the use of the KF or state-space models, and similar class of algorithms, because of the highly dynamic nature of the relationship between securities. The KF was first designed by Rudolf Kalman (1960's) to track a moving target. Since its introduction, the KF has found widespread applicability to other domains as well. For example, the KF is currently used in ballistic missile trajectory estimations, radar technology, global positioning systems (GPS), satellite tracking, robotics and econometrics. The KF is well equipped to handle multi-dimensions in both the state and observation matrices and is computationally efficient due to its recursive calculation scheme. The KF uses a set of equations to iteratively measure successive observations with increasing accuracy using only the previous estimate and current estimates of the latent variable(s) and variances. One of the benefits of its calculations being recursive is less memory and floating point operations are required, which ultimately increases speed, crucial to today's high frequency trading (HFT) environment. The KF quickly "filters out the noise" (Martinelli & Rhoads, 2010) and narrows in on the true value under investigation (in our case the hedge-ratio) by reducing the errors in the estimate and in the measurement and then using the Kalman Gain (KG) to weight the errors based on their marginal contribution to the estimate error (Van Biezen, 2015). Generally the application of ordinary least squares (OLS) linear regression to this type of problem requires a larger data set usually in combination with a moving window. Thus, the KF is ideal for real-time applications.

Engle and Granger (1987) in their seminal work studying pairs trading used cointegration and vector error correction (VEC) models to examine pairs of related equities. Often, the spread between equities were found to depart from equilibrium temporarily. A trading strategy which would short the high-priced (overvalued) stock and go long on the low-priced (undervalued) stock until market forces pushed the pair into equilibrium or would mean-revert. Johansen and Juselius (1990) extended their work to include baskets of cointegrated equities in Finnish and Dutch market data. The estimation was conducted using vector auto regression (VAR) and used a maximum likelihood estimator (MLE) with a Wald test for hypotheses on alpha and beta.

Vidyamurthy (2004) describes the best way to pick the pairs of cointegrated equities. In his book, a variety of methods are employed, including arbitrage pricing theory (APT) and cointegration models. He found proof the KF is optimal when the state-space and observation equations are linear and the noise follows a Gaussian distribution (Vidyamurthy, 2004). The algorithm is linearized in most cases but there does exist several extensions of the KF. For example, the Extended Kalman Filter (EKF) and Unscented KF, both handle non-linearities and the latter, non-Gaussian distributions as well.

The underlying mathematical model used in KF, is rooted in L.E. Baum's (1972) Hidden Markov Model (HMM) and Baum-Welch algorithm which lay the foundation for state-transition and measurement equations when estimating unobservable or latent-state variables. Elliot et al. (2005) focused solely on the state-space models including the KF, and mean-reverting Markov Chain Model to "monitor spreads, profiting from investment decisions based on this spread"

Elliot et al. found it useful for market data—in this case dollar-neutral portfolios, hedge funds, and trading. The idea is when the spread widens to a predetermined degree, to then short the higher-valued security, and go long on the lower-priced security. This strategy has indeed been profitable. However, speed is a crucial element to not only identify an adequate deviation from the mean, but also enter a position before equilibrium is restored and the opportunity lost.

Gatev et al. (2006) used 40 years of daily US market data picked his pairs by minimizing the variance of two daily-price series once each was normalized. The strategy entails taking a long and short position when the prices between a pair deviate by more than plus or minus 2 standard deviations in anticipation of mean-reversion (De Moura, Pizzinga, & Zubelli, 2016). The current study follows these specifications, except for the errors which do not exceed two standard deviations enough to have many trades, I restricted the limit to one standard deviation.

Montana et al. (2009) explored the use of the KF in market making.

III. Pair Trading Strategy Example

Pairs trading as a strategy is straightforward. When a pair of securities is believed to have a mean reverting relationship that is subject to temporary deviations. It is at these times of sufficient deviations the trading signals are generated. Sufficient deviation in this study is defined by 1 to 2 \pm sd. Refer to Figure 1.



Figure 1. Cointegrated pair of securities. Morgan Stanley (MS) and Freeport-McMoRan, Inc (FCX)

The pair illustrated in Figure 1. is an example of an cointegrated relationship between securities. Shown are two observed price series which tend to move together around a shared mean. The exception, and arbitrage opportunity, lie in the periods where the series depart from this equilibrium. Cointegrated relationships are characterized by mean-reverting behavior. This

means the pair will return to equilibrium no matter how far they stray. With this knowledge, the individual investor can profit. By entering a long position on the underperforming security and simultaneously shorting the overpriced security. In calculated anticipation, the spread will indeed converge. There is a potential downfall to strategy, the amount of time the pair takes to mean-revert could become too costly to hold the positions for an extended period.

IV. Methodology

1. Data

The primary data source for the ETF/ETN data is Commodity Systems, Inc., LLC (CSI). The sample covers the period from June 30, 2016 to September 30, 2017 inclusive for in-sample estimation. The days of October 1, 2017 up to and including December 15, 2017 were reserved for out-of-sample validation. These securities were filtered to meet the following liquidity requirements:

- average daily trading volume greater than 200K shares
- a minimum share price of \$2 over the period

These restrictions were placed to include only those ETF/ETNs that are likely to be excluded as “untradeable” by larger market participants due to insufficient liquidity. All prices are split/dividend adjusted. The resulting sample included 366 ETF/ETNs.

To simulate achievable trading results transaction costs were estimated at 50 bps per single sided trade (1 percent round trip trade). These costs include:

- \$1.00 per trade (currently offered by interactive brokers)
- SEC Section 31 transaction fee \$0.0000218 per \$1.00
- Slippage @ \$0.05 (difference between signal price and fill price)

2. Cointegration

A total of 366 ETF/ETNs met the tradability requirements described above resulting in 66,795 unique pairs. Each pair was then screened using the Augmented Dickey Fuller (ADF) Test² for cointegration with an alpha value of 0.99% confidence.

3. Model

The Kalman Filter (KF) is a recursive algorithm used to estimate the value of dynamic, unobservable variables. In the context of pairs trading, the hedge-ratio and mean spread between two securities are the latent-state variables of interest. The hedge-ratio allows the investor to determine how much of each ETF to long or short once a threshold has been met. This threshold is 1 to $2\pm$ sd. The cointegrated pairs, determined by prescreening each pair with the ADF test, are then selected by running a nested loop³ to pair ETF's. This method of

² The Matlab 2016 cadf function contained in the econometrics toolbox was used to perform the ADF test.

³ Code to create distinct pairs:

selection ensures all pairs are unique, there are no repeats and are cointegrated. The hedge ratio, denoted β_t , is time-varying, making traditional OLS an unsuitable model. The KF is conducive to dynamic estimation, defined by a set of equations which continuously update with each time step. The equations used concurrently are the measurement and state-transition equations (see part b below).

a) Traditional Kalman Filter

$$\begin{aligned} \mathbf{z}_t &= \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t && \text{Measurement equation} \\ \mathbf{x}_t &= \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{B}_t \mathbf{u}_t + \mathbf{w}_t && \text{State transition equation} \end{aligned}$$

b) Pairs trading Kalman Specification (used in this study)

$$\begin{aligned} \mathbf{y}_t &= \mathbf{x}_t \beta_t + \epsilon_t && \text{Measurement equation} \\ \beta_t &= \mathbf{I} \beta_{t-1} + \omega_t && \text{State transition equation} \end{aligned}$$

Table 1. Traditional Symbols to Pairs Symbols Mapping

| Name | Description | Traditional | Pairs Notation |
|---------------------------|-----------------------------------------------------|--------------------|------------------------|
| Observation (Measurement) | Measure of true state | \mathbf{z}_t | \mathbf{y}_t |
| True state | | \mathbf{x}_t | β_t |
| Observation matrix | Maps current state into observed space | \mathbf{H}_t | \mathbf{x}_t |
| Observation noise | $\mathbf{v}_t = \epsilon_t \sim N(0, \mathbf{R}_t)$ | \mathbf{v}_t | ϵ_t |
| State-transition matrix | Transition measurements applied to previous state | \mathbf{F}_t | \mathbf{I} |
| Previous estimate | | \mathbf{x}_{t-1} | β_{t-1} |
| Control-input matrix | Updates measurements in control vector | \mathbf{B}_t | Intentionally left out |
| Control vector | Current control values | \mathbf{u}_t | Intentionally left out |
| Process noise | $\mathbf{w}_t = \omega_t \sim N(0, \mathbf{Q}_t)$ | \mathbf{w}_t | ω_t |

Note: In both systems \mathbf{R}_t and \mathbf{Q}_t are the covariance matrices of the observation noise and process noise respectfully.

```

for(int i=1; i<N; i++)
{
    for(int j=i+1; j<N; j++)
    {
        Stock(i) = beta * stock(j);
    }
}

```

In this study, I removed the control variable matrix often found in applications where the state equation is permitted to change by some input (i.e. velocity or acceleration). I do not allow the state model to change⁴. The state transition matrix in this study is the identity matrix.

$$\beta_t = I\beta_{t-1} + \omega_t$$

To begin the Kalman filter process, I first needed to provide an initial estimate and initial value for the error in the estimate. The method is very sensitive to the initial values provided for Q_t and R_t . I used a genetic algorithm⁵ adapting code provided in Kinlay (2015) to estimate these values. The first data input (measurement) is introduced into the system. From these values, three steps are performed iteratively with each new observation. (Van Biezen, 2015) They are as follows:

Step 1. Calculating the Kalman Gain (KG). Both the error in the estimate (original for first iteration), and error in the measurement are used in computing the KG. Calculated as follows:

$$KG = \frac{\text{Error in estimate}}{\text{Error in estimate} + \text{Error in the measurement}}$$

The KG acts as a weight in determining which value, either the current measurement or previous estimate, to place more confidence in when entering the next step.

Step 2. The main step, updating the current estimate. Taking the previous estimate (First iteration this is the original estimate value) and summing this value with the weighted difference of the new measurement and previous estimate.

$$Est_t = Est_{t-1} + KG[Mea - Est_{t-1}]$$

It is in this step the KG plays its leading role. From step one's calculation if the errors in the estimates are large this means there is a lot of uncertainty in the previous estimate and the KG will be large. Clearly, the researcher wants this value containing inaccuracies to have a smaller impact going into 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 measurement and previous estimate, will retain a larger portion of the measured value when feeding into the current estimate. Alternatively, if the KG is small, this means the errors in the measurement were very large and less importance will be placed on the measured value to minimally impact the current estimate's calculation.

⁴ Further research may explore the use of the control component to handle such things as earnings announcements and other news items that have a transitory effect.

⁵ Source: <http://jonathankinlay.com/2015/02/etf-pairs-trading-kalman-filter/>

```
kfBetaLik = @(QR) kfBetaLikelihood(X(:,1),X(:,2), QR);
options = gaoptimset('Generations', 1000, 'TolFun', 1E-10);
[QR, Lik] = ga(kfBetaLik, 2, [], [], [], [], [0 0], [0.1 0.1], [], options);
```

Step 3. Calculate the new error in the estimate. The error in the estimate is updated by multiplying the previous error in the estimate by the difference between one and the KG.

$$E_{est_t} = \frac{(E_{mea})(E_{est_{t-1}})}{(E_{mea}) + (E_{est_{t-1}})}$$

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

This means if the previous error in the estimate was greater than the error in the data, the KG will be large. Subtracting from one will yield a small number. When multiplied by the previous error in the estimate quickly reduces the error in the estimate for the next iteration. If the KG was small it would indicate the errors in the measurement are large. Now we do not want the new estimates to be heavily influenced by the incoming observations. The error in the estimate therefore decreases at a slower rate.

V. Results

Table 2 shows the combined results of the 12 best pairs for the period from June 30, 2016 to December 12, 2017. The best pairs were selected based upon results estimated from June 30, 2016 to September 30, 2017 (in sample results). The out of sample results cover the period from Oct 2, 2017 to December 12, 2017.

| <i>ETF A</i> | <i>ETF B</i> | <i>APR</i> | <i>Sharpe</i> | <i>Cum. Return</i> | <i>Max. DD</i> | <i>Max. DDD</i> | <i>Q/R</i> |
|--------------|--------------|------------|---------------|--------------------|----------------|-----------------|------------|
| DJP | DRIP | 0.25 | 1.20 | 0.39 | -0.09 | 137 | 0.05 |
| DJP | ERY | 0.20 | 1.40 | 0.31 | -0.06 | 207 | 0.04 |
| DJP | GUSH | 0.33 | 1.53 | 0.51 | -0.11 | 64 | 0.01 |
| EMLC | GUSH | 0.26 | 1.23 | 0.40 | -0.14 | 115 | 0.01 |
| EMLC | SOXL | 0.11 | 0.89 | 0.16 | -0.09 | 147 | 0.00 |
| EMLC | SOXS | 0.18 | 1.17 | 0.27 | -0.11 | 110 | 0.00 |
| FNDF | SOXL | 0.17 | 1.23 | 0.25 | -0.07 | 109 | 0.00 |
| GUSH | PGF | 0.33 | 1.26 | 0.52 | -0.13 | 224 | 0.09 |
| JDST | XRT | 0.47 | 1.11 | 0.74 | -0.38 | 105 | 0.43 |
| JNUG | TVIX | 0.41 | 0.98 | 0.64 | -0.33 | 168 | 0.04 |
| JNUG | UVXY | 0.53 | 1.15 | 0.85 | -0.32 | 168 | 0.01 |
| KIE | SOXS | 0.14 | 1.07 | 0.22 | -0.13 | 115 | 0.01 |

Table 2.

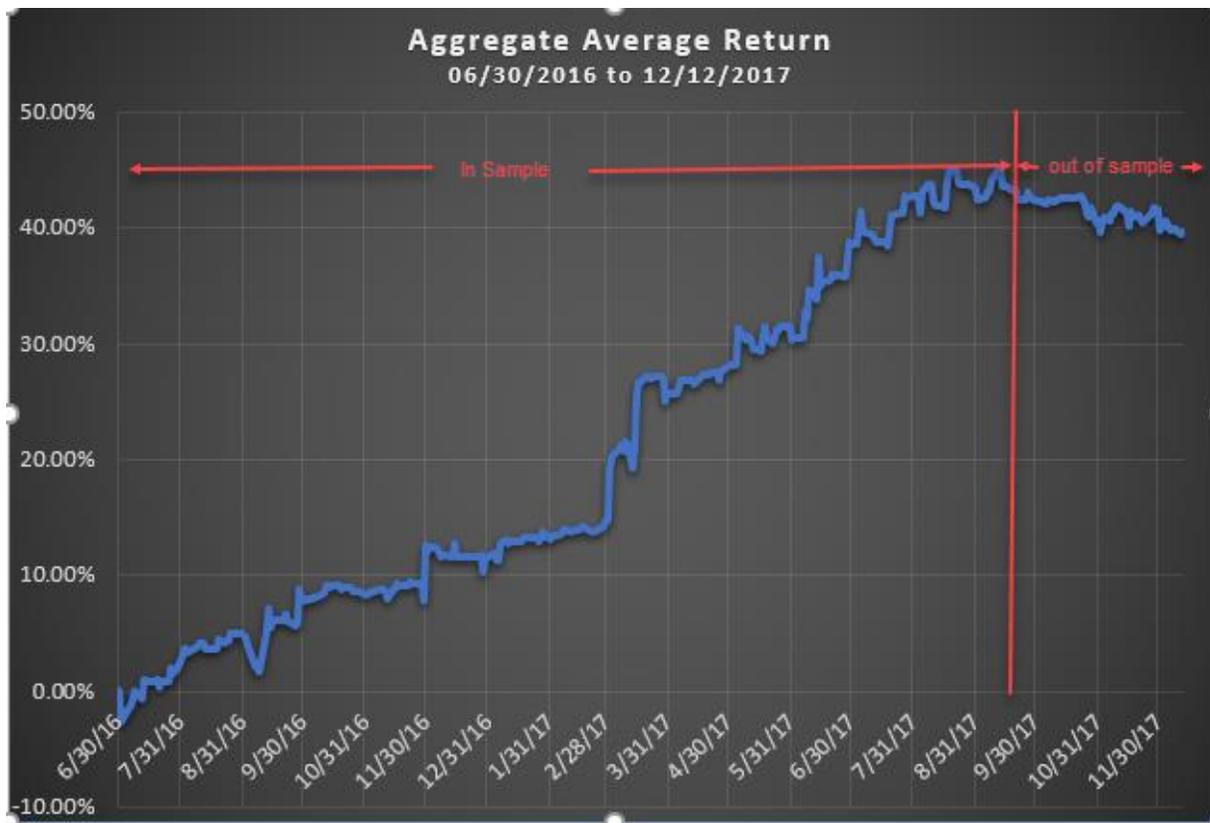


Figure 2. In-sample and out-of-sample results.

VI. Conclusions

Although the Kalman filter process produced excellent in sample results, the out-of-sample results proved disappointing and untradeable as depicted in Figure 2. The results may be because pairs trading strategies have gained widespread acceptance thus making profitability much more elusive.

That said, future investigations may explore using intraday data, such as 1 or 5 minute bars to see if profitable, tradeable strategies can be discovered.

References

- Benhamou, E. (2016). *Trend Without Hiccups - A Kalman Filter Approach*. Thomson Reuters.
- Chan, E. P. (2013). *Algorithmic Trading: Winning Strategies and Their Rationale*. Hoboken: John Wiley & Sons.
- De Moura, C. E., Pizzinga, A., & Zubelli, J. (2016). A Pairs Trading Strategy Based on Linear State Space Models and the Kalman Filter. *Quantitative Finance*, 16(10), 1559-1573.
doi:10.1080/14697688.2016.1164886
- Doucet, A., & Johansen, A. M. (2008). A Tutorial on Particle Filtering and Smoothing: Fifteen years later. Tokyo, Japan: Institute of Statistical Mathematics.
- Drakos, S. (2016). Statistical Arbitrage in S&P500. *Journal of Mathematical Finance*, 6, 166-177.
- Dunis, C. L., Giorgioni, G., Laws, J., & Rudy, J. (2010). *Statistical Arbitrage and High-Frequency Data with an Application to Eurostoxx 50 Equities*. Liverpool: Center for International Banking, Economics and Finance.
- Elliot, R. J., Van Der Hoek, J., & Malcolm, W. P. (2005). Pairs Trading. *Quantitative Finance*, 271-276.
doi:10.1080/146976680500149370
- Engle, R. F., & Granger, C. W. (1987, March). Cointegration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2), 251-276. Retrieved from <http://www.jstor.org/stable/1913236>
- Gatev, E., Goetzman, W. N., & Rouwenhorst, K. G. (2006). Pairs Trading: Performance of Relative-value Arbitrage Rule. *The Review of Financial Studies*, 19(2), 797-827.
- Halls-Moore, M. (2015, December 28). *State Space Models and the Kalman Filter*. Retrieved from QuantStart: <https://www.quantstart.com/articles/State-Space-Models-and-the-Kalman-Filter>
- Johansen, S., & Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand For Money. *Oxford Bulletin of Economics and Statistics*, 52(2), 169-210.
- Kim, P. (2011). *Kalman Filter for Beginners with MATLAB Examples*. Republic of Korea: A-JIN Publishing Company.
- Kinlay, J. (2015, February 23). *ETF Pairs Trading with the Kalman Filter*. Retrieved from Quantitative Research and Trading: <http://jonathankinlay.com/2015/02/etf-pairs-trading-kalman-filter/>
- Martinelli, R., & Rhoads, N. (2010, January). Predicting Market Data Using the Kalman Filter. *Technical Analysis of Stock and Commodities*.
- Miao, G. J. (2014). High Frequency and Dynamic Pairs Trading Based on Statistical Arbitrage Using a Two-Stage Correlation and Cointegration Approach. *International Journal of Economics and Finance*, Vol. 6(No. 3), 96-110. Retrieved from <http://dx.doi.org/10.5539/ijef.v6n3p96>
- Montana. (2009). *Dynamic modeling of mean-reverting spreads for statistical*.

PennState Elberly College of Science. (2017, December 8). *State Space Models*. Retrieved from Applied Time Series Analysis | STAT 510: <https://onlinecourses.science.psu.edu/stat510/node/27>

Tsai, Y.-C., & Lyuu, Y.-D. (2017). A New Robust Kalman Filter for Filtering the Microstructure Noise. *Communications in Statistics - Theory and Methods*, 46(10), 4691-4976.
doi:10.1080/03610926.2015.1096390

Van Biezen, M. (2015, October 27). Special Topics - The Kalman Filter (9 of 55) The Multi-Dimension Model 3: The State Matrix.

Vidyamurthy, G. (2004). *Pairs Trading: Quantitative Methods and Analysis*. Hoboken: John Wiley & Sons, Inc.