Statistical Arbitrage

Jeff Blokker, Emile Chamoun, Ibrahim Jreige,
Paris Georgoudis, Sameh Galal

Introduction

Statistical Arbitrage is the practice of using statistical measures to generate information that can be used to find consistent arbitrage opportunities in the financial markets. One technique discussed in a paper by Avellaneda and Lee (2008) is to isolate the idiosyncratic noise in the market and trade the noise, taking advantage of its mean reverting properties. In this project we follow that paper and analyze the U.S. equities market using Principal Component Analysis (PCA). Then we use Partial Least Squares (PLS) to extract residual noise on selected securities. Optimal trading of an Ornstein–Uhlenbeck process (O-U Process) is analyzed in a paper by Boguslavskaya and Boguslavskaya (2003). This optimal trading approach is explored using the residuals we generated from the market. Finally, a practical trading strategy is tested that combines the PLS of the market with optimal trading of the O-U process.

Principal Component Analysis

The standard model for analyzing the dynamics of equity prices is $\dfrac{dS_t}{S_t} = \mu dt + d\varepsilon_t$, where S_t is the price of the equity, μ is the average rate of growth, and $d\varepsilon_t$ is random noise. One refinement to this model seeks to separate the noise component into a systematic part and an idiosyncratic part. The systematic noise can be characterized as being due to p unknown Factors $F^{(j)}$, $j = 1, ..., p$. This model can be written as

$$d(\log S_t) = \mu dt + \sum_{j=1}^{p} \beta_j F^{(j)} + d\tilde{\varepsilon}_t$$
$$= \mu dt + \boldsymbol{\beta}\mathbf{F} + d\tilde{\varepsilon}$$

where $\tilde{\varepsilon}$ represents the idiosyncratic noise and $\boldsymbol{\beta}\mathbf{F}_t$ represents the systematic noise. In discrete time this can be written as

$$r_i = \log(S_i) - \log(S_{i-1}) = \mu\Delta t + \boldsymbol{\beta}\mathbf{F}_i + \Delta\tilde{\varepsilon}_i$$

where r_i is the log return. We choose the factors $F_t^{(j)}$ such that $E(\mathbf{F}) = 0$, $\text{cov}(\mathbf{F}) = \mathbf{I}$, $E(\Delta\tilde{\varepsilon}) = 0$, $\mathbf{F}$ independent of $\Delta\tilde{\varepsilon}_i$, and $\Delta\tilde{\varepsilon}_i$ are i.i.d. To fit this model we start with a sequence of n observed log returns

$$r_1 - \mu\Delta t = \beta_{11}F^{(1)} + \beta_{12}F^{(2)} + ... + \beta_{1p}F^{(p)} + \Delta\tilde{\varepsilon}_1$$
$$r_2 - \mu\Delta t = \beta_{21}F^{(1)} + \beta_{12}F^{(2)} + ... + \beta_{1p}F^{(p)} + \Delta\tilde{\varepsilon}_2$$
$$\vdots$$
$$r_n - \mu\Delta t = \beta_{n1}F^{(1)} + \beta_{n2}F^{(2)} + ... + \beta_{np}F^{(p)} + \Delta\tilde{\varepsilon}_1$$

In matrix form this can be written as $\mathbf{r} - \boldsymbol{\beta}\mathbf{F} - \mu\Delta t - \varepsilon + \Delta$.

$$(\mathbf{r} - \boldsymbol{\beta}\mathbf{F} - \mu\Delta t)(\mathbf{r} - \mu\Delta t - \boldsymbol{\beta}\mathbf{F})^T = \boldsymbol{\beta}\mathbf{F}\varepsilon)^T \varepsilon\varepsilon ( )^T + {}^T + {}^T$$
$$\text{cov}(\mathbf{r}) = \boldsymbol{\beta}E\left[(\mathbf{F} - \boldsymbol{\beta}\Delta t)(\mathbf{r} - \varepsilon\mu\Delta t)^T\right]\boldsymbol{\beta}^T = E[{}^T] \varepsilon\varepsilon E[{}^T] + E[{}^T] + E[{}^T]$$

Using the fact that $E(\mathbf{F}) = 0$, $\text{cov}(\mathbf{F}) = \mathbf{I}$, $E(\Delta\tilde{\varepsilon}) = 0$, $\mathbf{F}$ is independent of $\Delta\tilde{\varepsilon}_i$, and $\Delta\tilde{\varepsilon}_i$ are i.i.d. we can write

$$\text{cov}(r) = \boldsymbol{\beta}\boldsymbol{\beta}^T + \boldsymbol{\Psi}$$

where $\boldsymbol{\Psi}$ is the matrix of noise terms. Principal Component Analysis of the covariance matrix decomposes the log returns into a summation of Eigen values and Eigen vectors $(\lambda_i, \mathbf{e}_i)$.

$$\text{cov}(r) = \boldsymbol{\beta}\boldsymbol{\beta}^T + \boldsymbol{\Psi} = \sum_{i=1}^{p} \lambda_i \mathbf{e}_i \mathbf{e}_i^T + \boldsymbol{\Psi}$$

The largest Eigen values contribute most to the covariance matrix and correspond to the Factors with the greatest influence on the prices of the assets. By restricting the number of Eigen vectors used to m, $m < p$, we can utilize the most significant Factors and leave the contribution of the smaller Eigen vectors in the noise term $\hat{\varepsilon}_i$.

$$r_i = \mu dt + \sum_{j=1}^{m<p} \beta_j F^{(j)} + \hat{\varepsilon}_i$$

To evaluate the viability of this model we performed PCA on a set of 15 minute data between 2007 and 2009 on 500 stocks in the US markets. The first goal was to analyze the stability as time passes of the Eigen vectors generated from PCA.

Procedure

A set of Eigen vectors for an initial sample of 30 days of 15 minute data on 500 stocks in the US markets was generated. Then successive 30 day samples of data were formed by shifting the data one day into the future. The Eigen vectors of the successive samples were compared to the first set of Eigen vectors using a dot product.

$$\cos \theta_n = \mathbf{e}_0^T \mathbf{e}_n$$

There were two subtle problems that occurred. First, the algorithm that generates the Eigen vector may return an inverse Eigen vector. Secondly, as time passes, the order of the Eigen vectors may change. That is, the Eigen vector that was the largest in the first set may become the second largest at a later point in time. If care is not taken this appears as dispersion in the vector set that is not a reflection of reality.

Results

The following figures display the resulting Eigen vector dispersion after 30 days. We see that dispersion increases as the Eigen vectors become smaller.
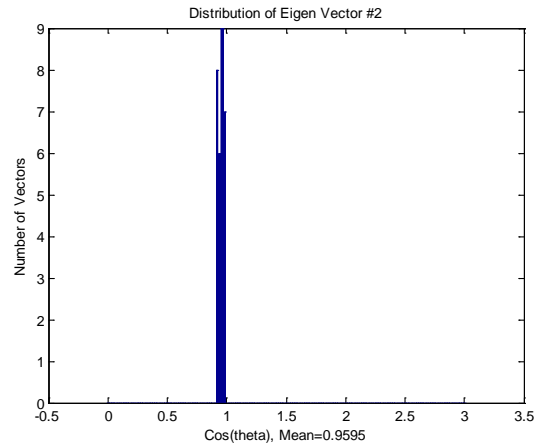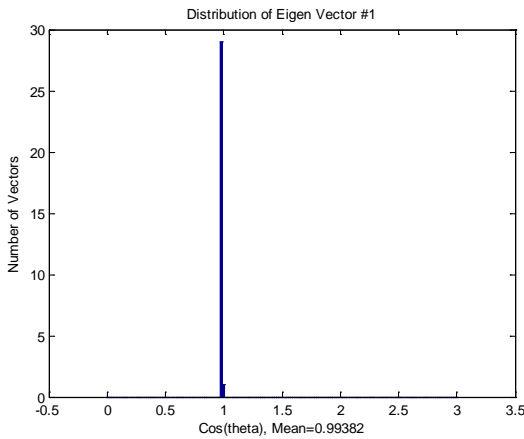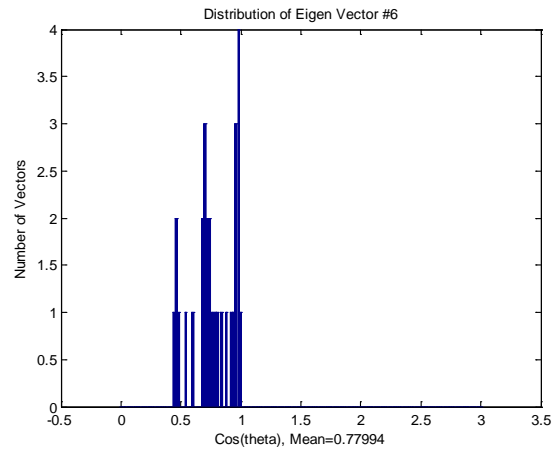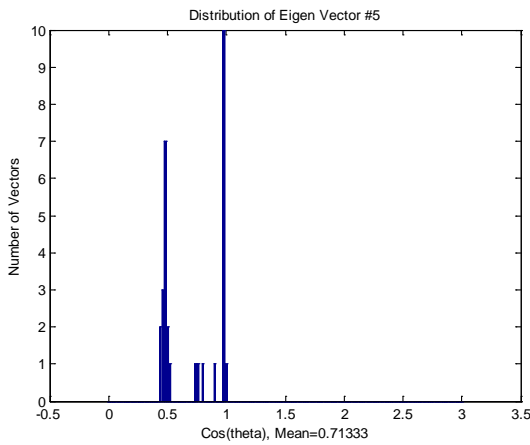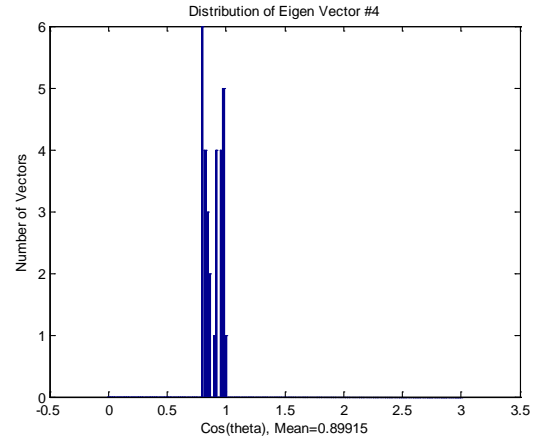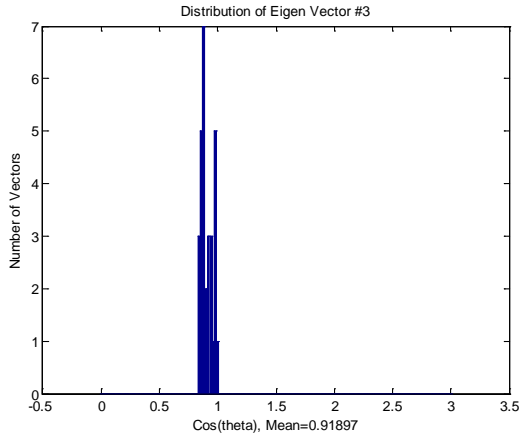
Mahalanobis Distance

Further analysis of the 15 minute dataset revealed that the first 45 minutes of the day produced the majority of the outlying data. To test this hypothesis we used the same 15 minute data set of 500 U.S. stocks and generated a training set of data that consisted of 100 days of data. Next we looked at the next 10 time points and measured their Mahalanobis distance (statistical distance) from the training data set. Mahalanobis distance is defined as

$$D_M = \sqrt{(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}$$

Next, the training data set was shifted 10 data points and a new test data set of 10 points was measured. This was repeated over the entire data set keeping track of the time of each distance measurement.
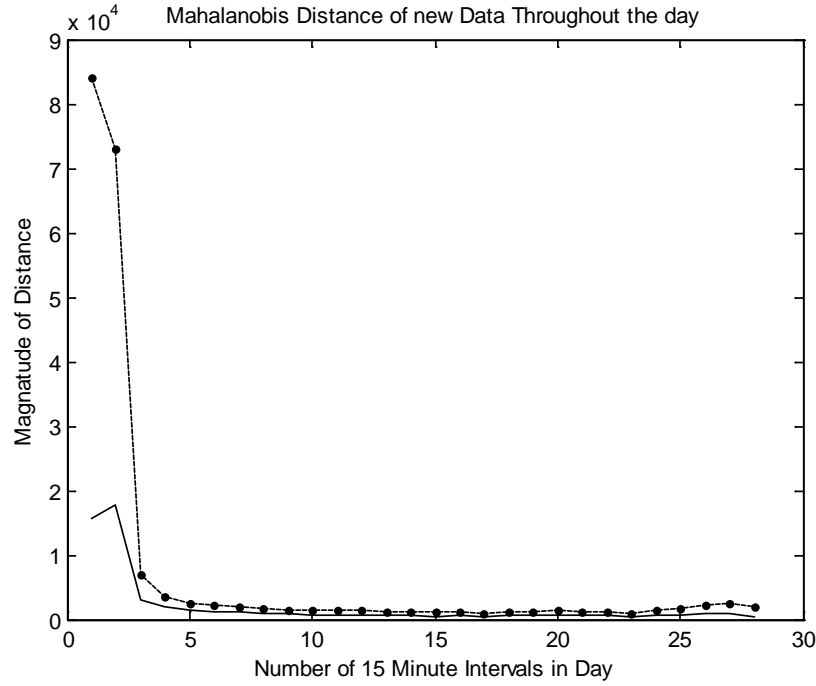
The solid line is the mean and the dotted line is the standard deviation of the distribution collected for each point of time in the day. The results show that the majority of the outliers occur in the first 45 minutes of the day. This indicates that the process that generates these returns is different in the first 45 minutes of the day than it is for the rest of the day.

Generation of Residuals

To extract the systematic variation from the data set we used a Partial Least Squares (PLS) algorithm. PLS is related to PCA. PCA finds the Eigen vectors that create the maximum variance in $E(\mathbf{X}^T\mathbf{X})$. PLS finds the Eigen vectors that create the maximum variance in $E(\mathbf{X}^T\mathbf{Y})$ where $\mathbf{Y}$ is an independent vector of returns from the desired stock. First, PLS finds the matrix information associated with the first Eigen vector. Then it subtracts this information from the return matrix $\mathbf{X}$. Then it finds the information for the second Eigen vector. This process repeats until all of the desired components are extracted. R^2 was used as a standard measure of the quality of the regression. $R^2 = 1 - \dfrac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{(\mathbf{y} - \mu)^T (\mathbf{y} - \mu)}$

From a training data set of 100 days of 15 minute log returns on 500 stocks, predictions were made on the next 10 points of data using PLS with the largest 9 Eigen vectors. Residuals were generated by subtracting the actual data from the predicted data. The training data set was then shifted 10 data points and the process repeated. Previously, we have shown that the first 45 minutes of the day has a different statistical profile than the remainder of the day. Therefore, we separated the residuals into two different regressions. Below are the results for the regression without the first 45 minutes of each day.
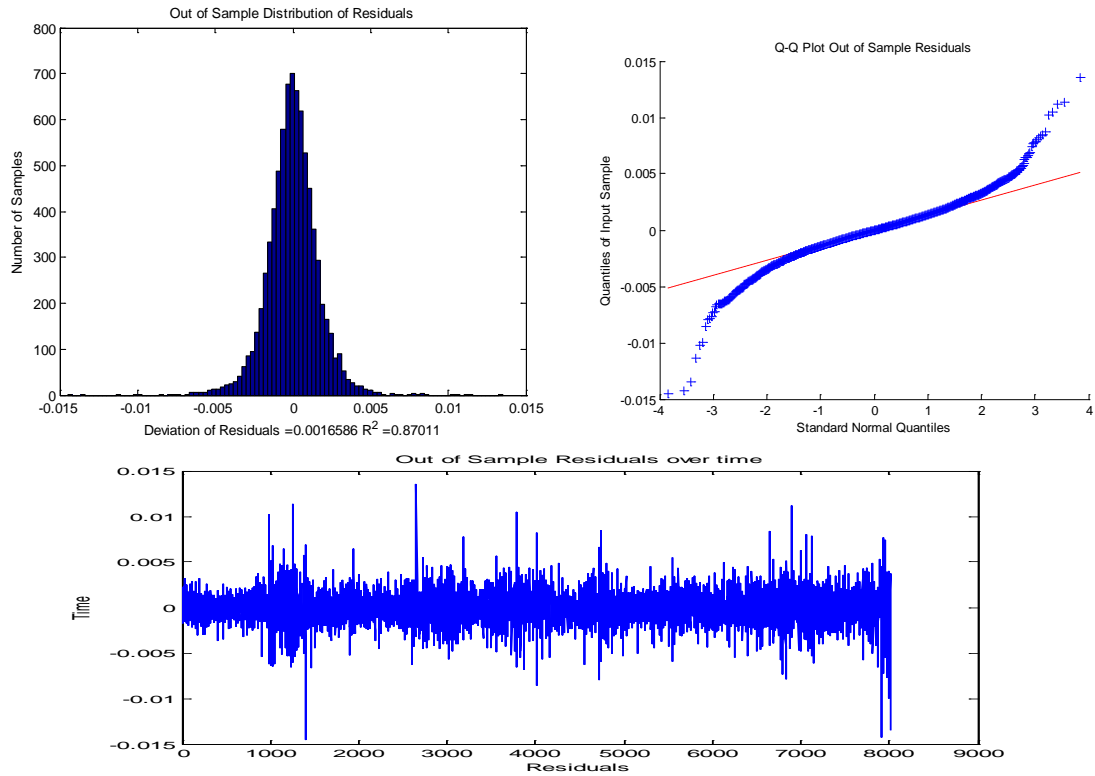
Below are the results for the regression using only the first 45 minutes of each day.
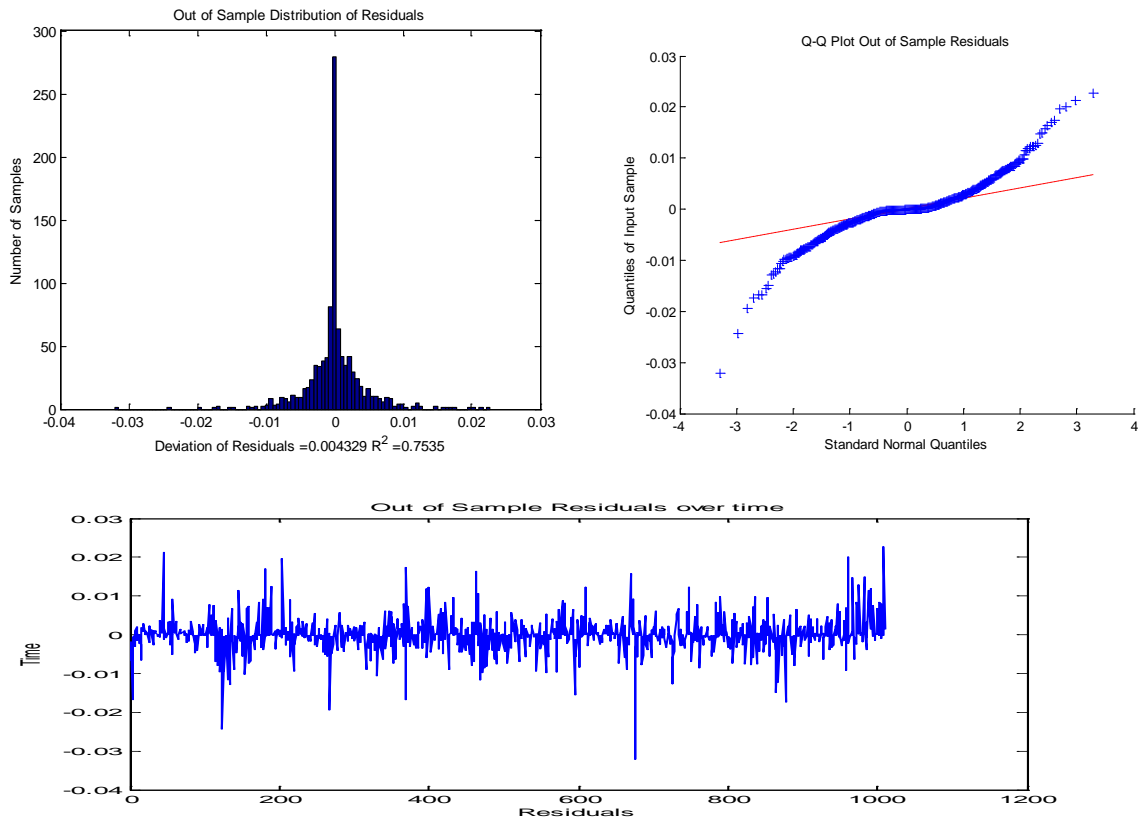
Optimal Trading the Ornstein-Uhlenbeck (OU) Process

We studied the optimal trading of residuals and explored the strategy in Boguslavsky-Boguslavskaya [ref]. The Ornstein-Uhlenbeck (OU) process in question is given by

$$dX_t = \kappa(\mu - X_t)dt + \sigma dB_t$$

Our OU process presents a drift component and satisfies (ref), but we show below that the resulting PDE and solution of the optimal control problem remains unchanged, so that we can adopt the same result for our calibrated OU process.

The problem considers the power utility function that satisfies $U = U(W_T) = W_T^{\gamma}/\gamma$, where γ is the risk aversion parameter. For a wealth process W_t, the change of wealth at time t satisfies $dW_t = \alpha_t dX_t$, where α_t is the assumed position in X_t at time t.

The optimal control problem therefore aims at finding the value function J that satisfies

$$J(W_t, X_t, t) = \sup_{\alpha_t} \mathbf{E}\left[\frac{1}{\gamma}W_T^{\gamma} \mid F_t\right].$$

Before solving the problem, we perform the following normalizations to simplify the form of our optimal control problem

$$X \to \frac{X - \mu}{\sigma}\sqrt{\kappa} = X'$$

$$\alpha \to \frac{\alpha}{\sqrt{\kappa}}\sigma = \alpha'$$

$$t \to \kappa t = t'$$

These normalizations leave the form of the wealth unchanged since

$$dW_t = \alpha_t dX_t = \frac{\sqrt{\kappa}}{\sigma}\alpha'_t \frac{\sigma}{\sqrt{\kappa}}dX'_t = \alpha'_t dX'_t$$

$$\frac{\sigma dX'_t}{\sqrt{\kappa}} = \kappa(\mu - \frac{\sigma}{\sqrt{\kappa}}X'_t - \mu)\frac{dt'}{\kappa} + \frac{\sigma}{\sqrt{\kappa}}dB_{t'}$$

So the model after normalization simplifies to:

$$dX_t = -X_t dt + dB_t$$

$$dW_t = -\alpha X_t dt + \alpha dB_t$$

Using Ito's formula, we see that:

$$dJ = J_t dt + J_x dX_t + J_w dW_t + \frac{1}{2}J_{ww}(dW_t)^2 + \frac{1}{2}J_{xx}(dX_t)^2 + J_{xw}(dW_t)(dX_t)$$

$$\boxed{dJ = (J_t - xJ_x - \alpha x J_w + \frac{1}{2}J_{xx} + \frac{1}{2}\alpha^2 J_{ww} + \alpha J_{xw})dt + (J_x - \alpha J_w)dB_t}$$

Applying the $\int_{t}^{t+h}$ operator to both sides, we obtain the following Hamilton-Jacobi-Bellman equation. We note that the dB_t term cancels out since it reduces to the expectation of a stochastic integral. This yields the Hamilton-Jacobi-Bellman equation satisfying

$$\sup_{\alpha}\left(J_t - xJ_x - \alpha xJ_w + \frac{1}{2}J_{xx} + \frac{1}{2}\alpha^2 J_{ww} + \alpha J_{xw} \right) = 0,$$

and the corresponding first order optimal condition

$$\alpha^*(w,x,t) = x\frac{J_w}{J_{ww}} - \frac{J_{xw}}{J_{ww}}.$$

By simple substitution, the PDE of the problem now satisfies

$$J_t - xJ_x + \frac{1}{2}J_{xx} - \frac{1}{2}J_{ww}\left(\frac{J_{xw}}{J_{ww}} - x\frac{J_w}{J_{ww}} \right)^2 = 0.$$

With the problem setup above, the optimal trading strategy in terms of the position α_t in at every time t is solved by:

$$\alpha^*_t = -wxD(\tau),$$

where $\tau = T - t$,

$$v = 1/\sqrt{1-\gamma},$$

$$C(\tau) = \cosh v\tau + v\sinh v\tau,$$

$$C'(\tau) = v\sinh v\tau + v^2 \cosh v\tau,$$

$$D(\tau) = \frac{C'(\tau)}{C(\tau)}.$$

Calibrating OU Process

According to literature by (Avellaneda, 2008) the estimation of the co-integrating residual is assumed to follow an Orstein-Uhlembeck process. The stochastic differential equation (SDE) is given by: $dX_t = \kappa(\mu - X_t) + \sigma dW_t$ (1), with κ the mean reversion rate, μ the long term mean and σ the volatility noise term.

Solving the SDE by applying Ito-Doeblin's formula we generate the following discrete equation:

$$X_{t+1} = X_t e^{-\kappa\delta} + \mu\left(1 - e^{-\kappa\delta}\right) + \sigma\sqrt{\frac{1-e^{-2\kappa\delta}}{2\kappa}}N_{0,1} \ (2),$$

which recursively can be expressed as a discrete AR(1) process.

Calibration using Least Squares Regression

Therefore, the continuous-time consecutive observations are linear with an iid normal random term, expressed as: $X_{t+1} = aX_t + b + \epsilon$ (3) , where $a = e^{-\kappa\delta}$,

$$b = \mu\left(1 - e^{-\kappa\delta}\right), sd(\epsilon) = \sigma\sqrt{\frac{1-e^{-2\kappa\delta}}{2\kappa}}, dW_t \sim N_{0,1}.$$

It becomes obvious that the coefficients in the O-U process in (1) take the following parameters based on (3):

$$\kappa = -\frac{\ln a}{\delta}, \mu = \frac{b}{(1-a)}, \sigma = sd(\epsilon)\sqrt{\frac{-2\ln a}{\delta(1-a^2)}}.$$

Assuming we want to calculate the parameters and the residuals are linear, the most common method is minimizing the sum of the square errors. Therefore,

$$a = \frac{n\sum_{t=1}^{n} X_{t-1}X_t - \sum_{t=1}^{n} X_{t-1}\sum_{t=1}^{n} X_t}{n\sum_{t=1}^{n} X_{t-1}^2 0 - \sum_{t=1}^{n} X_{t-1}}, b = \frac{\sum_{t=1}^{n} X_t - a\sum_{t=1}^{n} X_{t-1}}{n},$$

$$d(\epsilon) = \sqrt{\frac{nn\sum_{t=1}^{n} X_t^2 - \sum_{t=1}^{n} X_t^2 - a\left(n\sum_{t=1}^{n} X_{t-1}X_t - \sum_{t=1}^{n} X_{t-1}\sum_{t=1}^{n} X_t\right)}{n(n-2)}}.$$

Therefore, after generating a, b and $sd(\epsilon)$ we can easily calculate κ, μ and σ of the O-U process.

Calibration using Maximum Likelihood Estimates

A second way for estimating the parameters of the linear consecutive residual observations, is by including the linear model into the maximum likelihood equation to maximize the probability likelihood of these parameters. The standard setup assumes that the errors are normally distributed, but as we mentioned before this is not the case in our estimation scenario. However, the standard likelihood equation stands as $P\left(N_{0,1} = X\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}X^2}$. Solving the conditional probability density for X_{t+1} based on X_t we get the following probability function:

$$f(X_{t+1}|X_t; \kappa, \mu, \hat{\sigma}) = \frac{1}{\sqrt{2\pi\hat{\sigma}}}\exp\left[-\frac{\left(X_t - X_{t-1}e^{-\kappa\delta} - \mu\left(1-e^{-\kappa\delta}\right)\right)^2}{2\hat{\sigma}^2}\right], \text{ where } \hat{\sigma} = \sigma\sqrt{\frac{1-e^{-2\kappa\delta}}{2\kappa}}.$$

The maximization of the log-likelihood function $\mathcal{L}(\kappa, \mu, \hat{\sigma}) = \sum_{t=1}^{n} \ln\left(f(X_{t+1}|X_t; \kappa, \mu, \hat{\sigma})\right)$ yields the following solutions to the parameters:

$$\kappa = -\frac{1}{\delta}\ln\frac{\sum_{t=1}^{n}(X_t-\mu)(X_{t-1}-\mu)}{\sum_{t=1}^{n}(X_{t-1}-\mu)^2},$$

$$\mu = \frac{\sum_{t=1}^{n}\left[X_t - X_{t-1}e^{-\kappa\delta}\right]}{n(1-e^{-\kappa\delta})},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left[X_t - \mu - e^{-\kappa\delta}(X_{t-1}-\mu)\right]^2$$

By substituting κ into μ we can further simplify the equations above. The estimated parameters generated for κ, μ and σ under the maximum likelihood approach appear empirically to be very similar to the least squares approach.

However, according to Johnson (2009) the maximum likelihood approach usually overestimates the mean reversion speed κ and thus underestimates the noise σ. In addition, in the same paper the Kalman filtering method outperforms the maximum likelihood method. Another way to explain why over-fitting is causing problems to the trading positions also comes from Boguslavsky and Boguslavskaya (2009) suggesting that overly aggressive positions and understated standard deviation can lead quickly to bankruptcy.

Kalman Filtering-EM Algorithm

The Kalman filtering approach most closely resembles the least squares methodology in that it uses the assumed i.i.d. normal error component of the linear discrete equation to produce more accurate measurement of the noisy part.

The Kalman filter estimates the Markov state of a process given a sequence of noisy observation and a guess of the initial parameters. As a result the X_{t+1} based on X_t becomes the following: $X_{t+1} = A_{t+1}X_t + B_{t+1}u_{t+1} + W_{t+1}$, where A_{t+1} is the state's transition matrix from the previous state, B_{t+1} is the control-input model which controls vector u_{t+1} and W_{t+1} is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance Q_t. The whole point of the Kalman filter is that it acts recursively, which is seen in the following equations.

Predict

Predicted (*a priori*) state estimate	$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \mathbf{B}_k\mathbf{u}_k$
Predicted (*a priori*) estimate covariance	$\mathbf{P}_{k|k-1} = \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}} + \mathbf{Q}_k$

Update

Innovation or measurement residual	$\tilde{\mathbf{y}}_k = \mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$
Innovation (or residual) covariance	$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$
Optimal Kalman gain	$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{S}_k^{-1}$
Updated (*a posteriori*) state estimate	$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\tilde{\mathbf{y}}_k$
Updated (*a posteriori*) estimate covariance	$\mathbf{P}_{k|k} = (I - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}$

Since Kalman filtering requires initial guesses of the A, B and Q parameters which we ignore, we applied the expectations maximization algorithm strategy to obtain better estimates of these parameters. This two step algorithm goes as follows: Firstly, after we have applied Kalman filters to the current parameter estimates, we find the expected value of the log-likelihood of the data with respect to the condition distribution of the latent variables, given the observations under the current parameter estimates. Secondly, we find the parameters that maximize this log-likelihood, and repeat again this process.

The gain comes from the fact that the Kalman method filters out the noise from the residuals to produce estimates of the parameters of the latent OU process. This noise often results from market inefficiencies that could considerably affect the estimates of the OU parameters. The only drawback from using this approach is that it is much more computationally intensive than either the maximum-likelihood or least-squares methods, especially when the initial guesses are far from the final parameter estimates. Therefore, the running time of the EM algorithm is considerably reduced by using initial parameter guesses produced by the maximum-likelihood method. Figure 1 illustrates the cumulative wealth of our optimal trading strategy on the EvA residuals by using an initial wealth of $100 and a γ of 0 based on estimates from the EM algorithm (blue) and the maximum-likelihood method (red).
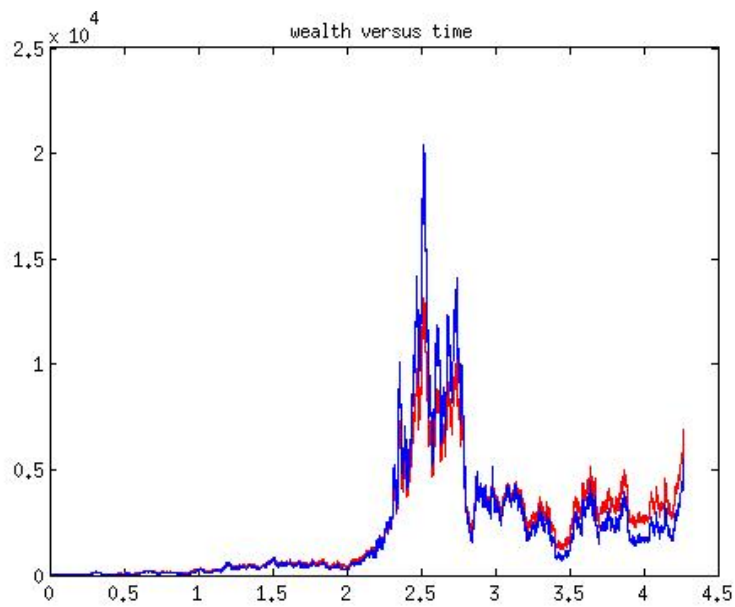
Figure 1

The difference between the cumulative curves obtained using the EM and maximum likelihood algorithms is illustrated in figure 2 for a better visualization of the advantages of our approach.
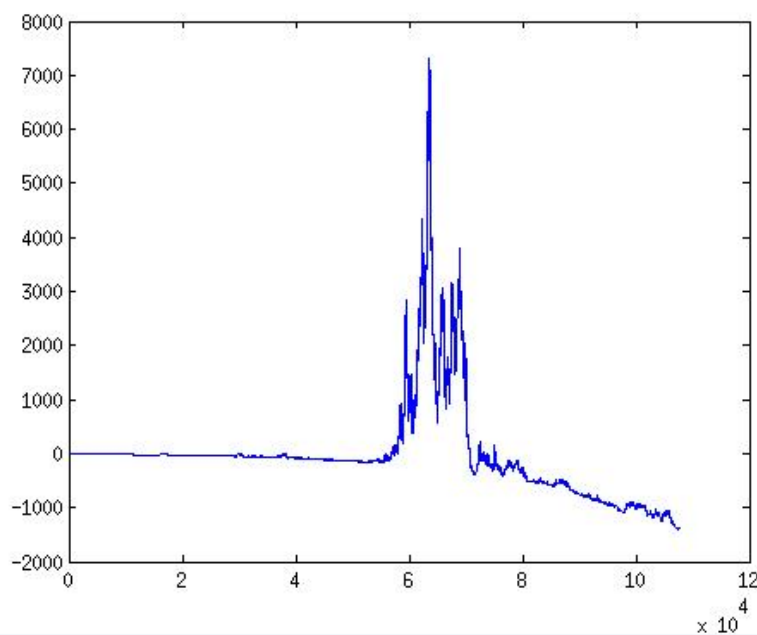

Figure 2

By analyzing figures 1 and 2, it is clear that the EM algorithm produces a much higher cumulative wealth at various times only to fall slightly below the maximum-likelihood approach later on. The higher cumulative wealth obtained by MLE in the second part of the graph can be explained by the fact that due to overestimation of the mean reversion speed, the maximum

likelihood approach takes more aggressive positions that can benefit its trading strategy at times but can also lead to bankruptcy relatively quickly.

Simulation Results

We tested the derived optimal strategy above on two sets of residuals: those provided by EvA for an unknown stock, and those derived for the energy stock Exxon Mobil (XOM) using the PLS method (described in the first part of this report) applied on the stock data provided by EvA.

For both residuals, we used a 70% training data to calibrate the mean-reverting process and the remaining 30% of the data to back-test the strategy. We herein present our results, with first ignoring transaction costs.

EvA residuals

Starting with an initial invested wealth of $100,000, and $\gamma=0$, the optimally traded EvA's 1-minute spaced residuals (first set) result in substantial profits over the trading period. Albeit the wealth process does not exhibit a continuously ascending pattern, the strategy is largely profitable overall. A risk neutral investor ($\gamma=0$) who traded the above strategy on the residuals would have his wealth peak at $4,300,000 into the third year, and end up with $3,700,000 at the end of the trading period. Our simulations also show that a value of $\gamma=0$ has achieved the best result on these residuals. Both risk-averse and risk-prone investors would have underperformed the risk-neutral investor on these residuals.
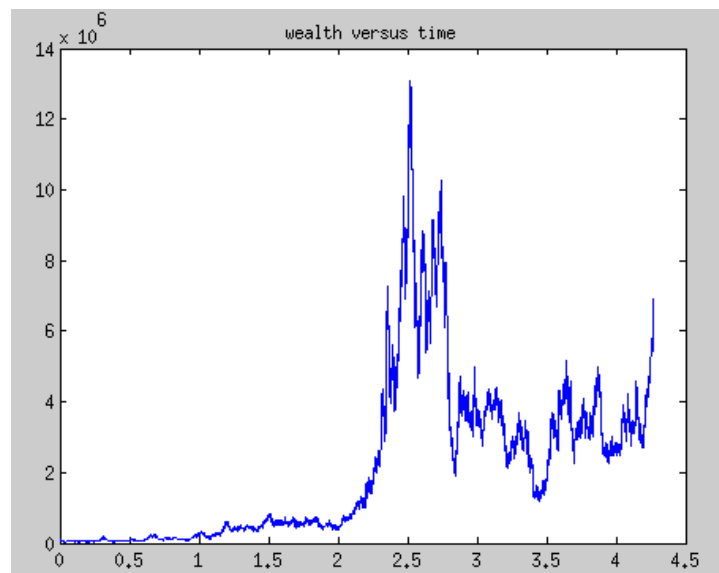

Fig. 3 – Wealth process for optimally traded EvA residuals (first set). $\gamma=0$, $\delta=1$min.

XOM residuals

We then implemented the same strategy on the residuals that we generated for XOM stock using the PLS method described earlier. For an initial invested wealth $W_0=\$100,000$, with $\gamma=0$ and 15-minute spaced residuals, the achieved wealth pattern has an ascending trend reflecting an obvious statistical arbitrage. A risk-neutral investor who traded these residuals would have seen her wealth increase five fold in one quarter's time and ended up with $490,000, in a transaction-cost-free world.
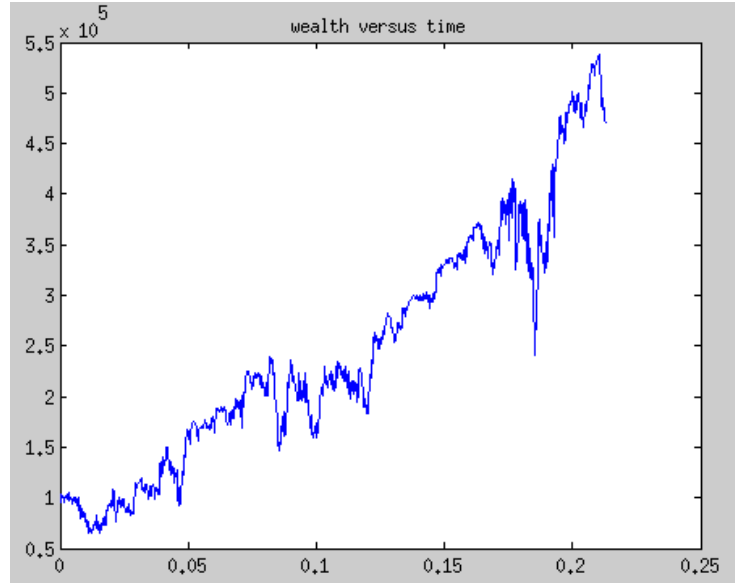
Fig. 4 – Wealth process for optimally traded XOM. $\gamma=0$, $\delta=15$min.

Calibration Window

One factor to consider when calibrating the OU process is the nature of the calibration window. We have adopted a common 70%-training-30%-testing breakdown, but a question arises on whether we should roll the window, i.e. calibrate continuously on each tick (while maintaining a fixed width) to adapt to the possibly changing statistical properties of the data, expand the window on each tick while maintaining the starting point, or just apply a one-time fixed calibration on the training data. Simulations showed that a one-time fixed calibration on 0.7x of the data yielded the best results for most residual sets, with a close performance from the other methods so that no one is a clear winner. We illustrate in the figure below the wealth process results of the three calibration methods applied on EvA second residuals set. Performance appears to be consistent across all three.
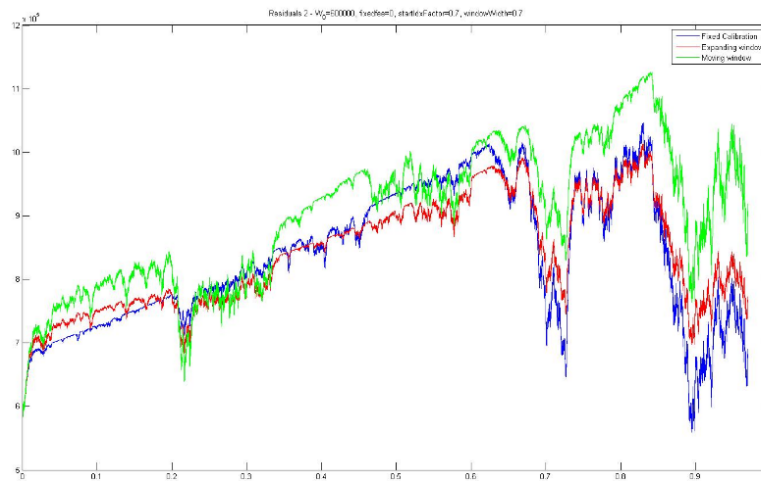

Fig. 5 – Wealth process for EvA residuals (2nd set) with different OU window calibrations (blue: fixed, red: expanding, green: moving).

Transaction Costs

The major factor that makes the above results hard to achieve in a real trading world is the existence of transaction costs. The subtraction of a fixed cost from the wealth at each trade rebalancing (and hence the application of a transaction-cost-free optimal trading solution on a wealth process that is adjusted by an actual transaction cost) has yielded suboptimal results, and often times a decreasing wealth process leading to quick bankruptcy in the event that the starting invested wealth W_0 is not sufficiently high. The green curve in figure [bob4] shows the evolution of the wealth process with 10x the fixed transaction cost charged at each tick (assuming that each trade involves an opposite position in 9 counterpart stocks.)

An approach to deal with the problem is to theoretically allocate a separate fund for the transaction fees so that the optimal control solution is still theoretically applied to a transaction-cost-free wealth process only. Practically, the evolution of the actual wealth accounts for the net difference between the two funds. Figure [bob5] shows the wealth evolution in this case, where the green curve again accounts for 10x the transaction cost charged at each trade.
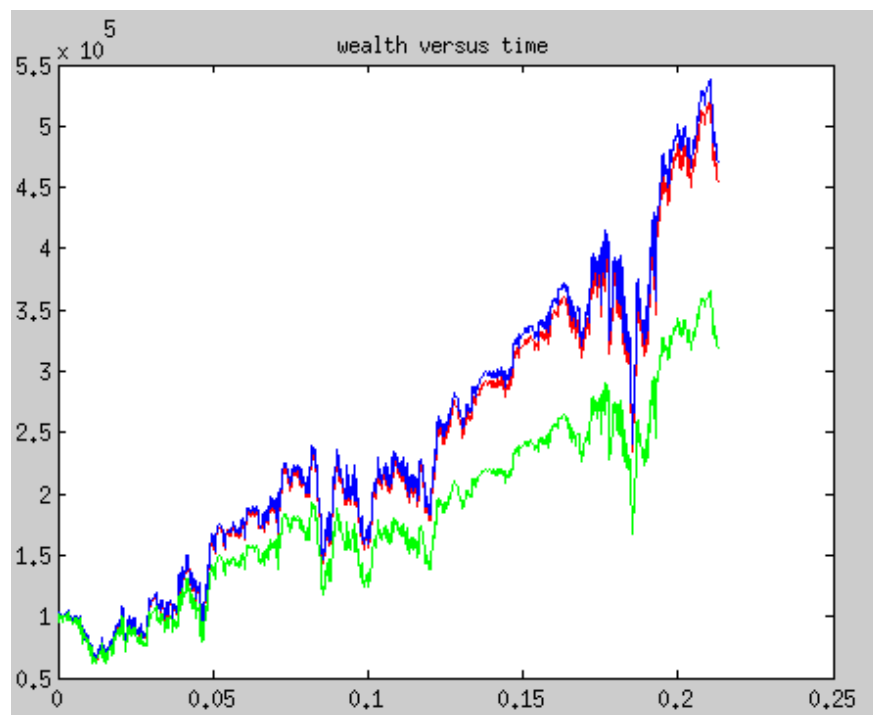

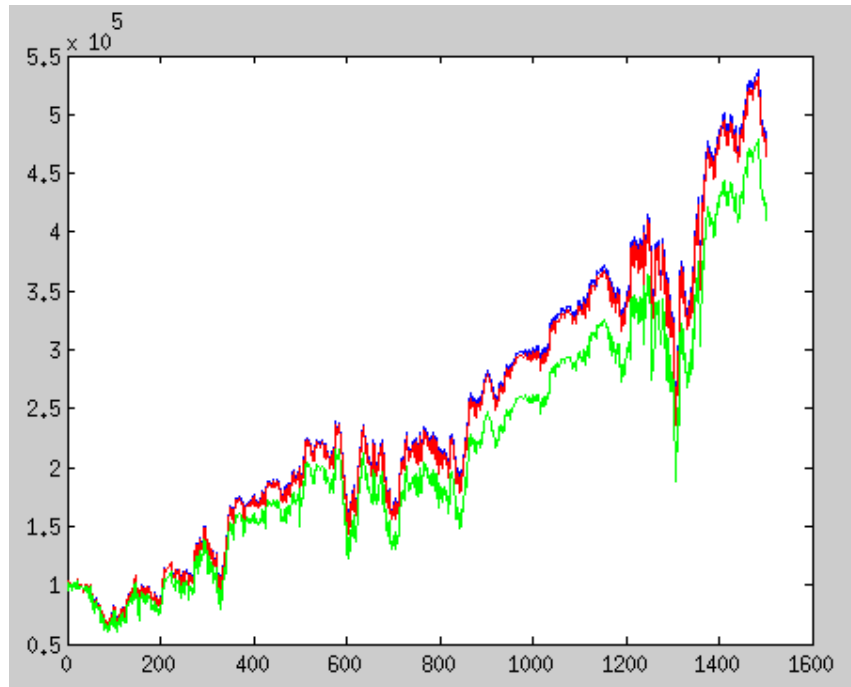Fig. 6 – Wealth process with $4 transaction cost (green: 10x cost at each rebalance, red: 1x cost at each rebalance).

Fig. 7 – Wealth process with $4 transaction cost paid from a separate fund (green: 10x cost at each rebalance, red:1x cost at each rebalance).
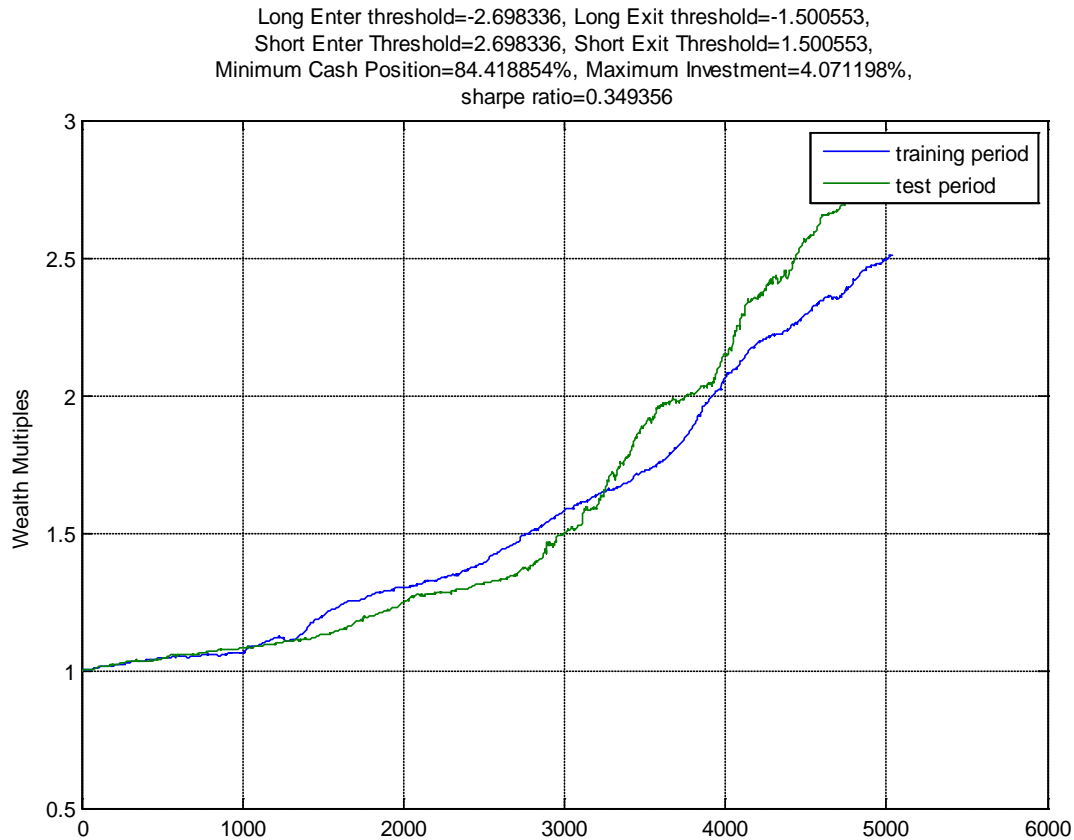
Portfolio Optimization with Residuals

After generating residuals, the next problem to consider is how to allocate capital for investment in different residuals based on mean reversion strategy. Since the returns of the residuals are not strongly correlated (they should be since they are the idiosyncratic part of the return), we could devise a strategy where the weights assigned to different residuals are the same. The only parameters needed are a threshold criteria for determining when to invest in the market and the leverage ratio. We devised a trading strategy summarized as follows:

```
For every 15 min
    Find residuals with returns exceeding certain threshold
    Invest equally in all residuals according to leverage and maximum individual residual percentage
    Generate the resultant portfolio using residuals replicating portfolios
    Rebalance portfolio by adjusting the differences between holding portfolio and required portfolio
end
```

The following graph shows the trading results for trading where a training period is used to find the optimal parameters while a subsequent test period was used for back testing.

Long Enter threshold=-2.698336, Long Exit threshold=-1.500553,
Short Enter Threshold=2.698336, Short Exit Threshold=1.500553,
Minimum Cash Position=84.418854%, Maximum Investment=4.071198%,
sharpe ratio=0.349356

Conclusions

We derived residual from PLS on a large set of 15 intra-day log returns data on the US equity market. We concluded that the first 45 minutes of the day has the largest number of outliers in the data. We analyzed optimal trading of an O-U process and explored several methods for estimating the parameters of this process. The EM method produced better estimation of the O-U parameters. Then using the derived residuals we simulated the trading process for various levels of risk aversion. Even after adding transaction costs we conclude that this trading strategy produces profitable results.

Bibliography

Avellaneda, Marco and Jeong-Hyun Lee.-H. (2008, July 11). Statistical Arbitrage in the U.S. Equities Market.

Boguslavsky, Michael and Elena Boguslavskaya. (2004, February). Arbitrage under Power.

Johnson, Thomas (2009, June 1). Approximating Optimal Trading Strategies Under Parameter Uncertainty: A Monte Carlo Approach. Chicago.