Time Series Prediction of US Swap Rates

Roland Burton
rolandb@stanford.edu

Shandor Dektor
sgd@stanford.edu

Dawn Wheeler
dawnw@stanford.edu

Abstract

In this project machine learning techniques were used to generate technical trading strategies in the US interest rate swap markets. Leela, a correlation algorithm that is closely related to autoregression, was developed to detect short term repeating patterns to predict future market moves. This algorithm exhibited a Sharpe Ratio of 1 when applied to a single swap. When additional swaps were used the Sharpe Ratio increased to 1.5. Finally, when PCA decomposition was employed, the time series of residuals exhibited a Sharpe Ratio in excess of 2.

Introduction

An interest rate swap is a contract between two parties, who agree to exchange fixed interest rate payments for floating rate payments[1]. The fixed interest rate is the quoted rate of a swap, and this rate primarily depends upon the duration or maturity of the contract. The fixed rate is the expected value of short term interest rates over the contract duration, and is often treated as a proxy for the borrowing costs of large corporations. The swap rate is measured in percent, and daily changes are measured in basis points, where one basis point is equal to 0.01%. Consequently profit and loss from holding a swap is most commonly quoted in basis points, a convention we adopt in this paper.

One of the uses of interest rate swaps is to hedge mortgage interest rate risk. When interest rates increase, mortgage servicers need to pay on a swap, further raising interest rates. Similarly, when interest rates decrease, servicers need to receive further decreasing rates. As the mortgage market is so large, the activities of servicers create temporary interest rate dislocations. It can take several days for the market to redistribute the new duration risk among other participants. It is hoped that the initial change in duration risk, followed by the subsequent redistribution of risk will create recognizable short term patterns in the data that can form the basis of a profitable technical trading strategy.

[1]Tuckman, B (2002) "Fixed Income Securities: Tools for Today's Markets". 2nd Ed. Wiley Finance

The performance of technical trading strategies is most frequently judged by looking at the Sharpe Ratio, a method originally devised for measuring the performance of Mutual Funds. The metric was further revised in 1994, and it is this version that will be used to measure trade performance[2]

$$\frac{E[R-R_f]}{\sqrt{\mathrm{var}[R-R_f]}}$$

where R denotes the annualized return and R_t denotes the annualized risk free rate. As swaps are derivative instruments requiring no capital, they can be fully cash collaterized, eliminating the need to incude the R_t term. Although quoted in annualized terms, the Sharpe Ratio is normally calculated using daily data. There are 251 trading days in the year, so for our purposes, the formula reduces to

$$\frac{E[r]}{\sqrt{\mathrm{var}[r]}}\sqrt{251}$$

and the daily return r is expressed in basis points. Trading strategies which achieve a Sharpe Ratio of over 1.0 are qualified as good. A Sharpe Ratio of over 2.0 is considered very good and Sharpe Ratios above 3.0 are outstanding[3].

1. Data

Barclays Capital provided daily interest rate swap quotes for swap contracts maturing in 2, 3, 5 and 10 years. Figure **??** shows a history of swap rates over the whole period. As can be seen interest rates have generally decreased over this whole period. A simple buy-and-hold strategy using ten year swaps returned a Sharpe Ratio of 0.25. However in reality a trader scales the size of his position based upon recently observed volatility. Such a strategy increases the Sharpe Ratio marginally to 0.27.

2. Preliminary Analysis

Initially each swap maturity was treated individually. The first algorithm to be tried was Ordinary Least Squares

[2]Sharpe, W. F. (1994). "The Sharpe Ratio". Journal of Portfolio Management 21 (1): 4958

[3]Masteika, Simutis (2006). "Stock Trading System Based on Formalized Technical Analysis and Ranking Technique". Computational Science ICCS 2006

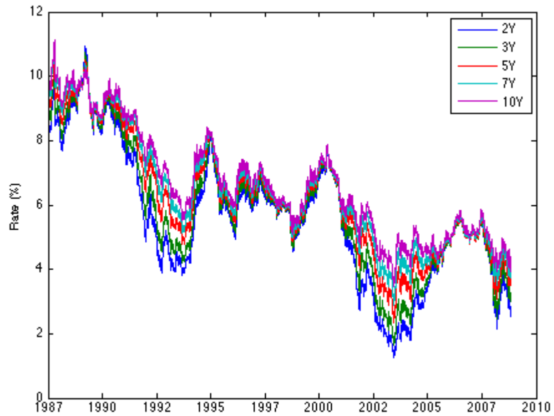

Figure 1. History of US swap rates from July 1987 to the present

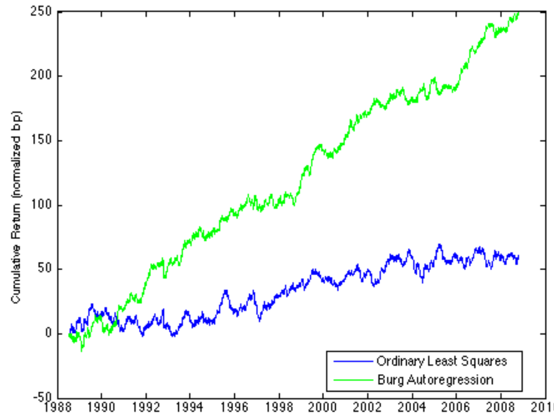

Figure 2. Linear Regression and Autoregression. Autoregression outperforms OLS both in return and in Sharpe Ratio. Visually, Sharpe Ratio is a measure of slope vs. stability of return.

(OLS) linear regression: The feature vector at time t is $x_t = [d_{t-1}\, d_{t-2} \ldots d_{t-p}]^T$, and the observation at time t is $y_t = d_t$ where d_t denotes the change in swap rate in basis points on day t. p is an integer that denotes how many days are considered for pattern matching. The coefficients, θ, are chosen such that $y_t \approx \theta^T x_t$ for all t. The trading signal for day $t+1$ is given by $\text{sgn}(\theta^T x_{t+1})$ and the return on the $t+1$ day is given by $r_{t+1} = d_{t+1}\,\text{sgn}(\theta^T x_{t+1})$. The data is heteroskedastic, and it is unreasonable to assume that θ is the same over all history, so θ is computed over a window of n days of recent data:

$$\theta_t = (X_t^T X_t)^{-1} X_t^T Y_t$$

where $X \in \Re^{n \times p}$ and the ith row of X is x_{t-i+1}^T, and $Y = [y_t\, y_{t-1} \ldots y_{t-n+1}]^T$. Reasonable values for n are from 63 to 1500, corresponding to ninety calendar days to six years. Recalling the motivation in the introduction, the idea is to detect short term patterns caused by temporary liquidity imbalances so reasonable values for p are 3, 4 or 5 (up to a full trading week).

An alternative approach to dealing with heteroskedasticity is to redefine the data based on local norms, so that $\tilde{x}_t = x_t/||x_t||$ and $\tilde{y}_t = y_t/||x_t||$, and then repeat the analysis as described as above. The θ used to compute r_{t+1} is based only on previously observed data, so the performance is out of sample. For the 10-year swap, after scaling the data to attempt to combat the heteroskedasticity, a Sharpe Ratio of 0.23 attained with $p = 3$ and $n = 251$, no better than a buy and hold. Without scaling, the Sharpe Ratio was zero.

Autoregressive (AR) methods were briefly investigated. Computation relied upon the built-in MATLAB function `ARBURG` which estimates parameters via the Burg Method. In a similar manner to linear regression, there are two parameters that can be chosen for the model, n and p. p is the order of the model and n is the window size used. The AR methods are very closely related to OLS, but they act upon correlation coefficients rather than on the data itself. Looking at the 10-year swap rate with $n = 502$ and $p = 4$, a Sharpe Ratio of 0.9 was attained, which is a much improved performance over OLS. Figure **??** compares the performance of the Burg AR algorithm and OLS regression.

3. The Leela Algorithm

Although the AR algorithm improved performance, it was unable to break through the barrier of a Sharpe Ratio of 1. A new correlation algorithm was proposed that is very closely related to the AR algorithm, which the authors have called Leela[4]. The return on day $t+1$ is given by:

$$r_{t+1} = d_{t+1}\,\text{sgn}(\sum_{i=t-n+1}^{t} \rho_{i,t}\tilde{y}_i)$$

where $\rho_{i,t} = \tilde{x}_t^T \tilde{x}_{t-i}$, and $\tilde{x}_t$ and $\tilde{y}_t$ are as previously defined. $\tilde{x}_t$ is the feature vector, and $\tilde{y}_t$ is trade selection. In words, the algorithm takes the last p days moves in the instrument, and then correlates this vector with all the previous p day moves in the past n days, giving n correlations, ρ. The trading signal, $\in \{-1, 1\}$, is the sign of the sum of all these correlations multiplied by the following days normalized move. The return is simply the trading signal based on data from day t and earlier, multiplied by the actual move on day $t+1$. By computing correlations (via normalized vectors), problems associated with heteroskedasticity are reduced. Another assumption in this algorithm is that patterns are symmetric. With respect to liquidity, symmetry assumes that a large seller in the market has the same

[4]The one eyed captain from Futurama

effect only reversed as a large buyer in the market. For equity markets this assumption is tenuous, but in fixed income markets sharp rallies and sell-offs are much more similar. In the following section two methods are examined to expand the feature vector and trade selection past a single swap.

4. Feature Vector and Trade Selection for Leela

We tested three main variants on the Leela algorithm to determine what feature vector and trading strategy held the most promise. Each of these algorithms inputs a different feature vector to the algorithm. The output from Leela was used to choose the best trade for each day.

4.1. Ten-year Feature Vector and Correlation

The 10y is the basic Leela algorithm: the feature vector is simply $d_t = d_t^{(10)}$. The remainder of the algorithm is identical to that described in section 2.

4.2. Multi-Instrument Feature Vectors

This variant is based on correlation using the three day performance of the overall market. This allows the algorithm to assess the recent trading environment more accurately. The feature vector for multi correlation consists of the normalized three day history for each swap:

$$x(t) = [d_{t-1}^{(2)}\ d_{t-2}^{(2)}\ d_{t-3}^{(2)} \dots d_{t-1}^{(10)}\ d_{t-2}^{(10)}\ d_{t-3}^{(10)}]$$

$$\tilde{x}_t = x_t / ||x_t||$$

$$\tilde{y}_t = [d_t^{(2)} \dots d_t^{(10)}] / ||x_t||$$

This algorithm selects the swap with the largest trading signal, as defined by:

$$\arg\max_s \operatorname{abs}\left(\sum_{i=t-n+1}^{t} \rho_{i,t} y_{i,s}\right)$$

Where $\rho_{i,t}$ is modified with:

$$\rho_{i,t} = \begin{cases} x(t)^T x(i) & \text{if } x(t)^T x(i) > 0.4 \\ 0 & \text{otherwise} \end{cases}$$

The algorithm considers how each swap performs given market conditions. Choosing the largest trading signal reduces the risk of making a poor decision. Using a ρ cutoff increases the similarity between current and past market conditions, but the bound must not be so strict that it eliminates relevant data.

4.3. Principal Component Decomposition

PCA was used to prefilter the data before passing it to the Leela algorithm. As was shown at the beginning of this project, the five different swap maturities appeared to show strong cross correlation. This was confirmed when the PCA decomposition was computed. Let $d_t^{(k)}$ denote the daily change (in basis points) of the k-year maturity swap on day t. Then define

$$D = \begin{bmatrix} d_{t-n+1}^{(2)} & \dots & \dots & \dots & d_{t-n+1}^{(10)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{t-1}^{(2)} & \vdots & \vdots & \vdots & d_{t-1}^{(10)} \\ d_t^{(2)} & d_t^{(3)} & d_t^{(5)} & d_t^{(7)} & d_t^{(10)} \end{bmatrix}$$

To compute the PCA factors we can use SVD, so $D = U\Sigma V^T$. Market convention is to set n=63, such that the PCA factors are computed over ninety calendar days. The Singular Values for a recent period are $\sigma = [211\ 33.6\ 12.4\ 7.2\ 2.4]$, showing how the first two factors dominate the decomposition. Most market practitioners use only the first two components, and it is this convention that is adopted in this project. It is also common to scale the two component vectors by their associated volatility, so define

$$F = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}^{1/2}$$

On the $t+1$ day, the least squares solution of $\hat{d}_{t+1} = Fw$, can be computed as $w = (F^T F)^{-1} F^T d_{t+1}$ giving the weights of the first two components ($w \in \Re^2$). Due to the scaling applied to F and orthogonality of U we expect $w \sim N(0, I)$.

Having computed the weights on each day, this time series can be input to the previous Leela algorithm. The performance with $n = 500$ and $p = 3$ returned Sharpe Ratios of around 0.8, a deterioration in performance.

Rather than looking at factors, component residuals can also be examined. The residuals are simply $\hat{d}_{t+1} - d_{t+1}$. These residual time series, one for each swap maturity, can then be input to the Leela algorithm. With $n = 750$ and $p = 4$, the results are plotted in Figure **??**, and summarized in the table below:

Year	Sharpe Ratio
2	0.4
3	1.4
5	1.2
7	2.2
10	0.8

While the 7y swap shows promise, the residual itself is not a trade-able instrument. Trading the 7y swap outright returns a Sharpe Ratio of close to zero. The best way to trade a residual is using a “butterfly” trade: If the algorithm signals to buy the 7y swap, the trader actually buys the 7y swap and sells the 5y and 10y swaps, with the amount of each swap dependent on the nullspace of F, so that the exposure to the first two factors is zero.

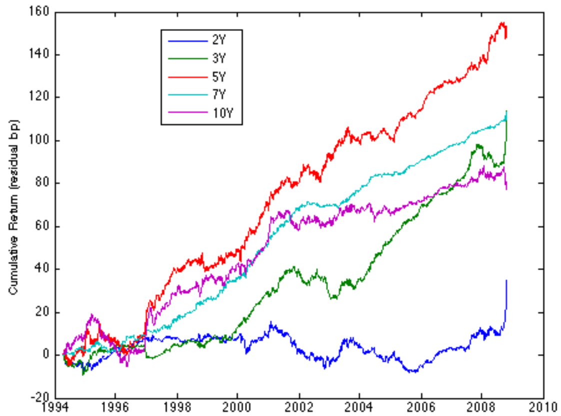

Figure 3. Trade Performance based upon PCA residuals. Again, the Sharpe Ratio is the most important metric rather than largest absolute return. The sharp spikes at the end of the time series are associated with the very recent market turbulence associated with bank insolvencies.

The strong performance of the 7y may not be a fluke of the data. One of the motivations for this report was the action of mortgage servicers hedging risk. The traditional duration of US mortgages is between five and ten years and many servicers use seven year instruments to hedge their duration exposure. The market is not deep enough to fully absorb the supply and demand, creating a dislocation. This dislocation is characterized by a high residual in the 7y point, which was shown to mean revert over a short period.

5. Results

The proposed algorithms were trained on the same market period, from approximately 1994 to late 2007. This was to ensure that all algorithms are exposed to the same market conditions making the results directly comparable. The year from late 2007 to late 2008 was used as test data to verify algorithm performance. The algorithms were trained for correlation window as this was the fuzziest parameter to determine. Initial tests showed that the three-day correlation window was superior for our algorithms. Varying the correlation window produced less convex results. Testing the data with a wide range of windows ensures we did not optimize for a random local maximum.

The algorithms were therefore compared with two goals: first, to find the correlation window with the largest Sharpe Ratio, and second to compare the yearly variation in IR[5].

[5]This project focused on empirical maximization of Sharpe Ratio. An excellent theoretical treatment can be found in Choey, M & Weigend, A - “Nonlinear Trading Models Through Sharpe Ratio Maximization” International Journal of Neural Systems 8 (3) (1997) 417-431

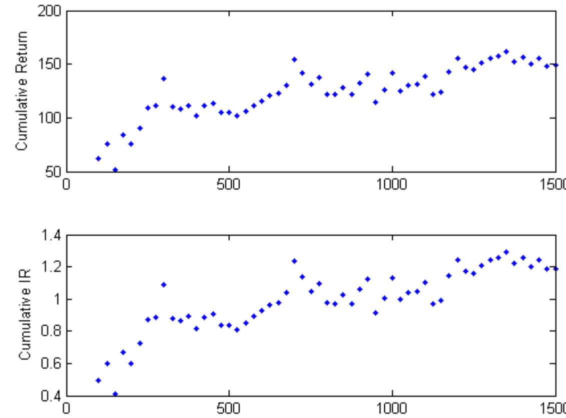

Figure 4. Training Results for 10y Swap Trade: Return and Sharpe Ratio vs. correlation window.

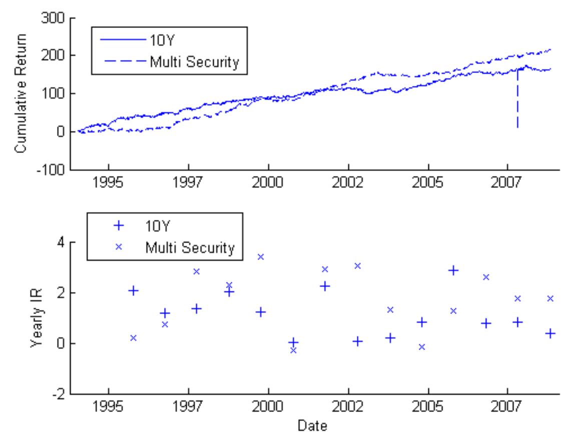

Figure 5. Best Single and Multi-Swap Trade Return and Yearly IR (300 day correlation window)

5.1. Correlation solely with 10y Swap

For the 10y swap, Figure **??** shows that as the correlation window increases the performance tends to increase, but with a Sharpe Ratio typically below 1.2. This is a decent result, and can be attributed to the character of the data. Using solely the 10y data, noise is likely to be a greater factor. Our data can be thought of as pseudo-random vectors of length 3; correlation between random unit vectors increases with decreasing length. Thus a longer correlation window allows us to ’average out’ the noise incipient to a short feature vector. This still does not provide a great Sharpe Ratio, however, and Figure **??** shows that the sharp ratio is fairly variable with time, though these sample location do not show any losses.

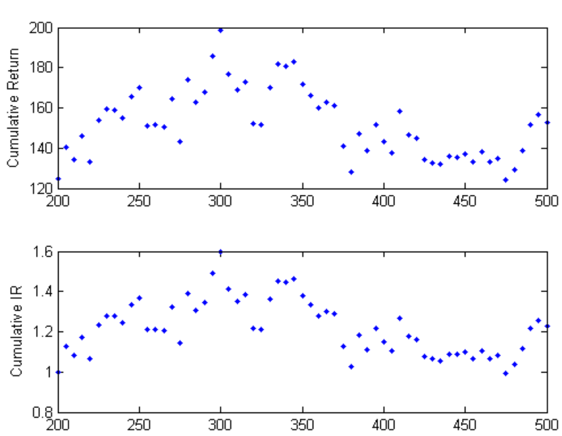

Figure 6. Training results for Multi-Swap Trade: Return and Sharpe Ratio vs. correlation window

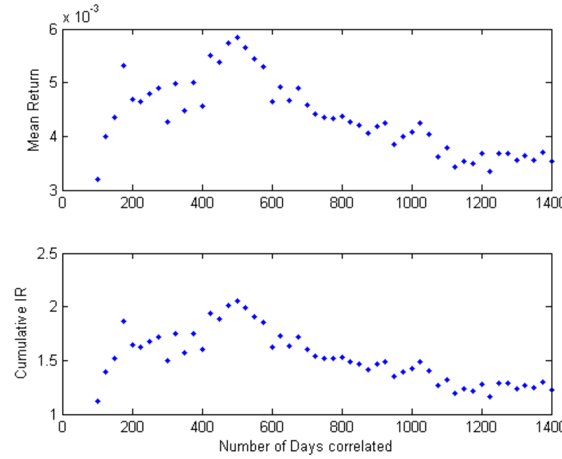

Figure 7. Training Results for PCA Trading: Return and Sharpe Ratio vs. correlation window

5.2. Correlation on 2y to 10y Swaps

For the multi-swap algorithm, Figure **??** shows that performance peaks at a correlation window of around 500 days, with a Sharpe Ratio of more than 1.5. This is a good and interesting result, as it indicates that the relation between the swaps changes with time. The algorithm weakens when correlated for much more than one and a half years, and the trade performs well over subperiods.

5.3. Correlation using PCA

PCA analysis has the strongest and most consistent correlation between performance and correlation window, with a peak at a Sharpe Ratio greater than 2 (Figure **??**). This is an excellent result for an automated program. This performance characteristic is likely due to the change in PCA components with time. As shown with multi swap correlation, the market character changes with a characteristic time on the order of a couple years. Therefore, PCA components are likely to change enough that comparison of residuals will no longer be ‘apples to apples’. Figure **??** shows that the yearly Sharpe Ratio performs at a consistently high level, with only one outlier with negative returns.

Conclusion

This project has shown that machine learning algorithms are able to generate trading rules that exhibit high ex-post Sharpe Ratios. While the simplest linear methods produced poor results, a correlation algorithm closely related to autoregression was shown to produce very strong returns. Furthermore when the data was preprocessed using Principle Components Analysis, the performance moved into a level warranting serious attention by technical traders.

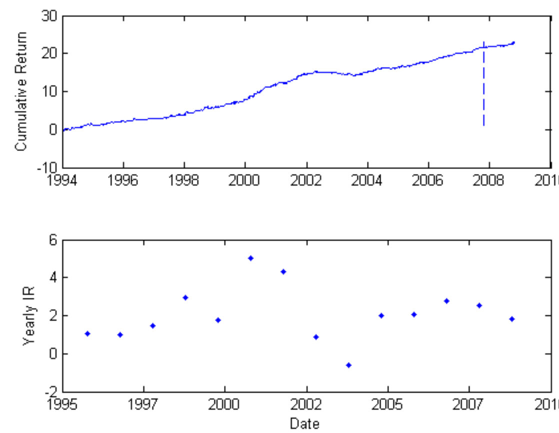

Figure 8. Best Peforming PCA trading: Return and Yearly IR (500 day correlation window)

Future Work

The algorithms presented in this paper are not tied to interest rate swaps, and are transferrable to other classes of financial instruments that exhibit similar properties. Specifically, the collection of instruments must be highly correlated but still exhibit temporary departures which create the trading opportunities. The most obvious extension is to similar fixed income classes, such as treasury, corporate, or mortgage debt. However in building a well balanced book, the portfolio manager would like to be active in as many uncorrelated markets as possible. One of the most natural markets to examine would be the G10 currency markets. Like rate swaps, currency markets offer liquidity and low margin requirements allowing for easy trading. Other applications would lie in a basket of energy futures or precious metals.