A Statistical Arbitrage Strategy

KUN ZHU

KTH Numerical Analysis
and Computer Science

Master's Degree Project
Stockholm, Sweden 2005

TRITA-NA-E05192

Numerisk analys och datalogi
KTH
100 44 Stockholm

Department of Numerical Analysis
and Computer Science
Royal Institute of Technology
SE-100 44 Stockholm, Sweden

A Statistical Arbitrage Strategy

KUN ZHU

TRITA-NA-E05192

Master's Thesis in Numerical Analysis (20 credits)
at the Scientific Computing International Master Program,
Royal Institute of Technology year 2005
Supervisor at Nada was Jesper Oppelstrup
Examiner was Jesper Oppelstrup

Abstract

Statistical arbitrage is a profit situation arising from pricing inefficiencies between securities. This is usually identified through mathematical modeling techniques. Hogan, Jarrow, and Warachka describe the dynamics of trading profits as a stochastic process. A test for statistical arbitrage can then be based on identification of the parameters of the process. This project implements such a test, and experiments on interest rates of deposits, FRA, and swap contracts from 2002 to 2005 by a defined trading strategy. We observe only one statistical arbitrage opportunity on the market by this trading strategy from test results.

Sammanfattning

Arbitrage är en handelsstrategi som utnyttjar prisskillnader mellan identiska tillgångar. Den innebär riskfri vinst och finns definitionsmässigt inte på effektiva marknader. Statistisk arbitrage är ett mindre restriktivt begrepp, som söker eventuella prissättnings-ineffektiviteter på marknaden som den ser ut. Hogan, Jarrow och Warachka modellerar handelsstrategin som en stokastisk process, och sökning efter statistisk arbitrage kan baseras på identifikation av process-parametrarna. I projektet har vi implementerat ett sådan test och prövat på svenska historiska data från 2002 till 2005. Data utgörs av dagspriser på olika finansiella instrument, som in/utlåningsräntor, "forward rate agreements", and bytes-kontrakt ("swaps"). Endast ett tillfälle kunde upptäckas, och med osäker signifikans. Det visade sig att data för marknader med store rörelser, som yen och USD, torde vara bättre testobjekt.

Acknowledgement

I would like to express my gratefulness to my brilliant advisor Dr. Claes R. Cramer, for your ideas and guidance throughout the work at Swedbank. Especially I would like to thank you for helping me on understanding all the financial theories, which give me a lot of inspirations and thoughts.

I would also like to express my thanks to my advisor Prof. Jesper Oppelstrup at Nada for your helps on numerical analysis, but more importantly for your kind encouragement during the work, making the completion of this thesis possible.

I would like to thank my dear friend Heng Chen for reading my thesis and all your precious suggestions for my report writing.

I would like to thank my parents. Thank you for everything, and I would never go this far without you.

Contents

1. Introduction

This project implements the test for statistical arbitrage proposed by S. Hogan, R. Jarrow, and M. Warachka [1] on derivative market outcomes using experimental data from the Swedish market. In mathematical financial theory, a portfolio is a collection of financial instruments traded on the market. As a portfolio trading strategy, Statistical Arbitrage is a quantitative and computational approach. There exists a variety of automated trading systems which commonly employ data mining, statistical methods, and artificial intelligence techniques. Some popular strategies such as "pair trading" and "arbitrage merging" on stocks have been used to create profits [19]. Here we focus on detecting statistical arbitrage opportunities by comparing new and historical data.

1.1 A short history of finance – from a mathematician's view

In 1900, Louis Bachelier described the variation of stock price using Brownian motion in his dissertation "Theorie de la Spéculation". He is the first to anticipate much of what later became standard fare in financial theory : random walk of financial market prices, Brownian motion and martingales, before both Einstein and Wiener [17]. Unfortunately, this first paper in the history of Mathematical finance was widely recognized only in the 1950's.

The modernization of finance would date to the year 1952 with the publication in Journal of Finance of Harry Markowitz's article, "Portfolio Selection". In this remarkable paper, Markowitz gave a precise definition of return and risk, as the mean and variation of the outcome of an investment. This therefore made the powerful methods of mathematical statistics available for the study of strategy of portfolio selection.

An issue that is the subject of intense debate among academics and financial professionals is the Efficient Market Hypothesis (EMH), which states that security (such as stock) prices fully reflect all available information at any time. This implies that there is no arbitrage opportunity in a perfectly "efficient" market, i.e., one can neither buy securities which are worth more than the selling price, nor sell securities worth less than the selling price. A significant development of EMH in 1960's by Eugene Fama [12] and his later work in 1998 [13] asserts that price movements in the market are unpredictable. The "Random Walk Theory" can be connected to Bachelier's work in 1900. Over more than half a century, much empirical research was done on testing the market efficiency, which can be traced to 1930's by Alfred Cowles [14, 15]. Many studies have found that stock prices are at least partially predictable [16]. The contradiction to market efficiency gives the possibilities to search for a statistical arbitrage opportunity. S. Hogan, R. Jarrow, and M. Warachka [1] demonstrate a method to test the existence of statistical arbitrage, and proved that it is incompatible with market efficiency.

1.2 Description of the project

This project implements the statistical method from [1] and experiments on real market data. Before we come to the description of the test, some background discussion about arbitrage and statistical arbitrage is necessary.

An *arbitrage* is a transaction or portfolio that makes a profit without risk. A portfolio is said to be an arbitrage if it costs nothing to implement, has a positive probability of a positive payoff, and a zero probability of a negative payoff. Loosely speaking, "buy low" and "sell high" trade.

To relax the condition of arbitrage, we can define a *statistical arbitrage*; in other words, an arbitrage is a special case of statistical arbitrage. One major distinction is that a statistical arbitrage is not riskless. Like an arbitrage, a statistical arbitrage costs nothing to implement. However, it has a positive expected payoff and a zero probability of a negative payoff only as time approaches infinity, and it's variance vanishes at time infinity.

The test for statistical arbitrage opportunity is applied to one trading strategy; specifically, the profits generated from the trading strategy every business day. This evaluation of profits of the trading strategy is based on a period of market data from the current trading day. It is believed that a statistical arbitrage opportunity appears as an abnormal behavior in the market. It then becomes essential to distinguish whether an abnormal market figure is a "potential opportunity" or simply erroneous data. This operation will be called "data cleaning". We use a 10 years yield curve for each day. A yield curve is a graphic representation of market value (yield) for a security plotted against the maturity of the security, the data to build the yield curve will be discussed shortly. We apply a trading strategy, valuate the profits generated based on these standard yield curves, and determine whether there exists a statistical arbitrage opportunity by the method proposed by S. Hogan, R. Jarrow, and M. Warachka. [1]. The profits generated from this trading strategy is described by a stochastic process with four parameters. A statistical arbitrage occurs when the identified four parameters satisfy the suggested conditions from [1].

To test our implementation, we use a standard 10 years yield curve for each day. This is constructed by employing a 1-month, 2-month, and 3-month deposit, 8 FRA contracts, and 10 swap contracts with maturities from 1 year to 10 years (definition for FRA and swap can be found in section 3.2.2 and 3.2.3 respectively). The interest rates for each instrument are obtained for every business day from 26 March 2002 to 8 August 2005 from the Swedish market. The data is acquired from Swedbank Market at Föreningsparbanken, Sweden. We have later concluded that a better choice of market data could be obtained from a more liquid market such as dollar, sterling, euro, or yen. We call those markets more liquid since they have more transactions traded each day, hence we expect to observe more statistical arbitrage opportunities in those markets.

1.3 Overview of this paper

- Chapter two gives a short description of "data cleaning", Savitzky-Golay filtering and how to use this smoothing technique to remove errors from the original data.
- Chapter three introduces the yield curves, the relevant financial theories involved, and the valuation of the trading strategy.
- Chapter four gives the mathematical derivation of the test for statistical arbitrage, and its parameter estimation problem, which is solved by minimizing the log-likelihood function by the Nelder-Mead algorithm or a Quasi-Newton method.
- Chapter five concludes the work and makes some further suggestions.

2. Data Cleaning

To test the model, we need to employ historical data from the market. Such data is used also to adjust the parameters for some of the algorithms involved on the implementation. Since real arbitrage opportunities are not expected to occur frequently, the performance of the testing can be severely affected by even a few erroneous data. These errors, also called “outliers”, might come from mis-typing, or a faulty operation during the data transfer. Since we are dealing with a stochastic time series, any relatively large variation can be from an error. It therefore becomes significant to eliminate these errors before actually testing on the historical data.

In this chapter we discuss using a low-pass filter, to search for the errors in the data. This requires setting parameters for the tolerance level between errors and accepted variation. These parameters can be configured through data training.

Section 2.1 gives an overview of the task for data cleaning. Section 2.2 introduces a filtering method called “*Savitzky-Golay filtering*”, which is widely used for outlier detection. Section 2.3 discusses the development of the cleaning procedure.

2.1 Remove outliers

We will exercise our data cleaning on the daily swap rates for one year. The swap rates are illustrated on figure 2.1.

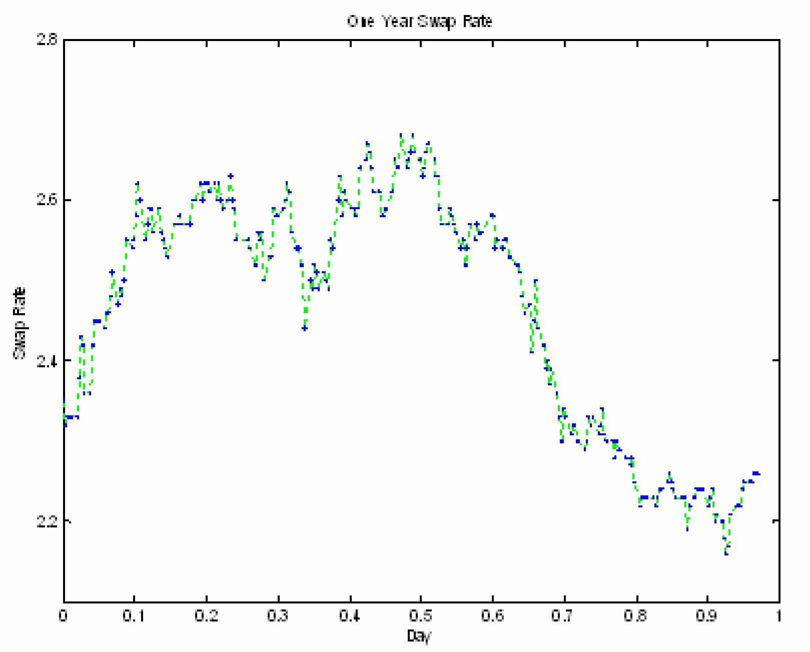

Figure 2.1. One year swap rate with erroneous data

The task is to find the “general trend curve” for the time series, then calculate the “spread”, the distance between each point and the trend curve. For this reason, we use low-pass filtering to find the trend. The most popular choice is the Savitzky-Golay low-pass smoothing filter.

2.2 Savitzky-Golay Smoothing Filters

Savitzky-Golay filtering can be thought of as a generalized moving average. The filter coefficients are derived by performing an unweighted linear least squares fit using a polynomial of a given degree. Assume we want to smooth a series of data points f_i, $i = 0,1,\ldots,$M. This filter replaces each data value f_i by a linear combination g_i of itself and some number of nearby neighbors.

$$g_i = \sum_{n=-n_L}^{n_R} c_n f_{i+n} \tag{2.1}$$

Here n_L is the number of points used "to the left" of a data point i, while n_R is the number of points used "to the right". The idea of Savitzky-Golay filtering is to find filter coefficients c_n that approximate the underlying function within the moving window not by a constant (whose estimate is the average), but by a polynomial of higher order, typically quadratic or quartic. For each point f_i, we least-squares fit a polynomial to all $n_L + n_R + 1$ points in the moving window, and then set g_i to be the value of that polynomial at position i. Consider how g_0 might be obtained: We want to fit a polynomial of degree M in i, namely $a_0 + a_1 i + \ldots + a_M i^M$ to the values $f_{-nL},\ldots,f_{nR}$. Then g1 will be the value of that polynomial at $i = 0$, namely a_0. The normal equations in matrix form for this least square problem is

$$(A^T \times A) \times a = A^T \times f \quad or \quad a = (\mathrm{A}^\mathrm{T} \times \mathrm{A})^{-1}(\mathrm{A}^\mathrm{T} \times f) \tag{2.2}$$

where

$$A_{ij} = i^j \qquad i = -n_L,\ldots,n_R, \quad j = 0,\ldots,M$$

Since the coefficient c_n is the component a_0 when f is replaced by the unit vector e_n, $-n_L \le n < n_R$, we have

$$c_n = \left\{(A^T \cdot A)^{-1} \cdot (A^T \cdot e_n)\right\} \tag{2.3}$$

We make no use of the value of the polynomial at any other point. When we move on to the next point f_{i+1}, we do a whole new least-squares fit using a shifted window. [9] A higher degree polynomial makes it possible to achieve a high level of smoothing without attenuation of "real" data features. This is illustrated in figure 2.2.

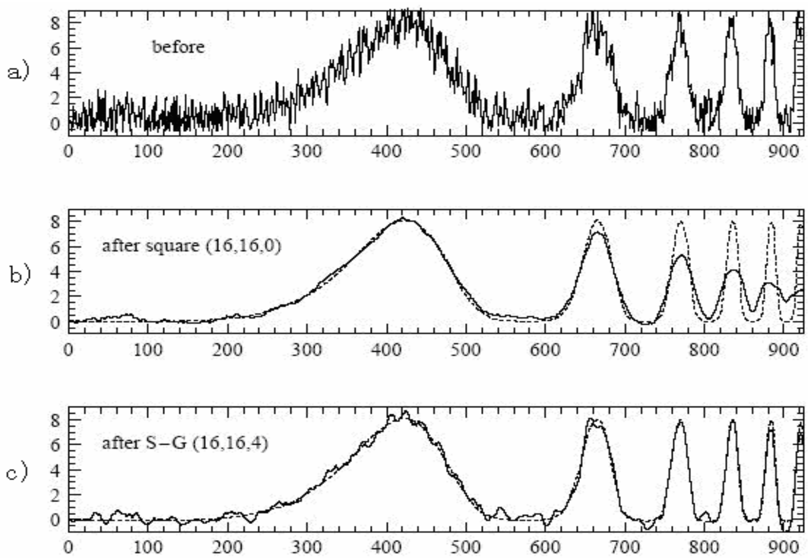

Figure 2.2 Curve smoothing by Savitzky-Golay filtering with quadratic and quartic polynomial.

Graph a) is a synthetic noisy time series consisting of a sequence of progressively narrower bumps, with additive Gaussian white noise. Graph b) is a simple moving average. The window extends 16 points leftward and rightward, for a total of 33 points. Note that narrow features are broadened and suffer corresponding loss of amplitude (we miss the actual trends!). The dotted curve is the underlying function used to generate the synthetic data. Graph c) shows a Savitzky-Golay smoothing filter of degree 4 using the same 33 points. While there is less smoothing of the broadest feature, narrower features have their heights and widths preserved [9].

2.3 Finding outliers by smoothing

We use Savitzky-Golay filter to smooth the swap rate data in figure 2.1. After some experiments, the filter is by degree 1 and span size 3. We shall discuss how to select the optimized degree and span size shortly. On figure 2.3, the dots are the original market data; the curve is the corresponding smoothed data, which represents the trend of the market data. We can clearly see the spread for each market data point from the trend. We then take the absolute value of all the spread which is illustrated from the second graph of figure 2.3.

The search for outliers uses the histogram of $(f_i - \overline{f_i})$ with M = *10* bins of equal width, see figure 2.4. We label a threshold T and define all f_i with $|f_i - \overline{f_i}| > T$ to be outliers. The next question is how to select the value M, and the threshold T. Suppose we are given a set of market data which contain previously known errors. Adjust M and T until we find proper pairs of M and T which can successfully find all the errors. We then can tune the parameters with more historical data from the same market. We can have an over-determined solution for the value of M and T by enough training data provided.

For this case, the partition size and threshold is 10 and 0.0327 respectively, gives a satisfactory result. Consequently we have 4 errors found, which are pointed out on figure 2.3.

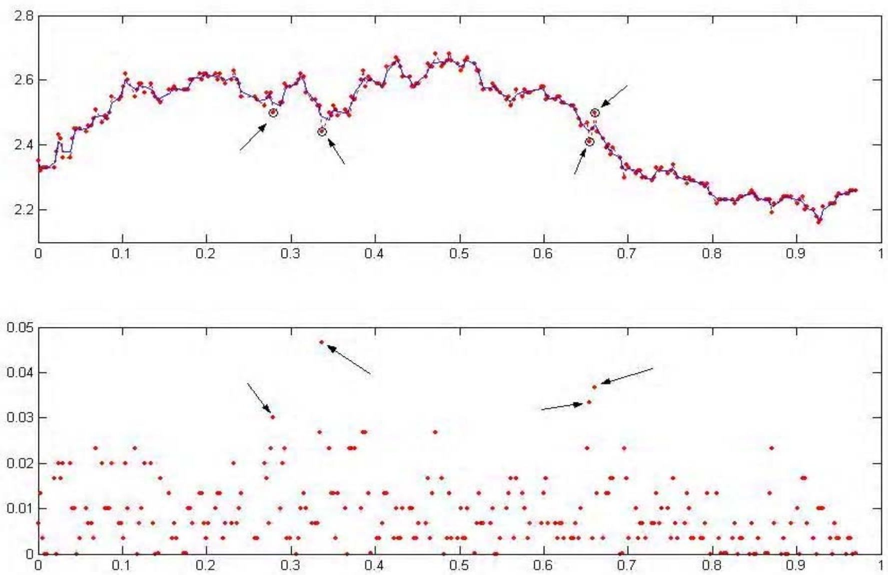

Figure 2.3. Remove outliers

Spread	0.0047	0.0093	0.0140	0.0187	0.0233	0.0280	0.0327	0.0373	0.0420	0.0467	>
Count	95	46	63	15	13	13	1	2	0	0	1

Table 1. Distribution of absolute value of residuals

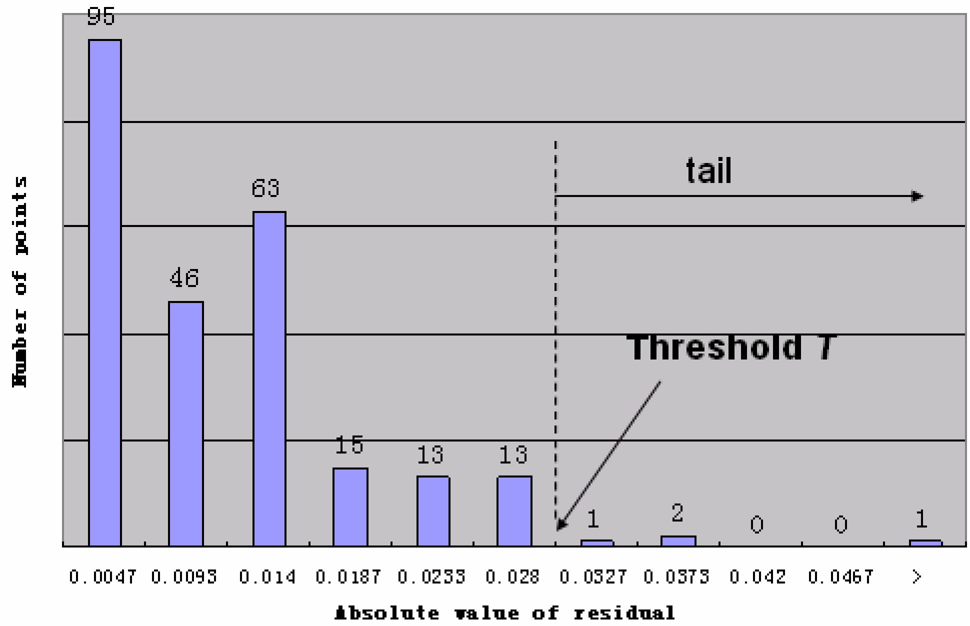

Figure 2.4. Residual Distribution

Outliers are replaced by interpolation. On the market, one common way to deal with error data is to replace it with the previous data, i.e., zeroth order interpolation. This method neglects the trend, while we usually expect movements on a liquid market. We can alternatively higher order interpolation. Higher order interpolation involves more points. If there is another erroneous point nearby, we may have an incorrect replacement. Implementation of the routine is described in appendix A.

Instead of utilizing much training data, an alternative to search for T is to iteratively smooth the data points. Step one, we choose a start T, say, T_0, and smooth the data according to M_0 and T_0. Second step we stop the iteration if the histogram has a "short tail", since we believe all the outliers are removed (see figure 2.4). Else we replace the outliers by interpolations, and repeat step one.

In the next chapter we will use the cleaned market to build standard yield curves to valuate the daily profits for one trading strategy.

3. Yield Curve Modeling

A yield curve is a graphic representation of market value (yield) for a security plotted against the maturity of the security. This chapter explains the yield curve, and valuation of the swap contract by using our experimental market data. I.e., the daily profit made from a swap contract in a certain time for a certain period. In the next chapter, we will use those profit values to test for statistical arbitrage.

Section 3.1 introduces the theory of future and present value of an investment. Section 3.2 introduces forward, FRA, swap contract, calculation of their interest rate, discount and the concept of yield curve. Section 3.3 gives the calculation of the daily profits from swap.

3.1 Financial Arithmetic

3.1.1 Future values

The *future value* of a sum of money (the *principal*) invested at a given annual rate of interest will depend on

1. whether interest is paid only on the principal (known as *simple interest*), or in addition on the interest that accrues (known as *compound interest*).
2. In the case of compound interest, the frequency with which interest is paid. For instance, annually, semi-annually, quarterly, monthly, daily, continuously.

With simple interest, the future value is determined by

$$F = P(1 + rT) \tag{3.1}$$

where

P = principal amount

F = future value

r = rate of interest

T = number of year

With compound interest, the future value (known as *compound value*) is determined by

$$F = P(1 + r)^T \tag{3.2}$$

In the case, interest is earned on the interest that accrues annually for T years. Sometimes compounding takes places semi-annually, quarterly, or even more frequently, say, *m* times per year. Then at the end of T years mT interest payments will have been made and the future value of the principal will be

$$F = P(1 + r/m)^{mT} \tag{3.3}$$

In the limit (m approach infinity) the interest is compounded continuously. This limit is derived as

$$\begin{aligned} F &= P[(1 + r/m)^{m/r}]^{rT} \\ &= P[(1 + 1/n)^n]^{rT} \end{aligned} \tag{3.4}$$

where n = m/r. Since $e = \lim_{n\to\infty}(1+1/n)^n$, we have

$$F = Pe^{rT} \tag{3.5}$$

in the case of *continuous compounding*.

3.1.2 Present values

If an amount F is to be received in T years' time, the *present value* of that amount is the sum of money P which, if invested today, would generate the compound amount F in T years' time. The process of finding present values is known as *discounting* and is the exact inverse of the process of finding future values. If we assume F to be the unit $1, the present value is known as *discount factor / discount function*. Obviously, the discount factor is always equal or less than 1.

With annual discounting, a sum of money F to be received in T years' time has a present value of

$$F = P(1 + r)^T \tag{3.6}$$

The T-year discount factor D is

$$df = (1 + r)^{-T} \tag{3.7}$$

Similar to the future value, we consider more frequent compounding. If discounting takes place m times a year, then the present value will be

$$P = F(1 + r/m)^{-mT} \tag{3.8}$$

The corresponding continuous discount factor is

$$df = e^{-rT} \tag{3.9}$$

3.2 Deposit, FRA and Swap contract

3.2.1 Forward rate and FRA contract

A *forward* contract is an agreement on selling or buying an asset in a certain future time for a certain price. A *forward interest rate* is the interest rate implied by current spot rates (the rates on the real market) for a specified future time period. For example, the *n-year spot rate* is the rate of interest earned on an investment that starts today and lasts for n years. If the five-year spot rate is 4% per annum, an investment of $100 for five years would grow to $100 \times 1.04^5 = 121.67$.

A *forward-rate agreement* (FRA) is a forward contract where the parties agree that a certain interest rate will apply to a certain principal during a specified future time period. The discount factor for FRA contract df_{fra} in the period between t_i and t_{i-1} is:

$$df^i_{fra} = \frac{1}{1 + r^i_{fra}(t^i_{fra} - t^{i-1}_{fra})/360} \tag{3.10}$$

where

t^{i-1}_{fra} is the starting date of the FRA contract in days.

t^i_{fra} is the current date of the FRA contract in days.

r^i_{fra} is the FRA rate.

df^i_{fra} is the discount factor of the FRA between t^{i-1}_{fra} and t^i_{fra}.

3.2.2 Swap contract

A *swap* is a contract by which two parties agree to exchange two cash flows with different features for a certain period. For example, Steven must pay a floating interest rate on his house mortgage issued by bank A. However, he would rather be paying a fixed rate due to the unpredictability for the floating interest rate each month. He could request a swap contract from bank B, which trades swaps; that he would be paying bank B a fixed amount of interest each month, while in return bank B would be paying the floating interest to bank A.

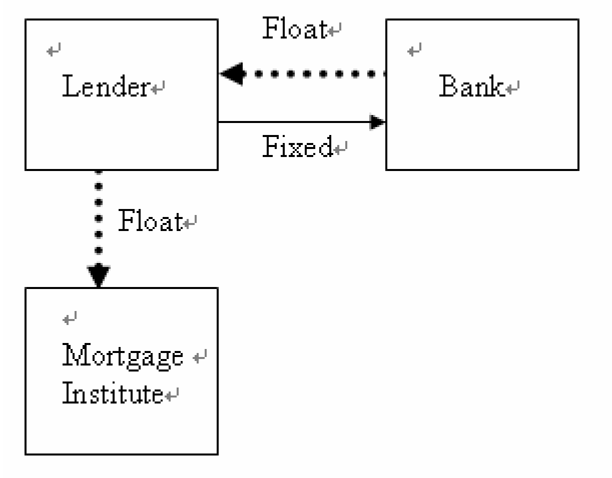

Figure 3.1 The buyer pays fixed interest and receives floating interest

Suppose we have swaps for a period of *N* years, the fixed rate S_i of the swap is:

$$s_i = \frac{1 - df_{swap}^i}{\sum_{j=1}^{i} df_{swap}^j} \qquad (3.11)$$

where df_{swap}^i is the *discount factor* for swap. where discount factor for swap contract on at t_i is:

$$df_{swap}^i = \frac{1 - r_{swap}^i \sum_{j=1}^{i-1} df_{swap}^j}{1 + r_{swap}^i} \qquad (3.12)$$

3.2.3 Yield curve

The statistical arbitrage model is based on analyzing daily profits generated from trading swap contracts in a certain time for a certain period. To estimate the profits, we need a set of standard daily yield curves with different maturities over a relatively long period. A yield curve is the curve obtained by plotting the yield, and hence the price, of financial instruments with different maturities against the maturity values. Typically, the yields for longer maturities are higher than the yield for shorter maturities, thereby result with an upward sloping as shown on figure 3.2 a. However, sometimes the short-term deposit yields more than the long-term FRA, and the curve is downward sloping, which is called *inverted*. The corresponding curves for discount factors is obtained by replacing the rates with the corresponding discount factor. The discount curve usually starts from value 1 if we calculate based on the maturity from the current day.

Standard yield curves for t his project are obtained from a 1-month, 2-month, and 3-month deposit, 8 FRA contracts, and 10 swap contracts with maturities from in 1 year to in 10 years. The interest rates for each instrument are obtained for every business day from 26 March 2002 to 8 August 2005 from the Swedish market, see figure 3.2. Our experiment for the test of statistical arbitrage requires discount or interest rate value on a daily change, we therefore need interpolation between any two consecutive points. An implementation of B-Spline interpolation is discussed in appendix A.

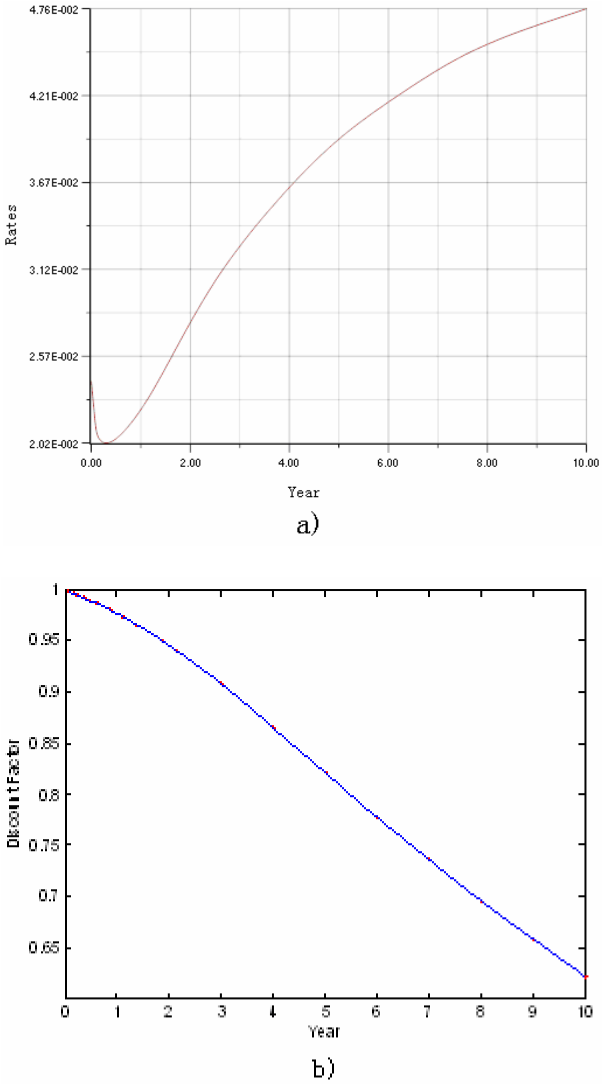

Figure 3.2 One typical yield curve for interest rate versus maturities (a). The corresponding discount factor curve versus maturities (b).

3.3 Evaluation of trading profits

The test of statistical arbitrage is based on analyzing the profits of a trading strategy. The strategy employed for this project is to trade a swap contract in a certain future time for a certain rate. For this purpose, we trade a forward contract at current day on the future swap rate. The swap contract has a quarterly cash flow for the floating payment. Since we can calculate the theoretical forward rate, and hence the swap rate. We need to calculate the profit for each day from current day (the day we trade the forward contract), up to the maturity of the swap contract.

Test of statistical arbitrage is executed on a daily count. From our experimental data, we have a ten years yield curve and corresponding discount curve, as one shown on figure 3.2. Suppose we trade the swap in six months and it matures after one year. We call it a "six month – one year swap contract". Suppose the day count is by unit of year. Then the current time $t_0 = 0$, the trading day of the swap is $t_1 = 0.5$, and the maturity of the swap is $t_2 = 1.5$. The discount value for T_0 is 1, the discount value for t_1 and t_2 are obtained from the ten years yield curve, say, df_1, df_2. The swap rate r_s is

$$r_s = \frac{-\ln(df_2 / df_1)}{t_2 - t_1} \tag{3.13}$$

Suppose we have the principal P. Before entering the swap, the profit v_i at day i is

$$v_i = (df_1 - df_2 - r_s df_2)P \tag{3.14}$$

After entered the swap contract, we need to price the swap. Since it is a one year swap with quarterly payment for floating sides, there are totally four cash flows after each three months. The profit is the difference between the value from fix side and float side. The payment from fix side can be easily calculated by the swap rate.

$$V_i^{Fix} = r_s df_i P \tag{3.15}$$

where

V_i^{Fix} is the payment (value) from the fix side

r_s is the swap rate

df_i is the discount factor for day i

P is the principal

And the payment from float side is sum of the floating cash flow in the future until the swap maturity.

$$V_i^{Float} = P\sum_{k=q}^{4} r_k \cdot (1/4) \cdot df_k \tag{3.16}$$

V_i^{Float} is the payment (value) from the float side

q is the number of next cash flow

r_s is the swap rate for k^{th} cash flow

df_i is the discount factor for k^{th} cash flow

P is the principal

Then the final profit if we receive fix and pay float is $v_i = V_i^{Fix} - V_i^{Float}$, and is $v_i = V_i^{Float} - V_i^{Fix}$ if we pay fix and receive float for the swap contract.

4. Test For Statistical Arbitrage

This chapter discusses the test of statistical arbitrage proposed by S. Hogan, R. Jarrow, and M. Warachka [1]. Sections 4.1 and 4.2 introduce the mathematical notion of statistical arbitrage and its implementation. Sections 4.3 and 4.4 present two numerical approaches to the implementation. Finally in section 4.5 we investigate this statistical arbitrage strategy by applying a swap contract for historical data.

4.1 Definition of statistical arbitrage

Let the stochastic process $(x(t), y(t) : t \geq 0)$ represent *trading strategy* involving *x(t)* units of stock and y(t) units of a market account. A market account is an investment that is initially randomly sampling changes in market variables in order to value a derivative. Define this trading strategy to be *zero initial* cost and *self-financing*. Zero initial cost means $x(0)S_0 + y(0) = 0$, where S_t is the price of stock, which can be interpreted as a portfolio of traded assets. Self-financing means no cash inflow or outflow after constructing the trading strategy. For example, the trading strategy consist a swap contract and a deposit account. Profits made from the trading strategy are put into the deposit account, and any further investments are paid by the deposit account. This trading strategy is formulated using only available information. Let the process *V(t)* denote the cumulative trading profits after the inclusion of transaction costs at time *t* that are generated by such as trading strategy $(x(t) : t \geq 0)$.

To motivate the definition of statistical arbitrage, we consider two examples of a portfolio of stocks. The first example illustrates a trading strategy that would not represent long horizon excess profit opportunities. The second example, in contrast, demonstrates how *time diversification* over a long trading horizon generates an arbitrage opportunity by imposing an additional condition on the variance of trading profits.

Example 1 Consider the classic model of stock price

$$\Delta S_i = rS_i \cdot \Delta t + \sigma \sqrt{\Delta t} z_i$$

i : = 1,2,... the day number

△Si : daily incremental profit generated by the portfolio

zi : are i.i.d. *N*(0,1) random variables

r : deterministic part

σ : volatility

Motivated by the classic model of stock price, we later assume a model for profits of our trading strategy.

□

Example 2 Suppose that the discounted trading profits over an intermediate trading interval $[t_{k-1}, t_k]$ can be represented as

$$v(t_k) - v(t_{k-1}) = \mu + \sigma z_k$$

where $\mu, \sigma > 0$ and z_k are independent identically distributed (i.i.d.) random variables with zero mean and variance *1/k*. This trading strategy has positive expected discounted profits over every interval (μ), but with random noise (σz_k) appended. The variance of the noise is decreasing over time. Again, $v(t_0)=0$ and the cumulative discounted trading profit at time t_n equals

$$v(t_n)=\sum_{k=1}^{n}[v(t_k)-v(t_{k-1})]=\mu n+\sigma\sum_{k=1}^{n}z_k$$

We have $E^P[v(t_n)]=\mu n$ and $Var\ [v(t_n)]=\sigma^2\sum_{k=1}^{n}\frac{1}{k}$. Therefore

$$\frac{Var^P[v(t_k)]}{n}=\sigma\frac{\sum_{k=1}^{n}\frac{1}{k}}{n}\to 0 \text{ as } t\to\infty$$

We say this trading strategy is "diversifiable".

As we will see shortly, it can also be shown that $P(v(t_n)<0)\to 0$ as $n\to\infty$, implying that the discounted trading profits are non-negative in the long time.

Across time, the random noise in this trading strategy is "diversifiable", thereby generating a limiting arbitrage opportunity. Indeed, this trading strategy has a positive discounted expected profit with a time averaged variance approaching zero. Over a long time horizon, this trading strategy represents a "statistical arbitrage".

□

As we see from this example, positive expected profits alone are not sufficient to declare a statistical arbitrage due to the unknown variation, i.e., risks; rather, more restrictive conditions must be provided to define the statistical arbitrage:

Definition 1 A *statistical arbitrage* is a zero initial cost, self-financing trading strategy $(x(t):t\geq 0)$ with *cumulative discounted value v(t)* such that:

1. $v(0)=0$
2. $\lim_{t\to\infty}E^P\left[v(t)\right]>0$
3. $\lim_{t\to\infty}\frac{Var^P\left[v(t)\right]}{t}=0$
4. $\lim_{t\to\infty}P(v(t)<0)=0$

□

Condition (1) of definition 1 implies it is a zero initial cost and self-financing trading strategy. Condition (2) suggests positive expected discounted profits. Condition (3) says a time averaged variance converging to zero, and Condition (4) gives zero probability of a loss

4.2 Basic Assumption

To test for statistical arbitrage, we begin by assuming the following stochastic process for incremental trading profits

$$\Delta v_i=\mu i^{\theta}+\sigma i^{\lambda}z_i \tag{4.1}$$

where z_i are i.i.d. $N(0,1)$ random variables. The time between equidistant increments is set to one; therefore i = 1,2,… is the day number. Reasons for imposing this assumption are discussed in [1]. A set of Δv_i's are shown in figure 4.1

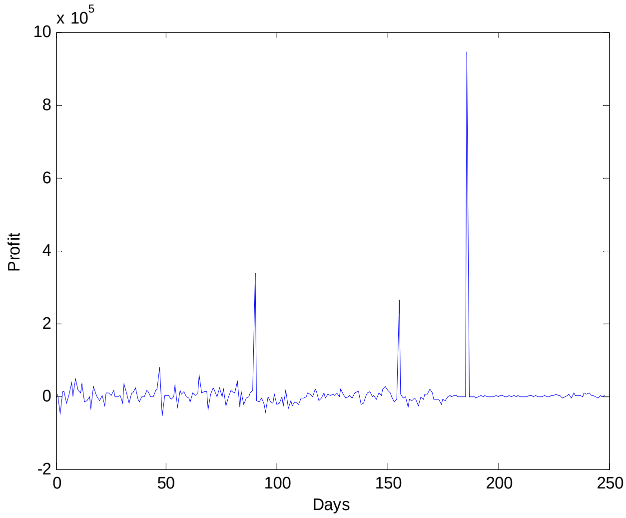

Figure 4.1 The daily profits generated from a trading strategy

A goodness-of-fit test for this model can be found in appendix B. This assumption of un-correlated normal error can be violated. However, this assumption has been shown to be dubious in empirical finance [2]. The assumption of a first order moving average process is discussed in appendix C. The implementation in this project will assume normality and independence. The initial quantities $z_0 = 0$ and Δv_0 are both zero by definition. The parameters σ and λ determine the volatility of incremental trading profits while the parameters μ and θ specify their corresponding expectation. [2]

This point estimate is done by the maximum likelihood method (MLE). We can express $z_i = (\Delta v_i - \mu i^\theta)/\sigma i^\lambda$ in term of observation v_i, i. Then we can test whether the observation can fit the normal distribution.

$$f(\Delta v_i \mid \mu,\theta,\sigma,\lambda) = \frac{1}{\sigma i^\lambda}\exp\left(\frac{(\Delta v_i - \mu i^\theta)^2}{2\sigma^2 i^{2\lambda}}\right) \tag{4.2}$$

By the assumption of independence, we have the likelihood function

$$\begin{aligned} L(\Delta v \mid \mu,\theta,\sigma,\lambda) &= \prod_{i=1}^{n} f(\Delta v_i;\mu,\theta,\sigma,\lambda) \\ &= \prod_{i=1}^{n}\frac{1}{\sigma i^\lambda}\exp\left(\frac{(\Delta v_i - \mu i^\theta)^2}{2\sigma^2 i^{2\lambda}}\right) \end{aligned}$$

[3]

And the corresponding log likelihood is

$$\log L(\Delta v_i \mid \mu,\theta,\sigma,\lambda) = -\frac{1}{2}\sum_{i=1}^{n}\log\left(\sigma^2 i^{2\lambda}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{1}{i^{2\lambda}}\left(\Delta v_i - \mu i^\theta\right)^2 \tag{4.3}$$

To estimate the four parameters σ^2, λ, μ, and θ, we solve the MLE problem by minimization algorithms discussed on sections 4.3 and 4.4. We can also calculate the first derivative with respect to them:

$$\frac{\partial \log L(\Delta v_i;\mu,\theta,\sigma,\lambda)}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{1}{i^{2\lambda}}\left(\Delta v_i i^{\theta} - \mu i^{2\theta}\right) \tag{4.4}$$

$$\frac{\partial \log L(\Delta v_i;\mu,\theta,\sigma,\lambda)}{\partial \sigma^2} = -\frac{n\sigma^2}{2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\frac{1}{i^{2\lambda}}\left(\Delta v_i - \mu i^{\theta}\right)^2 \tag{4.5}$$

$$\frac{\partial \log L(\Delta v_i;\mu,\theta,\sigma,\lambda)}{\partial \theta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{1}{i^{2\lambda}}\left(\mu^2 \log(i) i^{2\theta} - \Delta v_i \mu \log(i) i^{\theta}\right) \tag{4.6}$$

$$\frac{\partial \log L(\Delta v_i;\mu,\theta,\sigma,\lambda)}{\partial \lambda} = -\sum_{i=1}^{n}\log(i) + \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\log(i)}{i^{2\lambda}}\left(\Delta v_i - \mu i^{\theta}\right)^2 \tag{4.7}$$

The following proposition is the motivation for the empirical tests of statistical arbitrage on incremental trading profits under definition 1.

Proposition 1 A trading strategy generates a statistical arbitrage if the following conditions are all satisfied

1. $\mu > 0$
2. $\lambda < 0$
3. $\theta > \max(\lambda - \frac{1}{2}, -1)$

Which together with the MLE gives a constrained minimization problem: *min Log(L)* s.t. $\mu > 0$, $\lambda < 0$ and $\theta > \max(\lambda - \frac{1}{2}, -1)$ [2].

4.3 The Nelder-Mead algorithm

4.3.1 Description

The first approach to solve the minimization problem is the Nelder-Mead algorithm [9] (*fminsearch* function in Matlab). This method minimizes a function of n variables by the comparison of function values at the (n+1) vertices of a general simplex (in our case, n = 4), followed by the replacement of the vertex with the highest value by another point [5]. Four parameters must be specified to define a complete Nelder-Mead method: coefficients of *reflection* (p_r), *expansion* (p_x), *contraction* (p_c), and *shrinkage* (p_s). And those parameters must satisfy

$$p_r > 0,\ \ p_x > 1,\ \ p_x > p_r,\ \ 0 < p_c < 1, \text{ and } \ 0 < p_s < 1 \tag{4.8}$$

The default choices of those parameters is

$$p_r = 1,\ \ p_x = 2,\ \ p_c = \frac{1}{2}, \text{ and } \ p_s = \frac{1}{2} \tag{4.9}$$

The Nelder-Mead method is started with (n+1) points, defining an initial simplex. Suppose the point P_0 as the initial point, the other n points are

$$P_i = P_0 + \omega e_i,\ \ i = 1 \ldots n \tag{4.10}$$

where the e_i's are n unit vectors, and ω is our guess of the problem's characteristic length scale. We define $\omega = \|P_0\|_2 / n$ for this MLE problem.

Since the Nelder-Mead algorithm is for unconstrained minimization, we need to reformulate the three conditions from proposition 1.

1. $\mu = \tilde{\mu}^2$

2. $\lambda = -\tilde{\lambda}^2$

3. $\theta = \tilde{\theta}^2 + \max(\lambda, -\frac{1}{2}) - \frac{1}{2}$

Condition 1 ensures μ is positive and condition 2 ensures λ is negative. Condition 3 is illustrated in figure 4.2.

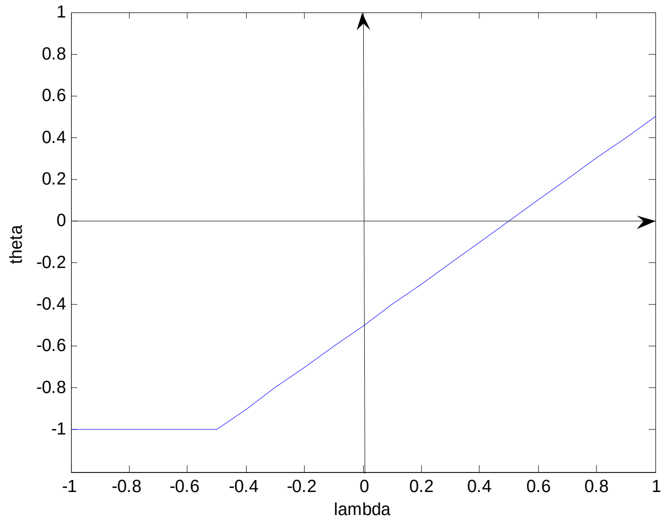

Figure 4.2 Reformulation of parameter θ

4.3.2 Result

The Nelder-Mead method is implemented in C#. To observe the performance of the estimation, we generate a series of simulations according to (4.1) by given sets of $\mu_t, \theta_t, \sigma_t, \lambda_t$, which satisfy the three conditions, i.e. there exists a statistical arbitrage. Then we use Nelder-Mead method to find $\mu^*, \theta^*, \sigma^*, \lambda^*$ that maximize (4.2). A successful result of $\mu^*, \theta^*, \sigma^*, \lambda^*$ will be very close to $\mu_t, \theta_t, \sigma_t, \lambda_t$.

The starting P_0 is always set as $\mu_0 = 0.01,\ \theta_0 = 0.01,\ \sigma_0 = 0.01, \lambda_0 = -0.01$, we run four different tests.

	Target Parameters				**Result**			
Test	μ	θ	σ	λ	μ	θ	σ	λ
1	0.8	0.5	1.2	-0.5	0.7995	0.5001	1.0861	-0.482
2	0.5	0.2	1.5	-0.35	0.4932	0.2023	1.7937	-0.376
3	0.5	0.2	1.5	-0.95	0.5	0.2	1.7668	-0.9751
4	0.5	0.2	0.25	-0.45	0.4985	0.2004	0.2503	0.4508

Table 1.1 Four tests for parameter estimate

We can see the Nelder-Mead algorithm is rather a robust method for our minimization problem.

4.4 The Quasi-Newton Method

The Nelder-Mead method requires only the objective function itself, and requires many evaluations. Since we can easily obtain the analytical gradient and even the Hessian matrix of the objective function, a rapidly convergent scheme such as Quasi-Newton scheme can increase the performance. Unfortunately, the Quasi-Newton method [8] did not line up the expectation. Even with a "good" starting guess, it failed to converge.

To understand the failure of the Quasi-Newton method, we study the sensitivity of the objective to variation of the parameters. Since it is a four-dimensional minimization, we simplify the problem by fixing parameters θ and λ. we change the value of μ and σ, and computing the value of log-likelihood (1)

$$\begin{aligned} L(\Delta v \mid \mu,\theta,\sigma,\lambda) &= \prod_{i=1}^{n} f(\Delta v_i;\mu,\theta,\sigma,\lambda) \\ &= \prod_{i=1}^{n} \frac{1}{\sigma i^{\lambda}} \exp\left(\frac{(\Delta v_i - \mu i^{\theta})^2}{2\sigma^2 i^{2\lambda}}\right) \end{aligned} \tag{4.11}$$

The result is shown in figure 4.3

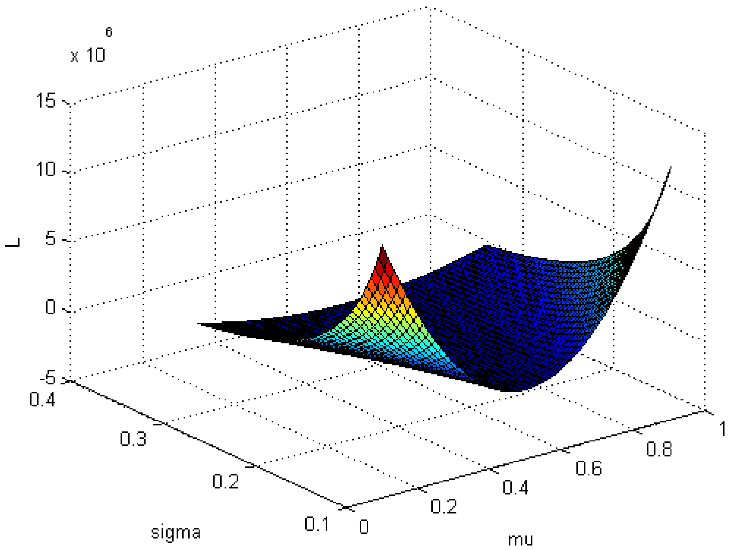

Figure 4.3 Surface graph of log-likelihood value against μ and σ

We see that there is clearly a minimal point along the μ axis. However the change in the σ direction at the minimum of μ is extremely small. To give better visualization, we draw the contour graph as on figure 2.

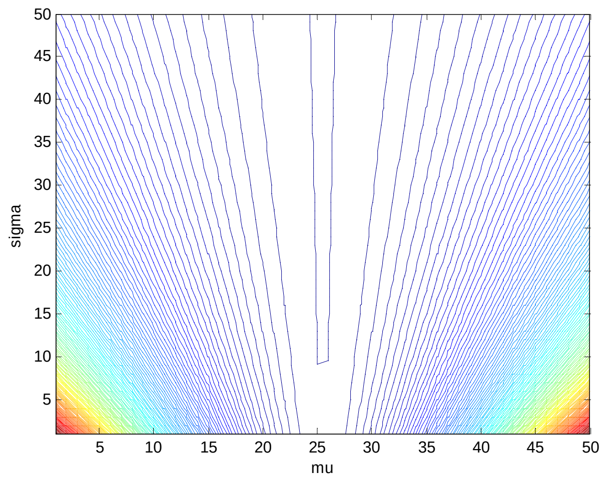

Figure 4.4 Contour Graph of log-likelihood value against μ and σ

There exists a minimum "valley" for the log-likelihood value on changing μ and σ. The Quasi-Newton method will be easily "fooled" in the valley by searching for the minimum of σ. Then we fix σ and λ, and plot the value of log-likelihood against μ and θ (figure 4.3 and 4.4). And similar difficulty is found for parameters θ and λ (see figure 4.5 and 4.6). Finally we conclude this parameter estimation problem is ill-conditioned. The Quasi-Newton method requires more parameter tuning than expected, and we use only the Nelder-Mead method.

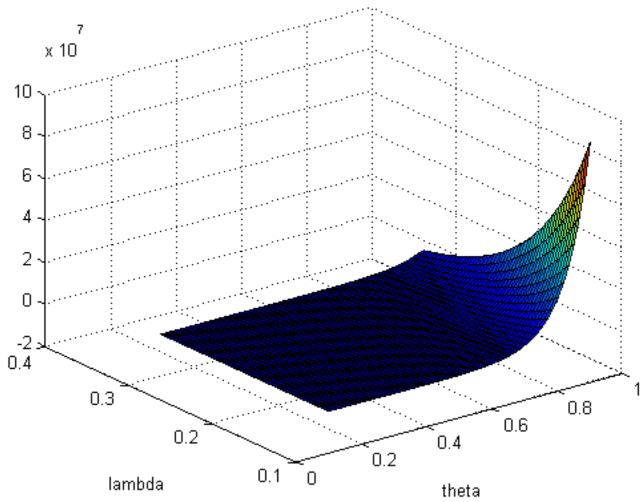

Figure 4.5 *Surface graph of log-likelihood value against* θ and λ

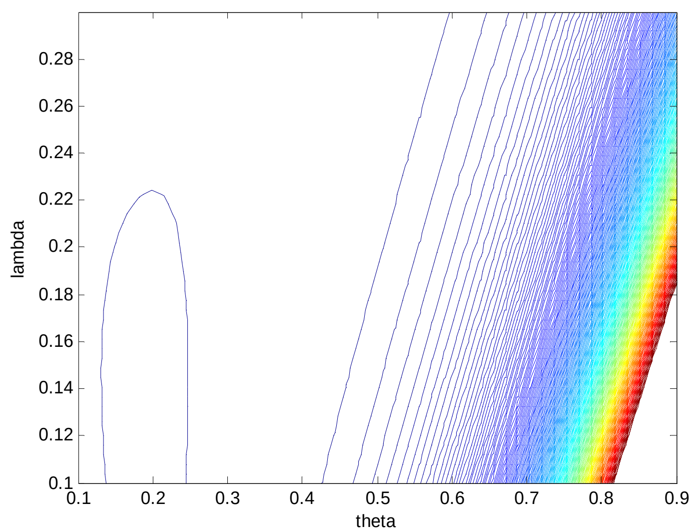

Figure 4.6 Contour graph of log-likelihood value against θ and λ

4.5 Test on experimental data

The standard 10 years yield curves for each day is constructed from interest rates of a 1-month, 2-month, and 3-month deposit, 8 FRA contracts, and 10 swap contracts with maturities from in 1 year to in 10 years, they are obtained for every business day from 26 March 2002 to 8 August

2005 from the Swedish market. Suppose in every day t_0 we trade a swap contract in certain time t_1, and this swap contract mature until time t_2. Therefore the contract period of the swap is $t_2 - t_1$. Define $T_1 = t_1 - t_0$ and $T_2 = t_2 - t_1$. By this assumption, we can generate a theoretical profit value for each day for certain T_1 and T_2. We trade on six different types of swap contracts as shown on table 2

Contract Type	T1	T2
One year - One year	1	1
Six months - One year	0.5	1
Three months - One year	0.25	1
One year - Six months	1	0.5
Six months - Six months	0.5	0.5
Three months - Six months	0.25	0.5

Table 2. Six different types of swap contract

Applying our implementation to those experimental data, one successful result is observed; which is on 24 April, 2002, by trading a three months – one year swap contract. The four parameters we acquired from the statistical model are:

μ	θ	σ	λ
61.015	1.2529	1.0281e+005	-1.2443e-007

Table 3. Four successful parameters observed for statistical arbitrage model

Those parameters satisfy the condition for statistical arbitrage defined on proposition 1. The daily profits for this trading is shown in figure 4.7:

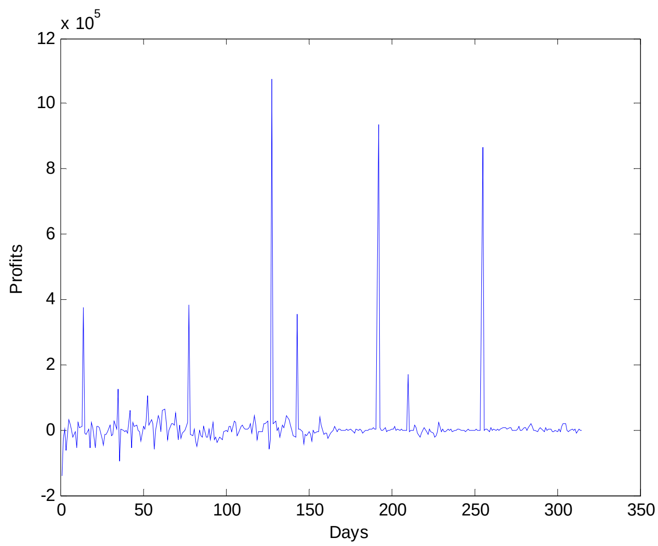

Figure 4.7 Daily profits for 3 months – 1 year swap contract at 24 April 2002

In figure 4.7, the unit for profits is Swedish kroner. Magnitude of parameter σ for volatility is so different from parameter μ and θ. Are these results significant?

5. Implementation

Data cleaning, yield curve building, valuation of various trading strategies, and parameter estimation are all implemented in C#. C# is an object-oriented programming language designed for building a wide range of enterprise applications. Since C# does not have a library for scientific computing, a library for matrix representation and operations, and some basic linear algebra routines are developed in C #. For data cleaning and parameter estimation, Savitzky-Golay smoothing filter and Nelder-Mead method are also implemented in C#. In order to utilize some complex linear algebra routines such as generalized least squares for yield curve building in section 3, the Lapack library in C is also interfaced to C#. We also developed a curve chart to display different yield curves, and some graphical user interfaces (GUI) for testing purpose. Some of the screenshots are shown below:

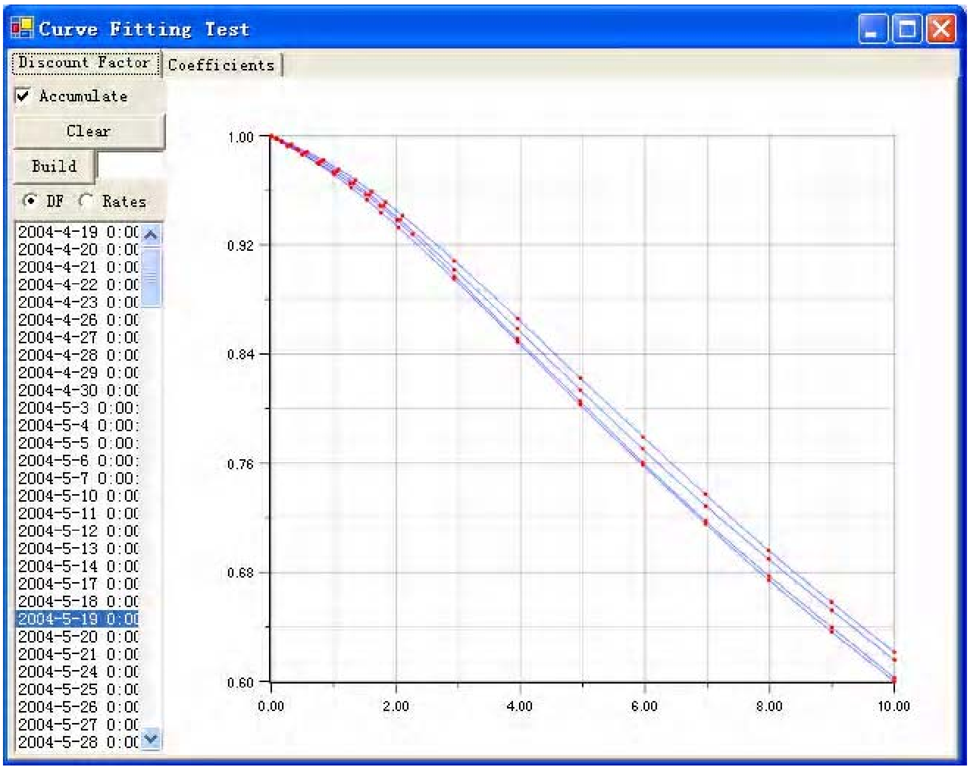

Figure 5.1 GUI for curve fitting test

Figure 5.2 GUI for data cleaning

In figure 5.1 it is an application to fit the yield curves fro each day by B-Spline. In figure 5.2 it is an application remove the outliers of one series of data points.

C# is widely used in the industry such as banking, online-stores. An implementation of mathematical model in this language can bring powerful tools developed in science to the industry. And make scientific application more easily commercialized.

6. Conclusion

Constructing financial trading strategy from use of different portfolio, we search for statistical arbitrage opportunity, by applying the test for statistical arbitrage. This project utilizes data cleaning, constructing yield curves on experimental data from Swedish interests market, and valuation of the trading strategy, to facilitate the test of statistical arbitrage. After six sets of experiments, this implementation observed one statistical arbitrage. Unlike the ordinary arbitrage strategy, trading based on a statistical arbitrage often make consistent profits but with occasionally spectacular losses [1]. It theoretically converges to an arbitrage in the limit, i.e., as time approaches infinity. The contribution of this project is the production of a mathematical model in the software package C#. Supported by carefully developed techniques including data cleaning and yield curve building, investor may apply this implementation to a trading strategy, and test whether there might exists a statistical arbitrage opportunity, or for a decision support purpose.

Further works for this project involves more experiments with more different portfolios, and with data from other financial market, such as sterling, dollar markets. Unfortunately our current experiments are not able to prove the significance of the results, and the consistence of our assumed model (4.1) with the real market.

Glossary

Arbitrage: An arbitrage is a transaction or portfolio that makes a profit without risk. A “buy low” and “sell high” trade

Forward: A forward contract is an agreement on selling or buying an asset in a certain future time for a certain time

FRA (Forward Rate Agreement): A FRA is a forward contract where the parties agree that a certain interest rate will apply to a certain principal during a specified future time period.

Maturity: The time when one security expires.

Portfolio: A portfolio is a collection of investments held by an institution or a private individual

Statistical Arbitrage: Statistical arbitrage, as opposed to arbitrage, is the mis-pricing of one or more assets based on the expected value of these assets. an arbitrage is a special case of statistical arbitrage.

Swap: A swap is a contract by which two parties agree to exchange two cash flows with different features.

Yield Curve: A yield curve is a graphic representation of market value (yield) for a security plotted against the maturity of the security

Reference

[1] S. Hogan, R. Jarrow, M. Warachka. *Statistical arbitrage and tests of market efficiency*. (2002) Working paper.

[2] Testing Market Efficiency using Statistical Arbitrage with Applications to Momentum and Value Strategies, (2004). Journal of Financial Economics 73, 525-565. With S. Hogan, R. Jarrow and M. Teo.

[3] W. Krzanowski. *An introduction to statistical modeling*. (1998) Edward Arnold.

[4] B. Øksendal. *Stochastic differential equations: An introduction with applications*. (1992) Springer-Verlag.

[5] J. A. Nelder and R. Mead. *A simplex method for function minimization*. Computer Journal 7 (1965), p 308-313.

[6] J. C. Hull. *Options, futures and other derivatives*. (2000) Prentice-Hall International.

[7] Alan Jennings and J.J. McKeown. *Matrix Computation*. (1992) Chichester : Wiley, cop.

[8] Lapack Documentation, The MathWorks.

[9] W. H. Press, S, A, Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in C*. Cambridge University Press. Ch 2.6, p 59-71.

[10] G. E. P. Box, G. M. Jenkins. *Time series analysis: forecasting and control*.(1970) Holden-Day.

[11] P. Priaulet. Exploring some alternative fixed-income strategies. HSBC-CCF & University of Evry (2004).

[12] Fama, E. F. (1965), Random walks in stock market prices, Financial Analysts Journal, 21 (5): 55-59.

[13] Fama, E. F. (1997). "Market Efficiency, Long-Term Returns, and Behavioral Finance," Center for Research in Security Prices Working Paper 448, University of Chicago.

[14] A. Cowles. *Stock market forecasting*. Econometrica, 12, (1944).

[15] A. Cowles. *A revision of previous conclusions regarding stock price behavior*. Econometrica, 28(4), (1960).

[16] B. G. Malkiel. *The efficient market hypothesis and its critics*. Working paper (2003).

[17] M. H. Millet. *The history of finance – an eyewitness account*. Journal of Portfolio Management, vol. 25 Issue 4 (1999).

[18] P. L. Bernstein. Capital ideas : the improbable origins of modern wall street. (1992) The Free Press.

[19] S. Reverre. The complete arbitrage deskbook. (2001) McGraw-Hill Professional.

Appendix

A Yield Curve By B-Spline Model

A.1 B-Spline Model

A.1.1 Basic assumption

The B-Spline model is a piecewise cubic curve fitting model. Suppose we have N data points to fit $\{(t_0, y_0), (t_1, y_1), ..., (t_{N-1}, y_{N-1})\}$. By given M knots $\{k_0, k_1, ..., k_{M-1}\}$, we want

$$f(\theta) = B\theta \quad \text{and} \quad \underset{\theta}{Min}\big(f(\theta) - y\big) \tag{A.1}$$

Where

$$B = \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_M(t_1) \\ B_1(t_2) & B_2(t_2) & \cdots & B_M(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(t_N) & B_2(t_N) & \cdots & B_M(t_N) \end{pmatrix} \tag{A.2}$$

is the basis for the B-Spline model. $\theta = \{\theta_0, \theta_1, ..., \theta_{M-5}\}^T$ are the coefficients for B-Spline model.

$y = \{y_0, y_1, ..., y_{N-1}\}^T$ is the data points.

A.1.2 Knot selection

B-Spline requires consideration of the knots:

- *The number of knots*
- *The placement of knots*

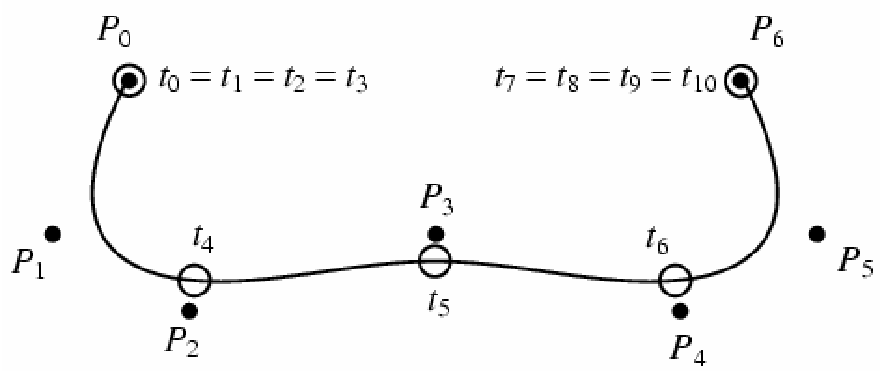

Figure A.1 Example of knots. P_0 to P_6 are the data points to fit. t_0 to t_{10} are the knots.

There are usually n = M-6 knots insides in the interior region $[t_0, t_{N-1}]$, three knots reside on the left and three on the right. Typically a choice $5 \le n \le 9$ works well, and more than 9 knots is seldom necessary, and it maybe give unexpected oscillation to the curve, and a even worse situation when too many knots are supplied is to yield a under-determined system.

For the placement of knots, the first consideration is multiplicity. Since we are dealing with cubic spline, degree of 4, we expect a C^2-continuity if there are no multiple knots. Since for a spline with degree d, we have

$$\textit{number of continuity at } k_i + \textit{number of knots at } k_i = d \tag{A.3}$$

Multiplicity 1 gives a C^2-continuity, and so on. For many applications, we use equally spaced fixed knots to give the basis B.

A.2 Build B-Spline Model

A.2.1 Model setup

This model involves deposit, FRA contract, and swap contract consecutively. Suppose we have four deposits, which are 1, 2, 3, and 6 months respectively. Starts from (the settle date) 4/19/2004 (MM/DD/YYYY) as t_{dep}^0. The FRA contract starts from the third Wednesday of March, June, September, or December. The first FRA contract starts from earliest day of those four particular dates after t_{dep}^0. Suppose there are eight FRA contracts. The first discount factor is calculated by doing an exponential interpolation from discount factor of deposits between t_{dep}^1 and t_{dep}^2. An exponential interpolation is a linear interpolation in the log space.

$$df_{fra}^1 = df_{dep}^{1\ \frac{t_{dep}^2 - t_{fra}^1}{t_{dep}^2 - t_{dep}^1}} df_{dep}^{2\ \frac{t_{fra}^1 - t_{dep}^1}{t_{dep}^2 - t_{dep}^1}} \Big/ \left(1 + r_{fra}^1 \frac{t_{fra}^2}{t_{fra}^1}\right) \tag{A.4}$$

We have swap rates from 1 year to 10 years. Since the last FRA contract ends at 6/21/2006, and there is no overlapping between FRA and swap. Hence the first swap contract starts from 4/30/2007 (since 4/29/2007 is Sunday, the actual day will consequently move to the followed Monday), which is the 3 year swap contract. The discount factor for year 1 and 2 can be found by interpolation from the discount factor curve of deposits and FRA together. The formula for discount factor is

$$df_{si} = \frac{1 - r_{si}\sum_{j=1}^{i-1} df_{sj}}{1 + r_{si}} \tag{A.5}$$

The B-Spline model for deposits and FRA is $B\theta = df$ where

And for swap is $S(D - Bt\theta) = 0$ where

$$S = diag(r_{swap}^i),\quad D_i = \sum_{j=1}^{i} df_{swap}^j$$

$$Bt = \left(\sum_{j=1}^{i} B\right) = \begin{pmatrix} B_1(t_1) & B_2(t_1) & \cdots & B_M(t_1) \\ B_1(t_2) + B_1(t_1) & B_2(t_2) + B_2(t_1) & \cdots & B_M(t_2) + B_M(t_1) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} B_1(t_i) & \sum_{i=1}^{N} B_2(t_i) & \cdots & \sum_{i=1}^{N} B_M(t_i) \end{pmatrix} \tag{A.6}$$

The final B-Spline model is

$$\underset{\theta'}{Min} \| A\theta' - f \| \tag{A.7}$$

where $A = \begin{pmatrix} B \\ Bt \end{pmatrix}$ and $f = \begin{pmatrix} df_{dep,fra} \\ D \end{pmatrix}$.

One way to have an accurate fitting is to choose the knots as the same value as t_i. This leads a least square problem $\| A\theta - f \|^2$ where A is rank deficient. Thereby the solution is not unique, and we select the L2-minimal solution, computed by the singular value decomposition [7]. A cheaper but sometimes less accurate alternative to the SVD is QR decomposition with pivoting. For more detail about the algorithms, please refer to [7]. The rank deficient least square solver by SVD and QR with pivoting are both available in Lapack.