

Financial Market Microstructure and Trading Algorithms

M.Sc. in Economics and Business Administration
Specialization in Applied Economics and Finance

Department of Finance
Copenhagen Business School 2009

Jens Vallø Christiansen

Submitted January 9, 2009

Advisor: Martin Richter

1 SUMMARY

The use of computer technology has been an important part of financial markets for decades. For some time, banks, hedge funds and other sophisticated market participants have used computer programs, known as trading algorithms, to trade directly in the market. As electronic trading has become more widespread, computer-based access to the markets has become more broadly available as well. Many banks and brokerages offer their clients, including private investors, access to the financial markets by means of advanced computer systems that route orders to the optimal price. The consequence has been increased trading volume, better liquidity, and tighter spreads. On major stock exchanges such as NASDAQ in the United States, trading algorithms now represent the majority of daily volume. This means that the majority of trading takes part without direct contact between human traders.

There are two main types of trading algorithms, those that are used for optimal execution, i.e. obtaining the best possible price for an order, and those used for speculation. This paper describes both from a theoretical perspective, and shows how two types of speculative algorithms can be designed. The first is a strategy that uses exponential moving averages to capture price momentum. The second is a market neutral relative-value strategy that trades individual stocks against each other known as pairs trading. Both are tested using empirical data and the results are encouraging. Despite the widespread use of algorithms in the markets, evidence remains of positive excess returns. In particular, the results of Gatev et al (2006) based on pairs trading are confirmed using more recent data from the London Stock Exchange. The idea of univariate pairs trading is extended to a multivariate framework in two ways. The second is based on state space methods. The results show that for the data sample used, higher transaction costs outweigh any benefits from this extension. The theoretical foundation of trading algorithms is market microstructure theory. This theory deals with the dynamics of trading and the interaction that takes place between market participants. Among the important issues are the existence of asymmetric information and the adjustment of market prices to new information, either private or public. The methodology of Hasbrouck (1991) is used to analyze the information content of high-frequency transaction data, also from the London

Stock Exchange. The results obtained show that the main conclusions in Hasbrouck's paper remain valid.

The concept of optimal execution is given a theoretical treatment based on Almgren and Chriss (2001) and McCulloch (2007). The first paper shows that the problem of minimizing implementation shortfall can be expressed as a quadratic optimization problem using a simple utility function. The second paper shows how the intra-day volume-weighted average price can be used as a benchmark for optimal execution.

2	CONTENTS	
1	Summary.....	2
3	Introduction	6
3.1	Overview	6
3.2	Literature.....	7
3.3	Notation	8
4	Background Part I: Market microstructure theory.....	9
4.1	Introduction and overview.....	9
4.2	Liquidity pools and aggregators	10
4.3	The information content of stock trades	10
4.3.1	Bins and transaction time	11
4.3.2	Model specification.....	12
4.3.3	Vector autoregression	13
4.3.4	A simple microstructure model	16
5	Background Part II: Trading algorithms.....	17
5.1	Overview	17
5.2	Optimal execution algorithms.....	19
5.2.1	Optimal execution with a quadratic utility function.....	20
5.2.2	Extensions to the optimal execution model: Drift.....	25
5.2.3	Extensions to the optimal execution model: Serial correlation	26
5.2.4	Sub-conclusion	26
5.2.5	Optimal VWAP Trading	27
5.3	Speculative algorithms	30
5.3.1	Overview	30
5.3.2	Pairs trading	32
5.3.3	Cointegration	33
5.3.4	Pair selection and trading signals	35
5.3.5	Multivariate pairs trading	36
6	Background Part III: State space models	39
6.1	Introduction.....	39
6.2	The linear Gaussian state space model.....	39
6.3	An example of a structural model.....	41

6.4	The Kalman filter	42
6.4.1	The Kalman smoother and Disturbance smoothing	44
6.4.2	Maximum likelihood estimation	45
6.4.3	Standard error of maximum likelihood estimates	47
6.4.4	The expectation maximization algorithm	47
7	Analysis Part I: High-frequency equity data	49
7.1	Data description	49
7.1.1	Description of a representative stock: Anglo American PLC	49
7.2	Dealing with transaction time	51
7.3	Empirical analysis of tick data	53
8	Analysis Part II: Speculative algorithms.....	58
8.1	Momentum strategies using moving averages.....	58
8.1.1	EMA-momentum strategy with different sampling frequencies.....	63
8.2	Univariate pairs trading.....	66
8.3	Multivariate pairs trading.....	77
9	Conclusion	84
10	Sources.....	86
11	Appendix.....	89
11.1	Event types in the FTSE 100 tick data	89
11.2	Volume and open interest of Bund and Euribor futures	90
11.3	Additional co-integration tests.....	91
11.4	Derivation of Kalman filter recursions and illustration.....	92
11.5	Empirical analysis of tick data	95
11.6	Derivation of the exponential moving average.....	96
11.7	EMA momentum strategies	97
11.8	Univariate pairs trading.....	98
11.9	Multivariate pairs trading.....	103

3 INTRODUCTION

3.1 Overview

Over the last few decades algorithmic trading has become an important part of modern financial markets. As the use of computer technology has become more broad based, investors are demanding faster, cheaper, more reliable, and more intelligent access to financial markets. Banks and hedge funds are taking advantage of this trend and have begun an arms race towards creating the best electronic trading systems and algorithms. Algorithmic trading represents an ever-growing share of trading volume, and in some markets the majority (KIM, K., 2007).

The theoretical foundation for algorithmic trading is found primarily in the fields of financial econometrics and market microstructure. The study of the time series properties of security prices is among the most pervasive subjects in the financial literature, and has grown rapidly in tandem with cheaper access to computing power. The field is characterized by the vast amount of data available to researchers in the form of databases of historical transaction data. The statistical theory needed to analyze such data is different from the datasets known from conventional economics with less frequent observations, and is often much more computationally intensive. Traditional methods in exploratory data analysis such as vector autoregression have been joined by new methods such as autoregressive conditional duration models, to take into account the nonsynchronous nature of high-frequency transaction data (ENGLE, R. and Russell, J., 1998).

The primary ambition of this paper is to provide a brief introduction to an extensive subject. The content has been selected with the aim of covering core areas of the theory while maintaining a coherent whole. The papers and models that will be covered are particularly well suited to empirical testing as opposed to much theory in the microstructure literature. The secondary aim of the paper is to determine whether it is possible to create speculative trading algorithms that earn positive excess returns.

The analysis will focus on two main areas, the microstructure of financial markets and trading algorithms. The two are connected in the sense that microstructure theory provides the theoretical basis for the development of trading algorithms. The distinction between microstructure theory and financial econometrics is often blurry, and elements from each field will be used as deemed

appropriate. The structure of the paper is thus divided into two main parts, a part that presents theoretical background, and an empirical analysis using market data.

In the first part, sections 4-6 present the background for the empirical analysis. Section 4 presents a method for the analysis of high-frequency equity transaction data. Section 5 covers trading algorithms, both for optimal execution and speculation. Section 6 outlines the general theory of state space models which can be used to design speculative algorithms.

In the second part, sections 7 & 8 apply the theory to market data. Section 7 carries out an empirical analysis of equity tick data based on the methodology of section 4. Section 8 proceeds to test two kinds of speculative algorithms, one based on security price momentum, the other based on the relative value of securities. Section 9 concludes.

3.2 Literature

The academic literature on the subject of market microstructure is vast. In this paper the main sources used were Hasbrouck (1991) and Hasbrouck (2007). For optimal execution the main sources were Almgren & Chriss (2001) and McCulloch & Kazakov (2007). Academic work on speculative trading algorithms is scarce, and pairs trading in particular, but Gatev et al (2006) gives a useful overview. The empirical analysis was done using MATLAB¹, and to this end Kassam (2008) was a great help.

Shumway & Stoffer (2006) was the source for general theory of time series analysis, and Campbell et al (2006) for financial econometrics, including a chapter on market microstructure. Durbin & Koopman (2001) was the main reference for state space models.

¹ The author may be contacted at jenschristiansen@gmail.com for the MATLAB code used in the empirical analysis.

3.3 Notation

The following notation has been used throughout the paper unless otherwise specified:

- t denotes a particular point in time while τ denotes a time increment $\Delta t = t_k - t_{k-1}$ for $k = 1, 2, \dots$
- $\sim N(\mu, \sigma^2)$ denotes a normally distributed random variable with mean μ and variance σ^2 .
- $\sim iidN(\mu, \sigma^2)$ denotes a random variable which is distributed with independent and identical increments from a normal distribution with mean μ and variance σ^2 .
- I_n denotes the $n \times n$ identity matrix.

4 BACKGROUND PART I: MARKET MICROSTRUCTURE THEORY

4.1 Introduction and overview

This section will present select parts of the theory of market microstructure, and prepare the reader for the empirical analysis in section 7. What follows is a brief description of the dynamics of a modern securities market.

The main market mechanism in modern electronic markets is the limit order book. The limit order book consists of a list of buy and sell orders at different prices and for different quantities. An example could be an order to ‘buy 100 shares at \$30.10 or less’ or ‘sell 300 shares at \$30.50 or more in 100 share increments’. By consolidating all such orders in a central system, as for example a stock exchange, it is possible to know the best bid (buy order) and best offer (sell order) at any given time. The difference between the bid and the ask is known as the spread. Trades take place when a trader is willing to ‘cross’ the spread, that is, to buy at the offer or sell at the bid of someone else. The market is typically anonymous, and trades may be made based only on the price and quantity being bought or sold. This is known as a continuous auction.

An *active* limit order that hasn’t yet been executed is known as a quote. Once an order is executed, i.e. a transaction takes place at a quoted price (either at the bid or ask), the quote disappears and is replaced by the next best available bid or ask. Note that a quote may be withdrawn before it is ever executed. The historical observations of interest to market participants are thus both historical *quotes* and historical *trades*. These are also the quantities that will be used for empirical analysis.

The effectiveness of continuous auction market depends on the amount of active market participants, and the amount of a given security they are willing to trade at any given time. This evasive concept is known as market ‘liquidity’. One of the main challenges of market microstructure theory is to define and quantify it. In a well-functioning market there are many participants that trade significant amounts of a security with each other – continuously. The price of the security being quoted at any given point in time will thus reflect what many participants believe it should be – otherwise the orders would be filled, and the market would move up or down in the order book. This process is known as price discovery. See Hasbrouck (2007) for further details.

The following section will touch briefly upon some of the main institutional features of modern electronic markets.

4.2 Liquidity pools and aggregators

Modern markets are characterized more by their fragmentation than their consolidation. This is despite of the progress of technology and electronic trading in particular. The main reason is that market participants constantly seek cheaper and better venues for their trading. A good example is the recent success of multilateral trading facilities (MTFs) such as Chi-X and Turquoise in Europe. MTFs are hybrid trading venues in the sense that they connect market participants to decentralized 'liquidity pools'. The term liquidity pool is used to describe the existence of a market in a security outside a central exchange. Liquidity pools are typically operated by large banks, and are intended to lower the cost of trading by moving it off the central exchanges (TURQUOISE, 2008).

The obvious implication of this is market fragmentation, and this is the gap that MTFs bridge by connecting the various liquidity pools at low cost. Market participants communicate across the MTFs, centralized exchanges and private 'dark pools' using common communication protocols. The most popular is called the Financial Information eXchange (FIX) (WIKIPEDIA, 2008a).

Market participants naturally want to access as many different liquidity pools as possible, to obtain the best possible price available. To this end they use 'aggregators', computer systems designed to route orders to the best possible price, wherever this price may be quoted. Aggregators are typically provided by investment banks, for example Deutsche Bank's autobahn system (DB, 2008). The importance of price aggregators has been further increased by the need to provide clients with 'best execution' in accordance with securities regulation such as MiFID in Europe (Markets in Financial Instruments Directive) (WIKIPEDIA, 2008b).

4.3 The information content of stock trades

This section will provide the theoretical background for the subsequent empirical analysis of tick data from the London Stock Exchange. The theoretical framework is that of Hasbrouck (1991) and involves the use of vector autoregression to extract the information content of stock trades.

4.3.1 Bins and transaction time

A common way to analyze stock data is to observe the daily closing prices of a given stock for an arbitrary number of trading days. The frequency of price observation is merely a matter of scaling, however, and this paper will analyze stock data from a more detailed perspective, namely at the tick level. The highest possible frequency of observation for a given stock is the observation of every single *trade event*. This may be combined with the observation of every single *quote event* to give a detailed view of the intra-day trading process. Large company stocks trade very frequently² however, so trade events are typically grouped in fixed time intervals known as bins. Daily price observations for a given stock may be seen as the creation of one-day bins of transaction data. Each bin has an opening price and a closing price. The price may in this case be the bid, the ask, the quote midpoint (the average of the bid and the ask) or the last traded price. Typically the last traded price is used for daily observations. Possible bin sizes range from a year or more to a minute or less. Bins are particularly useful for graphing price data. Figure 4-1 shows a 'candle' graph of Anglo American PLC (AAL.LN) with hourly bins. The green candles indicate that AAL.LN closed at a higher price than the opening price of a given one-hour period. The red candles indicate the opposite and the black 'whiskers' at the top or bottom of a candle indicate the range in which the stock traded during the time interval.

² As an example, on March 3, 2008 between 8:00 and 16:30, Anglo American PLC (AAL.LN) listed on London Stock Exchange experienced 51,412 events of which 9,941 were trade events and 41,471 were quote events.



Figure 4-1: AAL.LN hourly bins March 3-12, 2008. Source: E*TRADE.

When every trade and quote event of a stock is observed, each event is a new point in what is known as transaction time. Clock time increments in transaction time can vary from event to event. On March 3, 2008, the average transaction time increment of AAL.LN was 0.5945 seconds with a standard deviation of 1.8413, a minimum of 0 and a maximum of 50.92.

4.3.2 Model specification

Following the analysis in Hasbrouck (1991), the primary price variable of interest is the quote midpoint. At time t the best bid and ask quote in the market are denoted by q_t^b and q_t^a respectively, and transactions are characterized by their signed volume x_t (purchases are positive, sales are negative). Define the value of the security at some convenient terminal time T in the distant future as ψ_T , and let ϕ_t be the public information set at time t . Then the symmetry assumption is:

$$E[(q_t^a + q_t^b)/2 - \psi_T | \phi_t] = (q_t^a + q_t^b)/2 - E[\psi_T | \phi_t] = 0 \quad (4.1)$$

I.e. the quote midpoint at time t , $q_t = (q_t^a + q_t^b)/2$, contains all available public information of the future value of the security, ψ_T . The information inferred from the time t trade (x_t) can then be summarized as the subsequent change in the quote midpoint:

$$r_t = (q_t^a + q_t^b)/2 - (q_{t-1}^a + q_{t-1}^b)/2 = q_t - q_{t-1} \quad (4.2)$$

Conveniently, due to the symmetry assumption in (4.1), the information impact of x_t is not affected by the transaction cost-based component of the spread. This specification is characterized by a trade impact that is fully contemporaneous and can be written as $r_t = bx_t + v_{1,t}$. In reality, quote revisions are likely to show a lagged response to trade innovations for various reasons related to the market microstructure. Hasbrouck mentions threshold effects due to price discreteness, inventory control effects, and lagged adjustment to information (HASBROUCK, J., 1991). Threshold effects are of psychological nature, market participants respond to some price levels different than other because of their numerical value (round numbers or historical highs and lows). For example, a stock that approaches the GBp 1,000 mark for the first time is likely to motivate a different kind of behavior from market participants than the behavior seen when the stock was range-bound between GBp 920 and GBp 960.³ Inventory control effects are probably less prominent in today's largely electronic markets compared to the market on the NYSE in 1989 that Hasbrouck was investigating. Lagged adjustment to information is likely to remain an issue, but has possibly diminished since 1989 due to the presence of more market participants, the emergence of computerized trading algorithms, and more efficient trading systems in general (HASBROUCK, J., 2007).

4.3.3 Vector autoregression

A more flexible structure that allows the current quote revision to be affected by past quote revisions and trades, can be made using vector autoregression. A vector autoregressive model of order p , VAR(p), is written as

$$v_t = \alpha + \sum_{j=1}^p \gamma_j v_{t-j} + w_t \quad (4.3)$$

where each γ_j is a $k \times k$ transition matrix that expresses the dependence of v_t on v_{t-j} . The vector white noise process w_t is assumed to be multivariate normal with mean-zero and covariance matrix

³ GBp is an abbreviation for one penny, which is 1/100 of a British pound sterling, the price unit used for stocks in Great Britain.

$E(w_t w_t') = \Sigma_w$. As an example, a bivariate VAR(1) model, i.e. $k = 2$, consists of the following two equations:

$$v_{1t} = \alpha_{10} + \gamma_{11}v_{1,t-1} + \gamma_{12}v_{2,t-1} + w_{1t} \quad (4.4)$$

$$v_{2t} = \alpha_{20} + \gamma_{21}v_{1,t-1} + \gamma_{22}v_{2,t-1} + w_{2t} \quad (4.5)$$

(SHUMWAY, R. H. and Stoffer, D. S., 2006). The relationship between quote revisions and trades may be modeled using the structure in (4.4). We get

$$r_t = a_1 r_{t-1} + a_2 r_{t-2} + \dots + b_1 x_{t-1} + b_2 x_{t-2} + \dots + v_{1,t} \quad (4.6)$$

where $v_{1,t}$ is a disturbance term. The quote revision at time t is expressed as a function of past quote revisions and past trades. This implies that there is serial correlation in the quote revisions. Since the symmetry assumption in (4.1) is incompatible with such serial correlation in the quote revisions, it is replaced by a weaker assumption:

$$\text{As } s \rightarrow T, E[(q_s^a + q_s^b)/2 - \psi_T | \phi_t] \rightarrow 0 \quad (4.7)$$

For some future time s , $t < s < T$, conditional on the information set at time t . This allows any deviation in the quote midpoint from the efficient price to be transient.

To allow for causality running from quotes to trades, trades may be modeled in a similar fashion:

$$x_t = c_1 r_{t-1} + c_2 r_{t-2} + \dots + d_1 x_{t-1} + d_2 x_{t-2} + \dots + v_{2,t} \quad (4.8)$$

The innovation, $v_{2,t}$, captures the unanticipated component of the trade relative to an expectation formed from linear projection on the trade and quote revision history. Jointly equations (4.6) and (4.8) comprise a bivariate vector autoregressive system. It is assumed that the error terms have zero mean and are jointly and serially uncorrelated:

$$\begin{aligned} E v_{1,t} &= E v_{2,t} = 0, \\ E v_{1,t} v_{1,s} &= E v_{2,t} v_{2,s} = E v_{1,t} v_{2,s} = 0, \text{ for } s \neq t. \end{aligned} \quad (4.9)$$

The expected cumulative quote revisions through step m in response to the trade innovation $v_{2,0}$ may be written as

$$\alpha_m(v_{2,0}) = \sum_{t=0}^m E[r_t | v_{2,0}]. \quad (4.10)$$

By (3.5) as m increases,

$$\begin{aligned}\alpha_m(v_{2,0}) &= \sum_{t=0}^m E \left[\frac{(q_t^a + q_t^b)}{2} - \frac{(q_{t-1}^a + q_{t-1}^b)}{2} \middle| v_{2,0} \right] \\ &\rightarrow E[\psi_T | v_{2,0}] - E\psi_T = 0\end{aligned}\tag{4.11}$$

That is, the expected cumulative quote revision converges to the revision in the efficient price. For this reason $\alpha_m(v_{2,0})$ can be interpreted as the information revealed by the trade innovation, and constitutes the underlying construct of Hasbrouck's framework.

A way of seeing why it is important to include lagged trades and quote revisions in the model is to consider an alternative to the vector autoregressive setup. This could be a simpler model that assumes the complete absence of any transient effects in the price discovery process, including liquidity effects. This can be written as $r_t = \hat{\alpha}x_{t-1} + \hat{v}_{1,t-1}$ where the " \wedge " symbol denotes that the model is incorrectly specified. The regression coefficient $\hat{\alpha} = Cov(r_t, x_{t-1})/Var(x_{t-1})$ is likely to overestimate the immediate effect of a trade on the quote revision due to inventory and liquidity considerations. Instead of capturing the lagged adjustment of the efficient price to the trade innovation, this oversimplified model will embed all short-term effects in the regression coefficient $\hat{\alpha}$. Hasbrouck (1991) shows that another alternative specification of the model, which does not include lagged versions of the dependent variable, will also be inferior to the full specification.

An important feature of the VAR model of equations (4.6) and (4.8) is the implication that all public information is captured by the innovation $v_{1,t}$ and all private information (plus an uncorrelated liquidity component) with the trade innovation $v_{2,t}$. The rationale for the first implication is that all public information is immediately reflected in the quotes posted by market makers (otherwise the market makers would be exposed to arbitrage). The second implication is due to the fact that trade innovations reflect information that was not already contained in the history of trades and quote revisions, and must therefore be based on external (or private) information or liquidity trading. In other words, public information is not useful in predicting the trade information. Letting ϕ_t be the public information immediately subsequent to the time t quote revision,

$$E[v_{2,t+k} | \phi_{t-1}] = 0, \quad \text{for } k \geq 0.\tag{4.12}$$

4.3.4 A simple microstructure model

The VAR setup as it has been presented so far is an econometric representation of a simple microstructure model. The following description is also adapted from (HASBROUCK, J., 1991). The model exhibits both asymmetric information and inventory control behavior. Let m_t be the efficient stock price, the expected value of the stock conditional on all public information. The dynamics of m_t are given by

$$m_t = m_{t-1} + z v_{2,t} + v_{1,t} \quad (4.13)$$

where $v_{1,t}$ and $v_{2,t}$ are mutually and serially uncorrelated disturbance terms, and interpreted in the same way as above. The z coefficient reflects the private information conveyed by the trade innovation $v_{2,t}$. The quote-midpoint price q_t has dynamics

$$q_t = m_t + a(q_{t-1} - m_{t-1}) + b x_t \quad (4.14)$$

where x_t is again the signed trade at time t and a and b are adjustment coefficients with $0 < a \leq 1$ and $b > 0$. Equation (5.14) has an inventory control interpretation. Say that at time $t = 0$, $q_0 = m_0$. If $x_1 > 0$, i.e. an agent purchases from the market maker at the existing quote q_0^a , the market maker will react by raising his bid q_1^b to elicit sales. If the spread $(q_t^a - q_t^b)$ remains constant, this implies that q_1 rises. The case of $a < 1$ is associated with imperfect inventory control: competition from public limit order traders, for example, forces q_t to move closer to m_t with the passage of time.

The final equation in the model describes the evolution of trades:

$$x_t = -c(q_{t-1} - m_{t-1}) + v_{2,t} \quad (4.15)$$

where $c > 0$ defines a downward sloping demand schedule, i.e. when the quote midpoint rises above the perceived efficient price (by a magnitude greater than half the spread), market participants react by selling. From an econometric viewpoint the efficient price m_t is unobservable, so an empirical model must involve only x_t and q_t . This forms the basis for the VAR framework. Section 7.3 presents an empirical analysis of historical transaction data from London Stock Exchange.

5 BACKGROUND PART II: TRADING ALGORITHMS

5.1 Overview

Algorithmic trading volume has increased dramatically in the past several years. The NYSE reports that in 2000, 22% of all trading was executed via trading algorithms, up from 11.6% in 1995. In 2004, that number had increased to 50.6% (KIM, K., 2007). There are several reasons for the emergence and relative success of trading algorithms. One reason is that much trading in the financial markets is done based on discretionary human decision-making without consistent adherence to specific rules or systems. Broadly speaking a trading rule is a set of instructions a trader follows that depend on the market price of one or more financial instruments. A trading rule takes market prices as input and gives orders to buy or sell at a given point in time as output. Most traders do use rules and systems that they believe have worked in the past, but they are hard to repeat consistently, and traders will be tempted to deviate from their rules once it appears that they may not be working anymore. Such small deviations in trading patterns may do a critical amount of damage to an otherwise successful trading strategy. Trades executed by trading algorithms, on the other hand, are based on rules that may be formulated mathematically or in computer code, and are therefore possible to repeat with perfect consistency.

We may distinguish between two main types of trading algorithms: optimal execution algorithms and speculative algorithms. Optimal execution algorithms seek to execute orders in the market at the lowest possible cost, whereas speculative algorithms take market risks in the hope of earning a profit. As depicted in Figure 5-1, the process of developing and then implementing trading algorithms starts with financial modeling in a suitable statistical software environment such as Excel or MATLAB. Based on historical data the algorithm is designed and back-tested using historical price data to the point at which it has attractive out-of-sample characteristics. In the case of speculative algorithms this could be a high risk-adjusted return or low correlation with the return from investing in the broader market. Developers of optimal execution algorithms will instead focus on achieving fast and cost-efficient execution. The implementation then proceeds by bridging the development environment with trading infrastructure that may execute orders generated by the algorithm. The algorithm now takes as input real-time data from the financial markets.

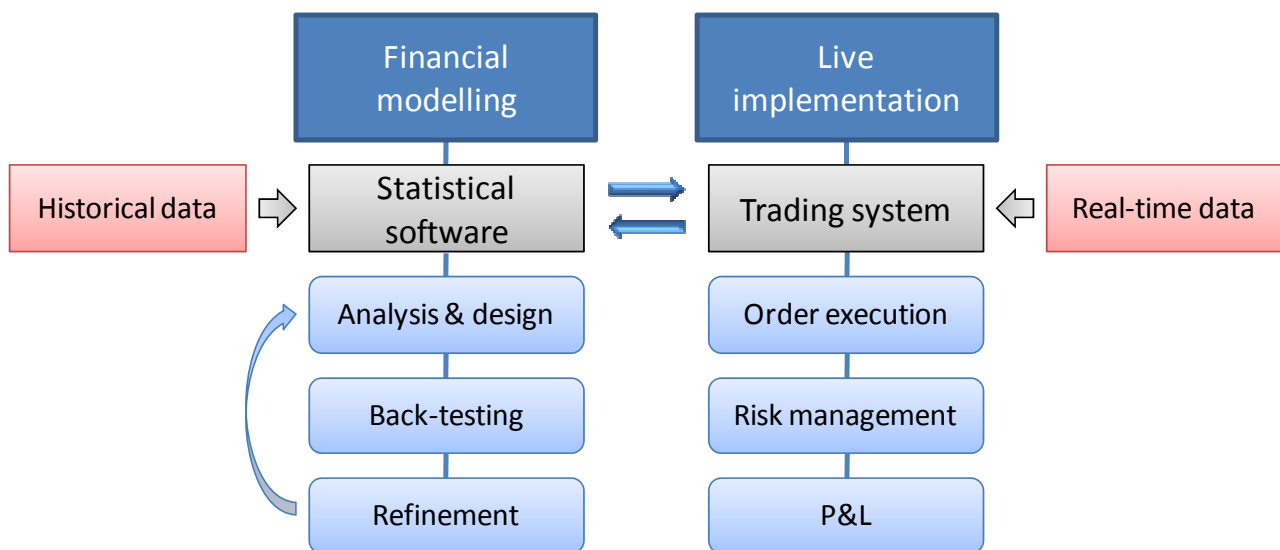


Figure 5-1: Trading algorithm development and implementation.

The implementation of both optimal execution and speculative algorithms requires that the algorithm can be programmed in a computer language, so the trading process can be fully automated. If the process of getting orders to the market at any stage requires human intervention, the reaction speed and accuracy of the algorithm will fall, and thus many of the appealing characteristics of trading algorithms will be impaired. On the other hand, the flexibility of the algorithm and its ability to adapt to a changing market environment will only be as good as the computer code it is based on – an obvious disadvantage compared to human traders. The speed with which a trading algorithm reacts to incoming real-time market data, processes the data and reacts by issuing new market orders or by waiting, will depend upon the computer infrastructure and programming language used. High-level programming languages such as MATLAB are suitable for the financial modeling of trading algorithms, but ‘faster’ languages are typically used to build trading systems. C++, C# and Erlang are examples of low-level programming languages. The latter is a concurrent language that facilitates simultaneous execution of several interacting computational tasks, and is particularly fast (WIKIPEDIA, 2008c). Such characteristics may improve the performance of the trading algorithm and decrease the time to market – which is a critical factor for both optimal execution and speculative algorithms. The faster

the algorithm can react to incoming real-time market data, the more likely it is to obtain the liquidity it seeks, i.e. successfully execute scheduled trades before other market participants.⁴

5.2 Optimal execution algorithms

The aim of optimal execution algorithms is to minimize the transaction costs involved in executing large orders, which is also known as execution costs. The literature on optimal execution usually focuses on the equity markets, as they are the most transparent and most thoroughly researched markets (many of the results found can be directly applied in other markets). The benchmark of execution costs is typically the *arrival price*, which is the average of the bid and ask price in the market when execution of an order begins (either a buy or sell order). The difference between the arrival price and the average price obtained for the order is known as the *implementation shortfall*. If the aim is to liquidate a given position, the implementation shortfall is the difference between the market value of the position at the beginning of liquidation, and the amount of cash obtained at the end of liquidation. The reason that implementation shortfall is different from zero, is the limitations imposed by market liquidity (or depth) and bid-ask spreads. At any given point in time there are buyers and sellers available in the market for a specific number of shares which may be much less than the size of the order to be executed. Once execution begins, the process of trading the order is likely to move the market price *against* the execution trader. When the aim is to buy a number of shares, the ask price will go up, and when the aim is to sell the bid will fall. This is known as market impact, and can be either temporary or permanent. An example of temporary market impact is when the bid-ask spread widens in response to a large trade. Typically the bid-ask will revert to its previous more narrow level once market participants have reacted to the change in price. Permanent impact is a change in the efficient price that will not immediately readjust to its previous level. As described in section 4.3.2, the efficient price can be changed by the act of trading alone if such an act is assumed to convey private information to the market, or if it is a lagged response to information already made public.

⁴ Time to market is a significant issue for hedge funds using algorithmic trading strategies. Some of those hedge funds are known to have placed their servers close to the New York Stock Exchange and other strategic venues to decrease the latency of data transfer (TEITELBAUM, R., 2007).

5.2.1 Optimal execution with a quadratic utility function

The implementation shortfall problem may be solved by creating an objective function for the execution trader that takes into account risk-aversion. Almgren & Chriss (2001) define such an objective function, and show that it may be minimized with respect to a quadratic utility function or a value-at-risk (VaR) measure. They show that there are two extremes in the approach to executing a buy or sell order. One is to execute the entire order immediately, and the other is to execute it at evenly spaced intervals throughout the trading *horizon*. The trading horizon places an upper limit on the amount of time the order execution may take. In between the two extremes there exists an efficient frontier in the space of time-dependent liquidation strategies. That is, for a given level of positive risk-aversion, there is a single trading strategy that dominates all other possible strategies. The derivation of the results that follow may seem daunting, but rests on a simple quadratic minimization problem.⁵ Using *linear* trade impact functions facilitates the derivation of explicit solutions to the minimization problem.

Following Almgren & Chriss (2001), suppose that we hold a block X of securities that we wish to liquidate before time T . We divide T into time intervals $\tau = T/N$, and define the discrete times $t = 0, \dots, N$ given by $t = k\tau$, for $k = 0, \dots, N$. We define a trading trajectory to be a list x_0, \dots, x_N , where x_t is the number of units that we plan to hold at time t . Our initial holding is $x_0 = X$, and liquidation at time T requires $x_N = 0$. We may equivalently define a strategy by the “trade list” n_1, \dots, n_N , where $n_t = x_{t-1} - x_t$ is the number of units that we sell between times $t - 1$ and t . Clearly, x_t and n_t are related by

$$x_t = X - \sum_{j=1}^t n_j = \sum_{j=t+1}^N n_j, t = 0, \dots, N.$$

A trading strategy can then be defined as a rule for determining n_t in terms of information available at time $t - 1$. An important point is that the optimal strategy is the same at all times t for $t = 0, \dots, N$ if prices are serially uncorrelated (see Almgren & Chriss (2001) for proof).

Now suppose that the initial value of our security is S_0 , so the initial market value of our position is XS_0 . The security’s price evolves according to two exogenous factors: volatility and drift, and one

⁵ The approach is similar to the derivation of the Capital Asset Pricing Model, although the CAPM is a maximization problem (CAMPBELL, J. Y. et al., 1997).

endogenous factor: market impact. These characteristics of price movement may be summarized as a discrete arithmetic random walk

$$S_t = S_{t-1} + \sigma\sqrt{\tau}\xi_t - \tau g(n_t/\tau), \text{ for } t = 1, \dots, N. \quad (5.1)$$

Here σ is the volatility of the asset, measured in standard deviations per year, ξ_t are iid normal random variates, and the permanent impact $g(v)$ is a function of the average rate of trading $v = n_t/\tau$. In equation (5.1) there is no drift term which we interpret as the assumption that we have no information about the direction of future price movements. Note that typically a continuous *geometric* random walk of the kind

$$dS_t = \mu S_t dt + \sigma S_t dz \quad (5.2)$$

is used to model stock prices, where dz denotes a Wiener process (HULL, J. C., 2006). (5.2) may be approximated in discrete time by

$$\begin{aligned} \Delta S_t &= \mu S_t \Delta t + \sigma S_t \Delta z \\ \Delta z &= \xi_t \sqrt{\Delta t} \quad \xi_t \sim iid N(0,1) \quad \Delta t = \tau \end{aligned} \quad (5.3)$$

Equation (5.2) models changes in the stock price S_t whereas (5.1) models the level of the stock price. For the purpose of modeling stock prices *intra-day*, (5.1) is a useful approximation of (5.3) and leads to tractable results.

Returning to the model in equation (5.1), we define temporary market impact as a change in S_t caused by trading at the average rate v , but we do not include this directly in the process of the efficient price. Rather it is added separately to the objective function. It can be expressed as $\tilde{S}_t = S_{t-1} - h(n_t/\tau)$. We now define the *capture* of a trajectory to be the full trading revenue upon completion of all trades,

$$\sum_{t=1}^N n_t \tilde{S}_t = XS_0 + \sum_{t=1}^N (\sigma\sqrt{\tau}\xi_t - \tau g(n_t/\tau)) x_t - \sum_{t=1}^N n_t h(n_t/\tau) \quad (5.4)$$

The first term on the right-hand side is the initial market value of our position. The second term is the effect of price volatility minus the change in price as a consequence of the permanent impact of trading, and the third term is the fall in value caused by the temporary impact of trading. The total cost of trading can be expressed as $XS_0 - \sum_{t=1}^N n_t \tilde{S}_t$ and is the implementation shortfall.

Prior to trading the implementation shortfall is a random variable with expectation $E(x)$ and variance $V(x)$. We readily obtain

$$E(x) = \sum_{t=1}^N \tau g(n_t/\tau) x_t + \sum_{t=1}^N n_t h(n_t/\tau) \quad (5.5)$$

$$V(x) = \sigma^2 \sum_{t=1}^N \tau x_t^2 \quad (5.6)$$

The objective function may then be expressed as

$$U(x) = E(x) + \lambda V(x) \quad (5.7)$$

where λ is a Lagrange multiplier that may be interpreted as a risk-aversion parameter. The objective of the analysis is to minimize the objective function in (5.7) for a given risk-aversion parameter λ , thus minimizing the expected shortfall while taking into account the uncertainty of execution. Using a linear impact function $g(v) = \gamma v$ the permanent impact term becomes

$$\sum_{t=1}^N \tau g(n_t/\tau) x_t = \frac{1}{2} \gamma X^2 - \frac{1}{2} \gamma \sum_{t=1}^N n_t^2$$

and the temporary impact term becomes

$$h(n_t/\tau) = \epsilon \text{sgn}(n_t) + \frac{\eta}{\tau} n_t$$

Where 'sgn' is the sign function. With linear impact equation (5.5) becomes

$$E(x) = \frac{1}{2} \gamma X^2 + \epsilon \sum_{t=1}^N |n_t| + \frac{\tilde{\eta}}{\tau} \sum_{t=1}^N n_t^2 \quad (5.8)$$

in which $\tilde{\eta} = \eta - \frac{1}{2} \gamma \tau$. Almgren & Chriss (2001) show that we may construct efficient strategies by solving the constrained optimization problem $\min_{x: V(x) \leq V_*} E(x)$ for a given maximum level of variance V_* . This corresponds to solving the unconstrained optimization problem

$$\min_x (E(x) + \lambda V(x)) \quad (5.9)$$

where, as already mentioned, the risk-aversion parameter λ is a Lagrange multiplier. The global minimum of (5.9), can be found by differentiating (5.7) with respect to each x_j yielding

$$\frac{\partial U}{\partial x_j} = 2\tau \left\{ \lambda \sigma^2 x_j - \tilde{\eta} \frac{x_{j-1} - 2x_j + x_{j+1}}{\tau^2} \right\} \quad (5.10)$$

where $\partial U / \partial x_j$ is the partial derivative of (3.15) for $j = 1, \dots, N - 1$. Setting (3.18) equal to zero and simplifying we obtain

$$\begin{aligned} \frac{1}{\tau^2} (x_{j-1} - 2x_j + x_{j+1}) &= \tilde{\kappa}^2 x_j \\ \tilde{\kappa}^2 &= \frac{\lambda \sigma^2}{\tilde{\eta}} = \frac{\lambda \sigma^2}{\eta \left(1 - \frac{\gamma \tau}{2\eta}\right)} \end{aligned} \quad (5.11)$$

The optimal trajectory x_j is now expressed as a linear difference equation and may be solved using the hyperbolic sine and cosine functions giving the expressions

$$x_j = \frac{\sinh(\kappa(T - t_j))}{\sinh(\kappa T)} X \quad j = 0, \dots, N \quad (5.12)$$

$$n_j = \frac{2\sinh\left(\frac{1}{2}\kappa\tau\right)}{\sinh(\kappa T)} \cosh\left(\kappa\left(T - t_{j-\frac{1}{2}}\right)\right) X \quad j = 1, \dots, N \quad (5.13)$$

where n_j is the associated trade list. See Almgren & Chriss (2001) for details on the derivation. The exposition here shows that it is possible to find a closed-form expression for the optimal trading trajectory using linear impact functions. The result is depicted in Figure 5-2 below as the efficient frontier of time-dependent liquidation strategies. Figure 5-3 shows the optimal trading trajectories for the three points A, B and C in Figure 5-2.

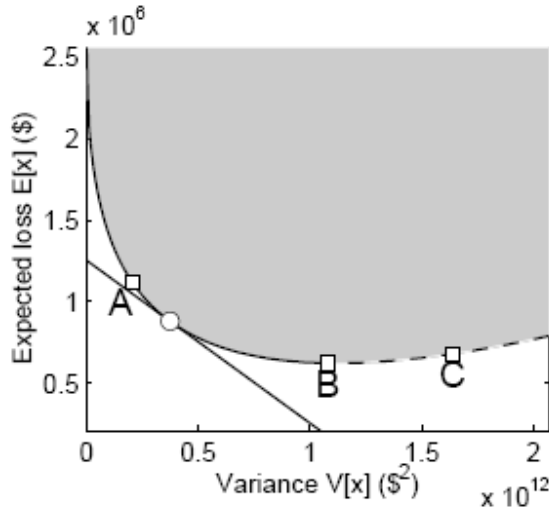


Figure 5-2: The efficient frontier of time-dependent liquidation strategies.
(ALMGREN, R. and Chriss, N., 2001)

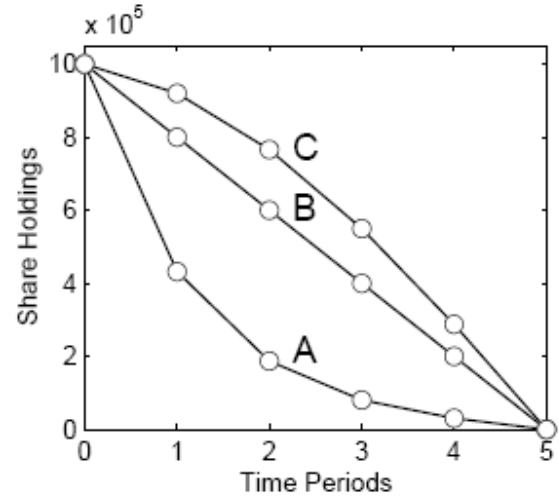


Figure 5-3: Three different trade trajectories for different values of λ . (A) $\lambda > 0$, (B) $\lambda = 0$, (C) $\lambda < 0$.
(ALMGREN, R. and Chriss, N., 2001)

The point 'B' is the naïve minimum variance strategy that corresponds to a risk-aversion parameter value of $\lambda = 0$. The strategy disregards the role of the variance of x and trades at a constant rate throughout the trading period. The point 'A' is an example of an optimal strategy for a trader with a positive risk aversion coefficient. For a relatively small (first-order) increase in expected loss $E(x)$, a relatively large (second-order) reduction in loss variance $V(x)$ is obtained. Point 'C' illustrates the optimal strategy of a trader who likes risk, and therefore has a negative risk-aversion coefficient. It is clear that all risk-averse traders will have convex trading trajectories.

Almgren & Chriss (2001) proceeds to show that equivalent results can be obtained by minimizing the liquidity adjusted value-at-risk (L-VaR). This is the maximum amount an execution trader is willing to lose, with a given statistical confidence over the trading period. The L-VaR objective function may be written as

$$\text{Var}_p(x) = E(x) + \lambda_v \sqrt{V(x)} \quad (5.14)$$

where the confidence interval is determined by the number of standard deviations λ_v from the mean by the inverse cumulative normal distribution function (we call λ from (5.7) λ_u to distinguish between the two). p is the probability with which the strategy will not use more than $\text{Var}_p(x)$ of its market value in trading. In other words, the implementation shortfall will not exceed $\text{Var}_p(x)$ a fraction p of

the time. Because $\text{Var}_p(x)$ is a complicated nonlinear function of the x_j , we cannot obtain an explicit minimizing solution such as (5.12)-(5.13). But once the efficient frontier has been calculated using (5.12)-(5.13) it is easy to find the value of λ_u corresponding to a given value of λ_v .

Extensions to the model in (5.8) can be made by expanding the information set of the trader. This can be done by including a drift term in price process (5.1) or by assuming that the error term ξ_t in (5.1) is serially correlated.

5.2.2 Extensions to the optimal execution model: Drift

A drift term α may be added to (5.1) to give

$$S_t = S_{t-1} + \sigma\sqrt{\tau}\xi_t + \alpha\tau - \tau g(n_t/\tau), \text{ for } t = 1, \dots, N. \quad (5.15)$$

The optimality condition (5.11) becomes

$$\frac{1}{\tau^2}(x_{t-1} - 2x_t + x_{t+1}) = \tilde{\kappa}^2(x_t - \bar{x}) \quad (5.16)$$

in which the new parameter $\bar{x} = \alpha/(2\lambda\sigma^2)$ is the optimal level of security holding for a time-dependent optimization problem. The optimal trading trajectory and corresponding trade list become

$$x_j = \frac{\sinh(\kappa(T - t_j))}{\sinh(\kappa T)} X + \left[1 - \frac{\sinh(\kappa(T - t_j)) + \sinh(\kappa t_j)}{\sinh(\kappa T)} \right] \bar{x} \quad (5.17)$$

$$\begin{aligned} n_j &= \frac{2\sinh(\frac{1}{2}\kappa\tau)}{\sinh(\kappa T)} \cosh\left(\kappa\left(T - t_{j-\frac{1}{2}}\right)\right) X \\ &+ \frac{2\sinh(\frac{1}{2}\kappa\tau)}{\sinh(\kappa T)} \left[\cosh\left(\kappa t_{j-\frac{1}{2}}\right) - \cosh\left(\kappa\left(T - t_{j-\frac{1}{2}}\right)\right) \right] \bar{x} \end{aligned} \quad (5.18)$$

$j = 0, \dots, N$. (5.17) is the sum of two distinct trajectories: the zero-drift solution in (5.12) plus a “correction” which profits by capturing a piece of the predictable drift component by holding a static position, \bar{x} , in the stock (ALMGREN, R. and Chriss, N., 2001). The difference between the solution in (5.12) and the one in (5.17) can be seen in a highly liquid market when $\kappa T \gg 1$. For a risk-averse trader, the optimal trajectory in such market conditions approaches strategy ‘B’ in Figure 5-3, as the importance of the risk-aversion parameter diminishes. In the case of (5.17), however, the optimal

trajectory approaches the optimal static portfolio holding \bar{x} . Near the end of the trading period this final holding is also sold to satisfy $x_N = 0$ at $t = T$.

Define x_j^* as the optimal solution using (5.12) and x_j^0 when using (5.17). The gain from the drift-enhanced strategy can then be expressed as

$$\alpha\tau \sum_{k=1}^N (x_j^* - x_j^0) = \alpha\bar{x}T \left(1 - \frac{\tau \tanh\left(\frac{1}{2}\kappa T\right)}{T \tanh\left(\frac{1}{2}\kappa\tau\right)} \right) \quad (5.19)$$

Since $\tanh(x)/x$ is a positive decreasing function, this quantity is positive and bounded above by $\alpha\bar{x}T$. Almgren & Chriss (2001) show that for any realistic values of the parameters, this quantity is negligible compared to the impact costs incurred in liquidating an institutional-size portfolio over a short period of time.

5.2.3 Extensions to the optimal execution model: Serial correlation

When ξ_t , $t = 0, \dots, N$, are serially correlated the optimal strategy becomes dynamic, that is, the best possible strategy at time $t > 0$, is no longer the same as the optimal strategy at time $t = 0$. If we denote the period-to-period correlation of ξ_t by ρ , we may express the maximum per-period gain as $(\rho^2 \sigma^2 \tau^2)/4\eta$. As in the case of the drift-enhanced trajectory, the gains are negligible when using realistic values for the parameters. Only in the case of an extremely liquid stock with extremely high serial correlation will the gains be significant for institutional trading (ALMGREN, R. and Chriss, N., 2001). But in reality those two characteristics are mutually exclusive.

5.2.4 Sub-conclusion

The optimal execution model of Almgren and Chriss (2001) gives a clear understanding of the tradeoff between execution uncertainty and market impact. It shows that a utility maximizing risk-averse trader who has one day to execute an order, will divide the order over the entire day to minimize market impact and make optimal use of ‘liquidity pockets’ during the day. Given the assumption of no serial correlation in prices, the optimal strategy is static and therefore unchanged throughout the trading period.

The central feature of the model is the creation of an efficient frontier of time-dependent execution strategies. The frontier is depicted in a two-dimensional plane whose axes are the expectation of total cost and its variance. Each point on the frontier corresponds to the optimal strategy of a trader with a given level of risk-aversion.

It is interesting to note that taking into account possible serial correlation or drift in prices, does not improve the performance of the strategy to a significant extent.

5.2.5 Optimal VWAP Trading

Another approach to the optimal execution problem is to use the volume-weighted average price as an execution benchmark. The volume-weighted average price is defined as

$$VWAP_N = \sum_i p_i V_i / \sum_i V_i$$

where p_i is the traded price of trade i , and V_i is the traded volume of trade i for $i = 1, 2, \dots, N$ being the trades in the VWAP period. An execution trader who has to execute a large buy order during the period of one trading day can split up the order into smaller bits and seek to obtain a final VWAP which is close to the market VWAP. This way he will know that the combined order was executed at reasonable prices given the volatility and liquidity conditions in the market. This is a more useful benchmark than the simple average price.

Define $v(t)$ as the strategy intra-day cumulative volume and $v(T)$ as total final volume. Market cumulative and total volume are denoted by $V(t)$ and $V(T)$, respectively. The strategy's intra-day relative volume can then be written as $x_t = v(t)/v(T)$, and market intra-day relative volume as $X_t = V(t)/V(T)$. Here the analysis is limited to buy orders, and unlike in section 5.2 above, the x_t variable is now normalized between 0 and 1, where 0 means that nothing has been traded, and 1 means that the entire order has been traded and the operation is done. An optimal VWAP strategy is a strategy that minimizes the expected difference between market VWAP and traded VWAP. This can be expressed as $\mathcal{V}(x_t) = VWAP(M) - VWAP(x_t)$ where x_t is the controlled trading strategy and M is the market.

Konishi (2002) derives a static optimal execution strategy that minimizes the L^2 norm of $\mathcal{V}(x_t)$

$$\min_{x_t} E\{[VWAP(M) - VWAP(x_t)]^2\} \quad (5.20)$$

In this framework, prices follow standard Brownian motion without drift of the form

$$dP(t) = \sigma(t, V(t))dB(t) \quad (5.21)$$

Under this assumption, for a single-stock trade, if price volatility is independent of market trading volume, the optimal execution strategy is determined only by the expected market trading volume distribution and is independent of expectations regarding the magnitude and time dependency of price volatility (KONISHI, H., 2002). The details of the analysis will not be pursued here. Instead a generalization of the approach that models intra-day volume as a Cox process will be presented.

Intra-day volume as a Cox process

McCulloch (2007) shows that if intra-day volume is modeled as a Cox (doubly stochastic) point process then intra-day *relative* volume may be modeled as a doubly stochastic binomial point process. Based on this idea, as well as the results of Konishi (2002), McCulloch and Kazakov (2007) derive an optimal VWAP trading strategy that takes price drift into account. Prices are assumed to evolve as a semi-martingale of the form $P_t = A_t + M_t + P_0$, where A_t is price drift, M_t is a martingale and P_0 is the initial price. The minimum VWAP risk trading problem is generalized into the optimal VWAP trading problem using a mean-variance framework as in section 5.2.1. The resulting optimal strategy is given by

$$x_t^* = \max_{x_t} [E[\mathcal{V}(x_t)] - \lambda \text{Var}[\mathcal{V}(x_t)]] \quad (5.22)$$

where λ is a Lagrange multiplier, and is interpreted as the VWAP traders risk-aversion coefficient. McCulloch and Kazakov (2007) show that for all feasible VWAP trading strategies x_t , there is always residual VWAP risk. The residual risk can be written as $\text{Var}[\mathcal{V}(x_t)]$ and is proportional to the price variance σ^2 of the stock and the variance of the relative volume process $\text{Var}(X_t)$. Empirical testing shows that relative volume variance is proportional to the inverse of stock final trade count K raised to the power of 0.44.

$$\min_{x_t} \text{Var}[\mathcal{V}(x_t)] \propto \sigma^2 \int_0^T \text{Var}[x_t] dt \propto \frac{\sigma^2}{K^{0.44}}$$

The relative volume X_t is the ratio of a random sum specified by the doubly stochastic binomial point process as the ‘ground process’ over the non-random sum of all trade volumes. It is assumed that

final trade volume is known in the information filtration of X_t , which is clearly an unrealistic assumption. In practice this value must be forecasted. This assumption is not made in Konishi (2002), and is a disadvantage of the approach in McCulloch and Kazakov (2007).

Implementation of the optimal VWAP strategy is done by dividing each day into bins. Bins are designed by dividing the VWAP trading period $[0, T]$ into b time periods with the bin boundary times for bin i denoted as $t = k_{i-1}$ and $t = k_i$. So $0 = k_0 < k_1 < \dots < k_i < k_{i+1} < \dots < k_b = T$. Each bin (time interval) must be large enough to allow the trading system to reach a specific proportional amount x_t^* of the total size of the order at the end of each bin.

One way of dividing the bins is to use equal-volume bins so $x^*(k_i) - x^*(k_{i-1}) = 1/b \quad \forall t \in [k_i, k_{i+1}]$.

Additional VWAP risk from using discrete volume bins depends on the number of bins b as $\mathcal{O}(b^{-2})$.

Figure 5-4 depicts the optimal trading trajectory using equal-volume bins or optimal bins with the continuous solution superimposed. Using optimal bins is a reasonable approximation to the continuous solution.

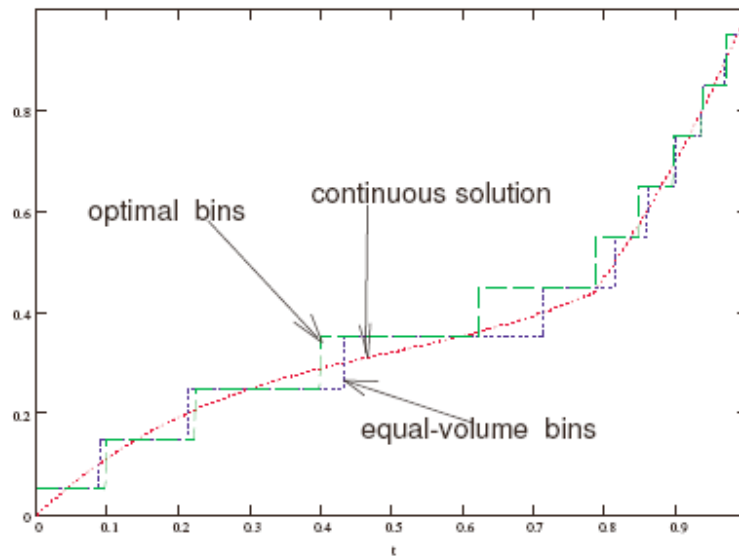


Figure 5-4: The continuous solution is compared to 10 optimal and 10 equal-volume bins. The x-axis is time and the y-axis is relative intra-day volume on the interval $\{0,1\}$. (MCCULLOCH, J. and Kazakov, V., 2007).

5.3 Speculative algorithms

5.3.1 Overview

The aim of speculative algorithms is to profit from changes in prices of traded assets by trading according to specific rules. These rules are typically based on market inputs, such as a live feed of market prices. The algorithm processes this live input and creates buy and sell signals accordingly. In order for the algorithm to be successful, the trade signals must be timely and reliable. The frequency of trading is arbitrary and depends on the nature of the algorithm. It may be a high frequency equity strategy that trades several times per minute, or a managed futures strategy that take a strategic position once per month. Speculative algorithms may be based on many different strategies. Three categories of typical strategies are momentum, relative-value and microstructure strategies.

Momentum strategies attempt to identify and follow trends in market prices by using statistical measures such as a moving average cross-over. They perform well in a market environment that is characterized by strong trends that are persistent over time. If market prices are moving 'side-ways' or show wildly oscillating behavior, momentum strategies will not perform well.

Relative-value strategies compare the price of one or more securities to the price of one or more other securities and trade them against each other when the price difference (or ratio or other relative measure) diverges from historical norms. The cheap or under-valued securities are bought (long position), and the dear or over-valued securities are sold (short position). The strategy is based on the idea that if the relative pricing of the securities has diverged from the historical norm, they will converge again in the future. The obvious risk is that the circumstances or factors that dictated the pricing of the securities in the past will no longer do so in the future. Therefore, it is possible that the relative pricing of the securities will never return to the historical level on which the trading strategy is based, and may continue to diverge. This is known as a regime change, and is the primary risk of relative-value trading.

Microstructure strategies attempt to exploit the mechanics of electronic markets. The architecture of some markets allows information to be extracted and acted upon in a way that is difficult to achieve without the help of an algorithm. A good example of this is the electronic limit order book. The 'depth' of the limit order book varies over time and between stocks and exchanges, but it is typically

reported as the five or ten best bids and offers at a given point in time, along with the quantities of shares to be bought or sold. This information gives an idea about the supply and demand in the market at different prices and how the balance changes over time.

In a fast and efficient market, such as the market for shares of large companies, the limit order book is updated very frequently, and it is a challenge for most traders to process this information, let alone observe it with accuracy. A computer algorithm may in this case be useful as it can rapidly process the information in the limit order book and execute trades based on it. A popular order type in electronic markets is the 'iceberg'. Iceberg orders reveal only a fraction of total volume at a time, replenishing as trades are executed. They are used to minimize trade impact by hiding the intentions of the trader from other market participants. Anecdotal evidence⁶ suggests that iceberg orders have become increasingly common in the market for futures contracts on interest rates. As a consequence, the limit order book of for example Bund⁷ or Euribor⁸ contracts reveals less volume at the bid and offer than what is actually readily available from market participants. In 2006-2007, Bund and Euribor contracts with maturity 1-3 months into the future typically had a total of 500-2000 contracts on the bid and ask. By December 2008 the volume had fallen to a few hundred contracts, largely because of iceberg orders. Open interest and traded volume in the contracts has fallen as well, but not nearly at the same rate, see Figure 11-1 in Appendix 11.2 for an illustration. This makes the analysis of historical limit order books more difficult (TRAGSTRUP, L., 2008).

An example of a microstructure trading strategy is to make use of limit order books that include stop-loss orders (TEITELBAUM, R., 2007). A stop-loss order is an order that traders submit to sell below the current best bid or buy above the current best offer to close a losing long or short position, respectively. An algorithm searches for a particularly large stop-loss sell order close to a psychologically significant price level such as \$15.00, when the stock is trading at, for instance, \$15.05. The algorithm then submits a substantial sell order at, say, \$15.01 hoping that the stop-loss order at \$15.00 will be hit. If the order at \$15.00 is executed, it will place significant downward

⁶ Author's interview with Senior Execution Trader Lars Tragstrup, Danske Markets (a division of Danske Bank) on December 30, 2008.

⁷ Futures contract for the delivery of EUR 100,000 notional principal of German government bonds with a maturity of 7-10 years at a future date.

⁸ Futures contract for a 3 month deposit with a notional value of EUR 1,000,000.

pressure on the stock. The algorithm then submits a take-profit order at a pre-determined level, for example at \$14.90. If \$15.01 is reached, but \$15.00 is never reached, the algorithm may submit a stop-loss order of its own, at for example \$15.05. The same algorithm may be used to trade many different stocks as it only requires that stop-loss orders are made publicly available, which may be the case for all stocks traded on a given exchange.

5.3.2 Pairs trading

Pairs trading will be used as an example of a speculative relative-value trading algorithm. The basic idea behind pairs trading is to trade two stocks that move together over time in a systematic way. If they drift apart to a specific pre-determined extent, the cheaper stock is purchased and the more expensive stock is shorted (sold). Proceeds from shorting the second stock should finance most of the initial purchase. Then the trader waits for the two prices to converge towards their historical price difference. If and when that happens the position is closed at a profit. The risk is that the two stocks continue to diverge further, and never return to their historical price difference. In that case the trader loses money.

The reason that the two stocks should move together is that they may share common factors in the sense of equilibrium asset pricing such as arbitrage pricing theory (CAMPBELL, J. Y. et al., 1997), or they are cointegrated in the sense of Engle and Granger (1987). Asset pricing can be viewed in absolute and relative terms. The pricing of a stock in absolute terms is done by discounting future cash flows at a discount rate that reflects the company's risk. It is notoriously difficult to find an accurate price and there is a wide margin of error due to the uncertainty involved in forecasting the cash flows and determining an appropriate discount factor. Relative pricing is based on the Law of One Price, which Ingersoll (1987) defines as the "proposition ... that two investments with the same payoff in every state of nature must have the same current value." (GATEV, E. et al., 2006). Two investments with similar future payoffs should therefore trade for a similar value. This may be true for two similar stocks. The similarity of value will vary over time, but the relationship will be mean-reverting to the extent that the fundamentals driving the two stocks don't change. This is a critical point in the evaluation of the risk of pairs trading, as the greatest vulnerability of the strategy is structural breaks. Structural breaks are points in time at which a significant change in the

fundamental valuation of a given security takes place. This is typically due to a company specific event such as a lawsuit, a new product invention, or a surprising earnings announcement. If the basis for the historical relationship between two stocks changes significantly, there is no reason why their relative value in the future should resemble the past.

5.3.3 Cointegration

A theoretical basis for pairs trading is found in cointegration. When a linear combination of two time series of stocks prices is stationary, while the two individual time series are non-stationary (which is usually the case), the two stocks are said to be cointegrated. Suppose that combinations of time series of stock prices obey the equation:

$$p_{it} = \sum (\beta_{il} p_{lt}) + \varepsilon_{it}, \text{ for } k < n \quad (5.23)$$

Where p_{it} is the price of stock i at time t , β_{il} is the regression coefficient of stock i on stock $l \neq i$, and ε_{it} is a covariance stationary error term in the sense of Shumway and Stoffer (2006). Assuming that p_{it} are covariance stationary after differencing once, the price vector p_t is integrated of order 1 with cointegrating rank $r = n - k$ (ENGLE, R. F. and Granger, C., 1987). Thus, there exist r linearly independent vectors $\{\alpha_q\}_{q=1, \dots, r}$ such that $z_q = \alpha_q' p_t$ are weakly dependent. In other words, r linear combinations of prices will not be driven by the k common non-stationary components p_l . The non-stationary components are in this case k stocks in the population of n stocks that are redundant in the process of creating cointegrating vectors α_q . The cointegrating rank of the individual price vectors is not used explicitly in the creation of pairs trading strategies in this paper. Yet the concept serves to show that in a given population of stocks, there may be several possible linear combinations of stocks with a cointegrating relationship.

Note that this interpretation does not imply that the market is inefficient, rather it says that certain assets are weakly redundant, so that any deviation of their price from a linear combination of the prices of other assets is expected to be temporary and reverting (GATEV, E. et al., 2006).

The idea that a linear combination of two stocks may be covariance stationary, may be interpreted as saying that a cointegrating vector may be partitioned in two parts, such that the two corresponding portfolios are priced within a covariance stationary error of each other. Given a large enough

population of stocks, this statement is empirically valid and provides the basis for identifying pairs of stocks suitable for pairs trading (GATEV, E. et al., 2006).

A research note on pairs trading from Kaupthing Bank recommends the use of cointegration tests for pair selection (BOSTRÖM, D., 2007). The argument is that the more significant the cointegration test is, the more likely it is that the pairs trade will work. Two tests for cointegration are mentioned, the Engle-Granger test and the Durbin-Watson test⁹. The Engle-Granger test is based on the Dickey-Fuller unit-root test for stationarity. It starts by regressing one time series, y_t , on another, x_t :

$$y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t \quad (5.24)$$

The residuals \hat{u}_t from this regression are then used to perform the regression:

$$\Delta\hat{u}_t = \hat{\beta}_1\hat{u}_{t-1} + \hat{\varepsilon}_t \quad (5.25)$$

where Δ denotes the difference operator and $\hat{\varepsilon}_t$ is a white noise error term. The t-statistic of the $\hat{\beta}_1$ parameter in (5.25) is called the tau-statistic¹⁰, the critical values of which can be found in Gujarati (2003) and in most statistical software packages. If the absolute value of the t-statistic obtained is larger than its critical tau value, the residuals are integrated of order 0, $I(0)$, that is, they are stationary, and the two series y_t and x_t are cointegrated.

Engle and Granger (1987) compare various measures of stationarity as a means of testing for cointegration and conclude that the Dickey-Fuller test is the most powerful in a statistical sense. They note that if the data is autocorrelated the *augmented* Dickey-Fuller (aDF) test should be used. The aDF test assumes that the observed time series y_t is driven by a unit-root zero drift process, i.e. an ARIMA(P,1,0) model with P autoregressive terms. It is based on the regression

$$y_t = \phi y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \dots + \zeta_p \Delta y_{t-p} + \varepsilon_t \quad (5.26)$$

for some AR(1) coefficient $\phi < 1$, and a number of lags p . If the observed time series y_t is a random walk with drift, a constant term can be included in the above regression. The stock price data used in section 8.2 is normalized, however, so the constant term is not different from zero in a statistical sense. Therefore the model in equation (5.26) is correctly specified.

⁹ See Appendix 11.3 for a definition of the Durbin-Watson test.

¹⁰ Note that there is no connection to Kendall's tau distribution.

Boström (2007) uses cointegration tests as one amongst a number of statistical tests to evaluate the quality of a pairs trade. I.e. the back-testing procedure described in the paper requires each pair to have Engle-Granger and Durbin-Watson test statistics above certain pre-specified levels. A similar method will be used in section 8.2, specifically the MATLAB function 'dfARTest', which performs an augmented Dickey-Fuller test assuming zero drift in the underlying process.

5.3.4 Pair selection and trading signals

There are several possible approaches to choosing pairs. One way is the method of testing for cointegration described above. Another is the use of a minimum distance criterion between the normalized prices of the stocks in a given population. This method is used by both Gatev et al 1997 and Perlin (2007). The first step of the method is to normalize the price series of each stock in the population of stocks (the population can be chosen arbitrarily as the members of a stock index or all stocks traded on a given exchange, for example). In this way, stocks with different price levels may be compared in a consistent way. This may be written as $P_{it}^* = [P_{it} - E(P_i)]/\sigma_i$ where P_{it} is the price of stock i at time t , $E(P_i)$ and σ_i are the mean and standard deviation of the price series of stock i , respectively, and P_{it}^* is the normalized price. We denote the normalized price of the *pair* of P_{it}^* by Q_{it}^* , and the difference between the two as $e_{it}^* = P_{it}^* - Q_{it}^*$. To find a suitable pair for a stock in a population of n other stocks, we find the stock that minimizes the sum of the squared differences (L^2 norm) for $t = 1, 2, \dots, T$: $\min_i \sum_{t=1}^T (e_{it}^*)^2$. In a given population of stocks, there may be several stocks that have a similarly low minimum distance value, but for the purpose of this paper, only the stock with the lowest value is chosen.

Although this method of finding pairs is strictly defined, it is possible to guess towards the characteristics of possible pairs. Depending on the chosen population of stocks, two candidates for a pairs trade are probably in the same industry, trade on the same stock exchange, and share other features such as scale of operation, geography, and market value.

Once pairs have been identified, the distance e_{it}^* between each target stock and its pair is evaluated on a daily basis. For some pre-specified constant b , when $|e_{it}^*| > b$ a pairs trade is opened.

Depending on whether e_{it}^* is positive or negative, the target stock is sold or bought, respectively, and the opposite position is taken in its pair. The unit of the difference in normalized prices can be vaguely

interpreted as the number of standard deviations between the two (because both prices have been normalized). Therefore a logical level for b would be 2, as this would imply a two standard deviation difference to the historical normalized spread, which can be considered a low-probability event (at approximately the 5% level if the normalized spread has a normal distribution). Another possibility is to define b as $b = d * std(e_{it}^*)$ for some constant d , where $std(.)$ denotes standard deviation. In this way, the barrier level b will be unique for each pair, and will reflect the specific volatility of e_{it}^* . Again, a logical value for d is 2.

The size of the two positions may be determined in various ways. One option is to use linear regression to determine the weight of the pair stock, β_i , as if it were a hedge:

$$P_{it}^* = \alpha_i + \beta_i Q_{it}^* + \varepsilon_{it} \quad (5.27)$$

Another option is for both to have the same initial market value. The size of the two positions will change over time as the market prices of the two stocks change. They may either be rebalanced or left alone, the risk is that one position will become much larger or smaller than the other, so the overall market exposure becomes either positive or negative. This is particularly clear when a large portfolio of pairs is traded simultaneously.

The regression method of weighing positions suggests an alternative approach to determining the price ‘distance’ e_{it}^* . The residual ε_{it} from equation (3.30) is the deviation of P_{it}^* from Q_{it}^* given the estimated parameters α_i and β_i . It may be used as a trading signal by defining a barrier level $b = d * std(\varepsilon_{it}^*)$ in the same way as described above. The method can be used on regular prices as well, that is P_{it} and Q_{it} , as the constant term α_i takes into account the difference in levels.

Section 8.2 will present empirical results based on the method of normalized prices and pair positions with equal market value.

5.3.5 Multivariate pairs trading

The idea of pairs trading can be extended to trading a portfolio of one more stocks against another portfolio of stocks (PERLIN, M. S., 2007b). Using the notation from above for normalized prices this can be expressed as $P_{it}^* = f(X)$ where $f(X)$ is some function of a matrix with information that explains P_{it}^* . This information could be any economic variable, but if it isn’t the price of a security, it cannot be traded. If this is the case it is still possible to create a strategy that trades P_{it}^* ‘outright’

against the ‘signal’ in $f(X)$, but the strategy will not be market neutral. Since this paper focuses on market neutral relative-value strategies, only tradable securities will enter the X matrix. The normalized price of target stock P_{it}^* may be compared to the normalized prices of a portfolio of n other stocks, yielding the expression

$$P_{it}^* = \alpha + w_1 Q_{1t}^* + \dots + w_n Q_{nt}^* + \epsilon_t \quad P^* \neq Q^* \quad (5.28)$$

where α is a constant, ϵ_t is an error term, and w_j , $j = 1, \dots, n$ are the weights given to each stock. To simplify,

$$P_{it}^* = \alpha + \sum_{j=1}^n w_j Q_{jt}^* + \epsilon_t. \quad (5.29)$$

The stocks in the Q_{jt}^* portfolio, which we shall call the M-pair portfolio, may be found in various ways. One approach is to compare P_{it}^* with each candidate for Q_{jt}^* individually, by means of a minimum distance criterion or a cointegration test. Alternatively OLS may be used on various combinations of stocks in Q_{jt}^* , to find which combination best explains P_{it}^* (highest R^2).

But since both P_{it}^* and Q_{jt}^* are non-stationary (the process of normalization doesn’t affect unit root non-stationarity), the problem of spurious regression arises, i.e. the ϵ_t might be non-stationary. One way to get around the problem is to use discrete or log returns of P_{it} and Q_{jt} (i.e. not the normalized series P_{it}^* and Q_{jt}^*),

$$\Delta \log(P_{it}) = \mu + \sum_{j=1}^n w_j \Delta \log(Q_{jt}) + \eta_t. \quad (5.30)$$

where η_t is a white noise error term.

But if there is a cointegrating relationship between P_{it}^* and its M-pair portfolio Q_{jt}^* , ϵ_t in (5.29) may be stationary. In that case, it is possible to use the w_j coefficients for statistical inference.

The M-pair portfolio may be created in several ways. One obvious approach is to use the n stocks with least (L^2 norm) distance to P_{it}^* *individually*, as was done for a single stock in section 5.3.4.

Another approach is to use an iterative procedure that maximizes the degree of cointegration between P_{it}^* and Q_{jt}^* . The first stock in Q_{jt}^* will be chosen using the minimum distance criterion. Each subsequent stock will be chosen to minimize the augmented Dickey-Fuller test statistic of the error

term ϵ_t in (5.29). The benefit of the method is that it ensures that P_{it}^* and Q_{jt}^* are cointegrated¹¹. The downside is that it is difficult to say a priori whether this will actually improve the algorithm. It is not certain that the additional stocks added to Q_{jt}^* contain useful information.

In order to trade the M-pair portfolio, we must scale the w_j in (5.29) so that they sum to one. This is done by dividing each weight by the sum of the weights to obtain $w_j^* = w_j / (\sum_{j=1}^n w_j)$. The w_j^* sum to one, so scaling the positions by these parameters ensures that the market exposure of the M-pair portfolio will be equivalent to the target stock P_{it}^* .¹²

Perlin (2006b) also suggests using a correlation weighting scheme that calculates the weights as $w_j = \rho_j / (\sum_{j=1}^n \rho_j)$ for $j = 1, \dots, n$. It is difficult to see what the advantage of this method should be, except that the weights will reflect the relationship between each Q_{jt}^* and P_{it}^* without the effect of the other Q_{jt}^* .

An alternative approach to multivariate pairs trading is to use state space methods to extract a signal from one or more stocks against which to trade P_{it} . State space models allow the estimation of unobserved processes based on observations. This unobserved process, which is called a signal, can be based on one or more stocks including the reference stock. Section 6 will present the concept of state space models, and section 8.3 will use such a model for multivariate pairs trading.

¹¹ To the extent that the augmented Dickey-Fuller test works.

¹² The strategy will only be market neutral initially. Price fluctuations from day to day will change the magnitudes of the various positions, bringing the strategy out of 'balance'. Unless the positions are rebalanced on a daily basis, net market exposure will be different from zero.

6 BACKGROUND PART III: STATE SPACE MODELS

6.1 Introduction

This section will provide general background on state space models, and illustrate the properties that are used in section 8.3 on multivariate pairs trading. State space models were originally invented as a tool to track the position and trajectory of space craft. Given a set of inputs from various sensors and tracking devices, such as velocity and azimuth, it was necessary to estimate the unobserved quantities of position and trajectory in a computationally efficient way. Furthermore, these estimates had to be continuously updated as new observations came in from the sensors and tracking devices. This led to the Kalman filter, a method of recursive updates of a system of equations which is the underlying construct of state space models. The benefits of state space models are their inherent flexibility and scope of application, and their computational efficiency is a major benefit in terms of numerical estimation of parameters. Examples of applications include structural models of trend and seasonality, exponential and spline smoothing, as well as stochastic volatility. State space models can be considered an alternative to the ARIMA system of analysis of Box and Jenkins (see for example Box et al 1994). ARIMA models require that the time series used as input are covariance stationary, so it is typically necessary to detrend data by taking one or more differences. This is not required in the state space approach where data characteristics such as trend and seasonality may be modeled explicitly. This is a fundamental difference between the two methods. Furthermore it is interesting to note that ARIMA analysis may be expressed and estimated in state space form.

6.2 The linear Gaussian state space model

This section will describe the linear Gaussian state space model. The description is adapted from Durbin & Koopman (2001), but will also include elements from Shumway & Stoffer (2006) and Welch & Bishop (2006). The general linear Gaussian state space model can be written in the form

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t & \varepsilon_t &\sim N(0, H_t) & t &= 1, \dots, n \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t & \eta_t &\sim N(0, Q_t) \end{aligned} \quad (6.1)$$

where y_t is a $p \times 1$ vector of observations and α_t is an unobserved $m \times 1$ vector called the state vector. The idea underlying the model is that the development of the system over time is determined by α_t according to the second equation of (6.1), but because α_t cannot be observed directly, an estimate is made of α_t based on the observations y_t . The first equation of (6.1) is called the observation equation, and the second is called the state equation. Z_t is a $p \times m$ matrix called the observation matrix, and T_t is the state evolution matrix with dimensions $m \times m$. In most applications including the ones in this paper, R_t is the identity matrix. The matrices Z_t , T_t , R_t , H_t and Q_t are either assumed known or estimated, depending on how the model is constructed. Typically, some or all of the elements of these matrices will depend on elements of an unknown parameter vector ψ , which can be estimated with an optimization algorithm. The error terms ε_t and η_t are assumed to be serially independent, and independent of each other over time. The initial state vector α_1 is assumed to be $N(a_1, P_1)$ independently of $\varepsilon_1, \dots, \varepsilon_n$ and η_1, \dots, η_n , where a_1 and p_1 may be assumed known or estimated. Note that the first equation of (6.1) has the structure of a linear regression model where the coefficient vector α_t changes over time. The second equation represents a first order vector autoregressive model, “the Markovian nature of which account for many of the elegant properties of the state space model.”

This general specification is a powerful and flexible tool that makes the analysis of a wide range of problems possible. The main point is that a vector of one or more underlying *signals* α_t can be estimated using a vector of observations y_t . It is possible to include additional known inputs in both the observation and state equation, but this is not used in the subsequent analysis and will therefore not be described here. The multivariate case is a straight-forward extension where the disturbances are written as

$$\varepsilon_t \sim N(0, \Sigma_\varepsilon) \quad \eta_t \sim N(0, \Sigma_\eta)$$

where Σ_ε and Σ_η are $p \times p$ and $m \times m$ covariance matrices. The disturbances may be independent (diagonal covariance matrices) or correlated instantaneously across series.

6.3 An example of a structural model

Shumway & Stoffer (2006) model the quarterly earnings of Johnson & Johnson using a simple structural model: $y_t = T_t + S_t + u_t$ where T_t is the trend component, S_t is the seasonal and u_t is a disturbance term. The trend is allowed to increase exponentially, that is $T_t = \phi T_{t-1} + w_{t1}$, where $\phi > 1$. The seasonal component is modeled as $S_t + S_{t-1} + S_{t-2} + S_{t-3} = w_{t2}$, which corresponds to assuming that the seasonal component is expected to sum to zero over a period of four quarters. To express this in state space form we define $\alpha'_t = (T_t, S_t, S_{t-1}, S_{t-2})$ as the state vector so the observation equation becomes

$$y_t = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} + u_t$$

and the state equation

$$\begin{pmatrix} T_t \\ S_t \\ S_{t-1} \\ S_{t-2} \end{pmatrix} = \begin{pmatrix} \phi & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} T_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ w_{t2} \\ 0 \\ 0 \end{pmatrix}$$

where $\Sigma_\varepsilon = \varepsilon_{11}$ and

$$\Sigma_\eta = \begin{pmatrix} \eta_{11} & 0 & 0 & 0 \\ 0 & \eta_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The parameters to be estimated are ε_{11} , the noise variance in the observation equation, η_{11} and η_{22} the model variances corresponding to the trend and seasonal components, and ϕ , the transition parameter that models the growth rate. An initial guess has to be given for the parameters. The initial mean of η_t is specified as $\mu'_0 = (.5, .3, .2, .1)$ with diagonal covariance matrix $\Sigma_{0ii} = .01$ for $i = 1, \dots, 4$. Initial state covariance is specified as $\eta_{11} = .01$ and $\eta_{22} = .1$, corresponding to relatively low

uncertainty in the trend compared with the seasonal. The measurement error covariance is started at $\varepsilon_{11} = .04$. Growth is about 3% per year so ϕ is started at $\phi = 1.03$. Using the expectation maximization algorithm (see section 6.4.4) the transition parameter stabilized at $\phi = 1.035$, which is exponential growth with an annual inflation rate of approximately 3.5% (see Shumway & Stoffer (2006) for further details). Note that the initial guess values for the parameters in the model are chosen rather arbitrarily, and that they may have a substantial impact on where the estimation algorithm converges. Because of this, an element of trial and error in the estimation of state space models is inevitable.

6.4 The Kalman filter

The Kalman filter is a recursion that allows the calculation of future state estimates based on the current estimate and observation, and an initial guess of the state mean and covariance. Based on the given parameters of the state space model in question, the Kalman filter derives optimal estimates of the unobserved signal(s) that are modeled as an autoregressive process in the state equation. This section will show what the Kalman filter recursions look like. The presentation closely follows Durbin & Koopman (2001) and uses elements of Shumway & Stoffer (2006) as well as Welch & Bishop (2006).

Denote the set of observations y_1, \dots, y_t by Y_t , then the Kalman filter allows the calculation of $a_{t+1} = E(\alpha_{t+1} | Y_t)$ and $P_{t+1} = \text{Var}(\alpha_{t+1} | Y_t)$ given a_t and P_t . Define v_t as the one-step forecast error of y_t given Y_{t-1} and $F_t = \text{Var}(v_t)$.

The recursion equations are then given by¹³

$$\begin{aligned} v_t &= y_t - Z_t a_t & F_t &= Z_t P_t Z_t' + H_t \\ K_t &= T_t P_t Z_t' F_t^{-1} & L_t &= T_t - K_t Z_t' \\ a_{t+1} &= T_t a_t + K_t v_t & P_{t+1} &= T_t P_t T_t' + R_t Q_t R' \\ && \text{for } t &= 1, \dots, n \end{aligned} \tag{6.2}$$

Note here that a_{t+1} has been obtained as a linear function of the previous value a_t and v_t . K_t is known as the Kalman gain. The key advantage of the recursions is that we do not have to invert a (pt x

¹³ See Appendix 11.4 for a derivation.

pt) matrix to fit the model each time the t th observation comes in for $t = 1, \dots, n$; we only have to invert the $(p \times p)$ matrix F_t and p is usually much smaller than n .

The update of the state estimate that takes place in $a_{t+1} = T_t a_t + K_t v_t$ can be seen as two discrete steps; first a projection of the current state into the future ($T_t a_t$), and then a correction that takes into account the new (or incoming) observation ($K_t v_t$).

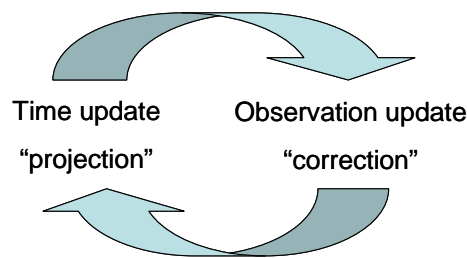


Figure 6-1: The Kalman filter recursion loop.

In a similar fashion, the error covariance P_t is first projected into the future, and then corrected.

Using $L_t = T_t - K_t Z_t$, the error covariance recursion $P_{t+1} = T_t P_t T_t' + R_t Q_t R'$ can be written as

$P_{t+1} = T_t P_t T_t' + R_t Q_t R' - T_t P_t (K_t Z_t)'$. This can also be seen as two discrete steps; first a projection of the current state error covariance into the future ($T_t P_t T_t' + R_t Q_t R'$), and then a correction that takes into account the new observation ($-T_t P_t (K_t Z_t)'$). The process is summarized in Figure 11-2 in Appendix 11.4.

Dimensions of state space model (3.12)				Dimensions of Kalman filter			
Vector		Matrix		Vector		Matrix	
y_t	$p \times 1$	Z_t	$p \times m$	v_t	$p \times 1$	F_t	$p \times p$
α_t	$m \times 1$	T_t	$m \times m$			K_t	$m \times p$
ε_t	$p \times 1$	H_t	$p \times p$			L_t	$m \times m$
η_t	$r \times 1$	R_t	$m \times r$			M_t	$m \times p$
		Q_t	$r \times r$	a_t	$m \times 1$	P_t	$m \times m$
a_1	$m \times 1$	P_1	$m \times m$	$a_{t t}$	$m \times 1$	$P_{t t}$	$m \times m$

Table 6-1: The matrix dimensions of the state space model and the Kalman filter.
(DURBIN, J. and Koopman, S., 2001)

To compute the contemporaneous state vector estimate $E(\alpha_t | Y_t)$ and its associated error variance matrix, which we denote by $a_{t|t}$ and $P_{t|t}$ respectively, the *contemporaneous filtering equations* can be used. They are a reformulation of the equations in (6.2).

$$\begin{aligned}
 v_t &= y_t - Z_t a_t & F_t &= Z_t P_t Z_t' + H_t \\
 & & M_t &= P_t Z_t' \\
 a_{t|t} &= a_t + M_t F_t^{-1} v_t & P_{t|t} &= P_t - M_t F_t^{-1} M_t' \\
 a_{t+1} &= T_t a_{t|t} & P_{t+1} &= T_t P_{t|t} T_t' + R_t Q_t R_t'
 \end{aligned} \tag{6.3}$$

for $t = 1, \dots, n$.

6.4.1 The Kalman smoother and Disturbance smoothing

As shown above, the Kalman filter is a *forward* recursion that derives an estimate for α_t given all observations up to time t . It is also possible to estimate α_t given the entire series y_1, \dots, y_n in a *backwards* recursion. This is called the Kalman smoother and, as the name suggests, provides a more accurate fit to the observed data. While the Kalman smoother is not used directly in the analysis in this paper, a particular aspect of it called disturbance smoothing is used for parameter estimation. The backward recursions for state smoothing are given by

$$\begin{aligned}
 r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t \\
 \hat{\alpha}_t &= a_t + P_t r_{t-1} & V_t &= P_t - P_t N_{t-1} P_t
 \end{aligned} \tag{6.4}$$

for $t = n, \dots, 1$,

with $r_n = 0$, $N_n = 0$, and N_t, V_t $m \times m$ matrices.

We write the smoothed estimates of the disturbance vectors ε_t and η_t as $\hat{\varepsilon}_t = E(\varepsilon_t | y)$ and

$\hat{\eta}_t = E(\eta_t | y)$. It can be shown that $\hat{\varepsilon}_t = H_t u_t$ where the *smoothing error* u_t is defined as

$u_t = F_t^{-1} v_t - K_t' r_t$. The smoothed estimate of the state disturbance, $\hat{\eta}_t$, is defined as $\hat{\eta}_t = Q_t R_t' r_t$. The recursions for the smoothed disturbances and their variance can be summed up as

$$\begin{aligned}\hat{\varepsilon}_t &= H_t(F_t^{-1}v_t - K_t'r_t) & \text{Var}(\varepsilon_t | y) &= H_t - H_t(F_t^{-1} + K_t'N_tK_t)H_t \\ \hat{\eta}_t &= Q_tR_t'r_t & \text{Var}(\eta_t | y) &= Q_t - Q_tR_t'N_tR_tQ_t\end{aligned}\quad (6.5)$$

for $t = n, \dots, 1$,

These definitions will be used for parameter estimation in the next section.

6.4.2 Maximum likelihood estimation

Following Durbin & Koopman (2001), the likelihood function of the Gaussian linear state space model can be written as

$$\log L(y) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n (\log |F_t| + v_t' F_t^{-1} v_t)$$

It is known as the *prediction error decomposition*. The quantities v_t and F_t output from the Kalman filter so $\log L(y)$ can be calculated simultaneously with the Kalman filter – speeding up the calculation process. This is convenient for numerical estimation of the unknown parameters, as many function evaluations are (typically) made before $-\log L(y)$ is minimized.

Define ψ as a vector of one or more elements of the system matrices Z_t , T_t , H_t , R_t and Q_t that have to be estimated using maximum likelihood. The dependence of the log-likelihood on ψ can be written as $\log L(y | \psi)$. In this paper a Quasi-Newton method of solving equations is used to minimize $-\log L(y)$ using the BFGS algorithm in MATLAB by means of the function ‘fminunc’.

Newton’s method solves the equation

$$\partial_1(\psi) = \frac{\partial \log L(y | \psi)}{\partial \psi} = 0$$

using the first-order Taylor series

$$\partial_1(\psi) \simeq \tilde{\partial}_1(\psi) + \tilde{\partial}_2(\psi)(\psi - \tilde{\psi}) \quad (6.6)$$

for some trial value $\tilde{\psi}$, where

$$\tilde{\partial}_1(\psi) = \partial_1(\psi)|_{\psi=\tilde{\psi}} \quad \tilde{\partial}_2(\psi) = \partial_2(\psi)|_{\psi=\tilde{\psi}}$$

with

$$\tilde{\partial}_2(\psi) = \frac{\partial^2 \log L(y|\psi)}{\partial \psi \partial \psi'}$$

By equating (6.6) to zero we obtain a revised value $\bar{\psi}$ from the expression

$$\bar{\psi} = \tilde{\psi} - \tilde{\partial}_2(\psi)^{-1} \tilde{\partial}_1(\psi)$$

This process is repeated until it converges or until a switch is made to another optimization method.

The gradient $\partial_1(\psi)$ determines the direction of the step taken to the optimum and the Hessian modifies the size of the step. It is possible to overstate the size the maximum in the direction determined by the vector

$$\tilde{\pi}(\psi) = \tilde{\partial}_2(\psi)^{-1} \tilde{\partial}_1(\psi),$$

and therefore it is common practice to include a line search along the gradient vector within the optimization process. We obtain the algorithm

$$\bar{\psi} = \tilde{\psi} + s \tilde{\pi}(\psi),$$

where various methods are available to find the optimum value for s , which is usually found to be between 0 and 1. The BFGS algorithm calculates an approximation of the Hessian $\partial_2(\psi)$ and updates it at each new value of ψ using the recursion

$$\bar{\partial}_2(\psi)^{-1} = \tilde{\partial}_2(\psi)^{-1} + \left(s + \frac{g'g^*}{\tilde{\pi}(\psi)'g} \right) \frac{\tilde{\pi}(\psi)\tilde{\pi}(\psi)'}{\tilde{\pi}(\psi)'g} - \frac{\tilde{\pi}(\psi)g^{*'} + g^*\tilde{\pi}(\psi)'}{\tilde{\pi}(\psi)'g}$$

where g is the difference between the gradient $\tilde{\partial}_1(\psi)$ and the gradient for a trial value of ψ prior to $\tilde{\psi}$ and $g^* = \tilde{\partial}_2(\psi)^{-1}g$. The BFGS method ensures that the approximate Hessian matrix remains negative definite. The 'score vector' $\partial_1(\psi) = \partial \log L(y|\psi) / \partial \psi$ specifies the direction in the parameter space along which a search should be made. The score vector takes the form

$$\begin{aligned} \left. \frac{\partial \log L(y | \psi)}{\partial \psi} \right|_{\psi=\tilde{\psi}} &= -\frac{1}{2} \frac{\partial}{\partial \psi} \sum_{t=1}^n \left[(\log |H_t| + \log |Q_{t-1}| + \text{tr} \left[\left\{ \hat{\varepsilon}_t \hat{\varepsilon}_t' + \text{Var}(\varepsilon_t | y) \right\} H_t^{-1} \right] \right. \\ &\quad \left. + \text{tr} \left[\left\{ \hat{\eta}_{t-1} \hat{\eta}_{t-1}' + \text{Var}(\eta_{t-1} | y) \right\} Q_{t-1}^{-1} \right] | \psi) \right]_{\psi=\tilde{\psi}} \end{aligned}$$

where $\hat{\varepsilon}_t$, $\hat{\eta}_{t-1}$, $\text{Var}(\varepsilon_t | y)$ and $\text{Var}(\eta_{t-1} | y)$ are obtained for $\psi = \tilde{\psi}$ as in section (6.4.1). ‘ \wedge ’ denotes that the parameter has been replaced by its maximum likelihood estimate.

6.4.3 Standard error of maximum likelihood estimates

The distribution of $\hat{\psi}$ for large n is approximately $\hat{\psi} \sim N(\psi, \Omega)$, where

$$\Omega = \left[-\frac{\partial^2 \log L}{\partial \psi \partial \psi'} \right]^{-1}.$$

Thus the standard error of maximum likelihood estimates are given as the square-root of the diagonal of Ω , $\sqrt{\text{diag}(\Omega)}$, where Ω is the inverse of the negative Hessian. The Hessian matrix is typically calculated using finite-difference methods, and may result in negative elements in the Ω matrix due to approximation and rounding errors. In that case it may be concluded that the model is misspecified, or the absolute value of $\text{diag}(\Omega)$ may be taken before the square-root, to obtain approximate standard errors (DURBIN, J. and Koopman, S., 2001).

6.4.4 The expectation maximization algorithm

An alternative method of parameter estimation is the expectation maximization (EM) algorithm. The EM algorithm initially converges faster than numerical optimization with BFGS, but slower near the maximum (or minimum) (SHUMWAY, R. H. and Stoffer, D. S., 2006). Therefore it is attractive to initially use EM and then switch to BFGS once the speed of convergence begins to fall. The analysis in this paper, however, only uses BFGS as complications arise with the EM algorithm when restricting which parameters in the system matrices should be changed, and which parameters should remain at a pre-specified value. The EM algorithm changes all elements in the error covariance matrices H_t and R_t , including the elements off the diagonals, which is not always desirable. For example,

independence between individual state and observation vectors cannot be forced, by setting values off the diagonals to zero.

7 ANALYSIS PART I: HIGH-FREQUENCY EQUITY DATA

This first part of the analysis investigates equity tick data from London Stock Exchange using vector autoregression as described in section 4.3. The aim is to extract information about the dynamics of trading from transaction data.

7.1 Data description

The data set used in the empirical analysis is tick data from London Stock Exchange obtained from Reuters Datascope. The data set includes all trade and quote events for Anglo American PLC during March of 2008. Every event is coupled with a time stamp and a qualifier (identification code). Error events and other abnormal events are also included in the data set and are also identified by qualifiers. The data set contains numerous electronically generated events such as an exchange calculated intra-day volume-weighted average price (VWAP). Appendix 11.1 provides details of the various event types.

7.1.1 Description of a representative stock: Anglo American PLC

Figure 7-1 is an example of tick data observations of Anglo American PLC from London Stock Exchange. The data are here shown as comma separated values, and the data values in rows 60, 75 and 76 are explained in Table 7-1.

7. Analysis Part I: High-frequency equity data

```

60 AAL.L,03-MAR-2008,08:00:33.651,+0,Quote,,,,,3159,3170,,,,
61 AAL.L,03-MAR-2008,08:00:33.717,+0,Quote,,,,,3159,3176,,,,
62 AAL.L,03-MAR-2008,08:00:33.797,+0,Quote,,,,,3159,3161,,,,
63 AAL.L,03-MAR-2008,08:00:34.673,+0,Trade,3170,,,,,OB Open[GEN],,
64 AAL.L,03-MAR-2008,08:00:36.900,+0,Trade,,,,,3170.0000,,,,,Mkt VWAP[GEN],,
65 AAL.L,03-MAR-2008,08:00:36.968,+0,Quote,,,,,3159,3161,,,,
66 AAL.L,03-MAR-2008,08:00:36.968,+0,Trade,3161,250,,,,,A[LSE],,
67 AAL.L,03-MAR-2008,08:00:36.968,+0,Trade,,,,,3169.8370,,,,,OB VWAP[GEN],,
68 AAL.L,03-MAR-2008,08:00:37.138,+0,Quote,,,,,3159,3161,,,,
69 AAL.L,03-MAR-2008,08:00:37.139,+0,Trade,3161,,,,,OB Low|Mkt Low[GEN],,
70 AAL.L,03-MAR-2008,08:00:37.139,+0,Quote,,,,,3159,3162,,,,
71 AAL.L,03-MAR-2008,08:00:37.139,+0,Quote,,,,,3159,3172,,,,
72 AAL.L,03-MAR-2008,08:00:38.340,+0,Quote,,,,,3159,3170,,,,
73 AAL.L,03-MAR-2008,08:00:40.401,+0,Quote,,,,,3159,3161,,,,
74 AAL.L,03-MAR-2008,08:00:40.431,+0,Quote,,,,,3159,3170,,,,
75 AAL.L,03-MAR-2008,08:00:40.431,+0,Trade,3161,502,,,,,A[LSE],,
76 AAL.L,03-MAR-2008,08:00:40.431,+0,Trade,,,,,3169.5280,,,,,OB VWAP[GEN],,
77 AAL.L,03-MAR-2008,08:00:40.494,+0,Quote,,,,,3159,3168,,,,
78 AAL.L,03-MAR-2008,08:00:42.220,+0,Quote,,,,,3159,3168,,,,
79 AAL.L,03-MAR-2008,08:00:46.412,+0,Quote,,,,,3160,3168,,,,
80 AAL.L,03-MAR-2008,08:00:46.413,+0,Quote,,,,,3160,3168,,,,
81 AAL.L,03-MAR-2008,08:00:46.413,+0,Quote,,,,,3160,3167,,,,
82 AAL.L,03-MAR-2008,08:00:46.451,+0,Quote,,,,,3160,3165,,,,
83 AAL.L,03-MAR-2008,08:00:46.681,+0,Quote,,,,,3160,3165,,,,
84 AAL.L,03-MAR-2008,08:00:49.473,+0,Trade,,,,,3169.5280,,,,,Mkt VWAP[GEN],,
85 AAL.L,03-MAR-2008,08:00:51.342,+0,Quote,,,,,3160,3165,,,,
86 AAL.L,03-MAR-2008,08:00:56.857,+0,Quote,,,,,3160,3165,,,,
87 AAL.L,03-MAR-2008,08:00:59.584,+0,Trade,3165,1131,,,,,A[LSE],,
88 AAL.L,03-MAR-2008,08:00:59.584,+0,Trade,3165,311,,,,,A[LSE],,
89 AAL.L,03-MAR-2008,08:00:59.584,+0,Trade,3165,333,,,,,A[LSE],,

```

Figure 7-1: Representative tick data for Anglo American PLC on March 3, 2008

Source: Reuters Datascope

Description	Reuters tag	Row 60 value	Row 75 value	Row 76 value
Stock quote	#RIC	AAL.L	AAL.L	AAL.L
Date	Date[G]	03-MAR-2008	03-MAR-2008	03-MAR-2008
Time	Time[G]	08:00:33.651	08:00:40.431	08:00:40.431
GMT offset	GMT Offset	+0	+0	+0
Trade / quote	Type	Quote	Trade	Trade
Traded price	Price	-	3161	-
Traded volume	Volume	-	502	-
VWAP	VWAP	-	-	3169.5280
Quoted bid	Bid Price	3159	-	-
Quoted bid size	Bid Size	-	-	-
Quoted ask	Ask Price	3170	-	-
Quoted ask size	Ask Size	-	-	-
Qualifiers	Qualifiers	-	A[LSE]	OB VWAP[GEN]

Table 7-1: Column labels for Anglo American PLC tick data.

A typical quote, trade and VWAP event is included. Source: Reuters Datascope

7. Analysis Part I: High-frequency equity data

As an example of a limit order book, the table below shows the 10 best bid and ask quotes at the time corresponding to row 74 in Figure 7-1:

Anglo American PLC Limit order book			
Date	March 3, 2008	Time	08:00:40.42
Bid	Volume	Ask	Volume
3159	419	3135	9344
3161	502	3179	5000
3155	4000	3122	11146
3170	23	3200	50
3151	3997	3120	10000
3172	666	3203	4085
3150	2150	3110	2000
3173	845	3204	1000
3145	1000	3100	1300
3176	1203	3240	150

7.2 Dealing with transaction time

There are at least four possible approaches to dealing with transaction time:

1. Hasbrouck's method (HASBROUCK, J., 1991): Give each trade or quote event a new t-index, but with two exceptions:
 - a. The appearance of a quote revision within 5 seconds prior to a transaction is a considered a reporting anomaly and the quote is resequenced after the trade
 - b. Quote revisions occurring within 15 seconds after a trade are given the same t-subscript as the trade
2. Simple method: Give each event in the tick data a new t-index. This implies trades and quotes will never be simultaneous, and therefore neither will trade events and quote revisions.
3. Alternative: Give each event in the tick data a new t-index. But trades and quotes with the same time stamp (at the millisecond level) are lumped together with the same t-subscript. Trades with the same time stamp and price are also combined.
4. Method (2) and (3) can be further modified by removing all quote events with no revision. This may reduce the concern of stale quotes, and can be used to gauge the importance of stale quotes by comparing results from a VAR analysis that uses all quote events, and one that

doesn't. This modification will obviously create larger clock time gaps between the transaction time indices.

The data used in Hasbrouck (1991) is from 1989, and is almost certainly less frequent than the FTSE 100 data from 2007-08 used in this paper. It is unfortunately not mentioned in Hasbrouck's paper how many observations there were in the data set, namely the 62 trading days of the first quarter of 1989. For the empirical analysis in this paper, modification (1.a) above seems dubious. If the quote revision 5 seconds before a given trade is, for instance, right after the previous trade, why should it then be sequenced after the second trade? Modification (1.b) seems biased towards emphasizing the causality running from trades to quote revisions.

Rewriting equations (4.6) and (4.8) using trade indicators x_t^0 instead of signed volume x_t and include constant terms we obtain the following VAR(p) system

$$r_t = a_0 + \sum_{i=1}^p a_i r_{t-i} + \sum_{i=1}^p b_i x_{t-i}^0 + v_{1,t} \quad (7.1)$$

$$x_t^0 = c_0 + \sum_{i=1}^p c_i r_{t-i} + \sum_{i=1}^p d_i x_{t-i}^0 + v_{2,t} \quad (7.2)$$

One particular problem that may arise due to modification (a) is that it may bias the magnitude of the coefficient of x_t^0 in model (7.1). Hasbrouck considers the positive coefficient of x_t^0 in the regression of r_t on lagged x_t^0 and r_t to be particularly important, as it is the average quote revision immediately subsequent to a trade (within 15 seconds). The coefficients of the subsequent lags (>0) are then the effect of trades on quote revisions beyond 15 seconds, each by an increment of transaction time 1. So the actual clock time between the transaction time increments changes considerably for lags greater than 0. This could be a source of bias in the analysis towards emphasizing the immediate impact of trades on quote revisions.

Another issue is how to deal with multiple trades with the same price and possibly the same time stamp. The trades are given each their own t-index because it cannot safely be assumed that the trades are made between the same two counterparties. It is possible that one of the two counterparties is the same for the simultaneous trades (same time stamp), but not necessarily both counterparties. Therefore each trade event must remain separate, as it contains information about

the behavior of agents in the market that will be lost if the trades are aggregated and given the same transaction time index. Each separate t-event can be seen as a unique decision by a market participant to initiate a trade or post a quote and therefore contains information.

It should be noted that the FTSE 100 data contains many repetitive price quotes and therefore many quote revisions will be zero. This is not a major concern, however, because the method of vector autoregression rests on the assumption of covariance stationarity which is not impaired by a time series dominated by zeros (or any other numerical value for that matter) (HASBROUCK, J., 2007).

7.3 Empirical analysis of tick data

To test the model specification of section 4.3, the VAR system (4.6)-(4.8) will be estimated using five lags. For quote revision r_t and signed trade indicator x_t^0 we write:

$$r_t = a_0 + \sum_{i=1}^5 a_i r_{t-i} + \sum_{i=1}^5 b_i x_{t-i}^0 + v_{1,t} \quad (7.3)$$

$$x_t^0 = c_0 + \sum_{i=1}^5 c_i r_{t-i} + \sum_{i=1}^5 d_i x_{t-i}^0 + v_{2,t} \quad (7.4)$$

The input data is all trade and quote events of Anglo American PLC in March 2008.

7. Analysis Part I: High-frequency equity data

Data input			
Company	Anglo American PLC		
Market value (bn £)	38.0	Avg trade volume	637.89
Stock quote	AAL.LN	Median trade volume	383
Start date	03-Mar-08	Low price	2,673.00
End date	31-Mar-08	High price	3,547.00
Trading days	19	Average price	3,101.40
Observations	1,253,636	VWAP	3,092.70
Events removed	195,692	Min spread	1
Events used	1,057,944	Max spread	72
Trade events	234,832	Avg spread	2.6
Quote events	823,112	Median spread	2
Nonevents	571,939	Min t-time increment	0
Total volume	149,037,283	Max t-time increment	76.516
Avg daily volume	7,844,068	Mean t-time increment	0.5452
Min trade volume	1	Median t-time increment	0.037
Max trade volume	2,053,739.00		
Volume in shares, prices in GBP and t-time increments in seconds			

The results from the estimation are the following:

	Lag	Coeff.	T-stat		Lag	Coeff.	T-stat
<i>a</i>	0	0.0004	0.7	<i>c</i>	0	-0.0010	-2.4
<i>a</i>	1	-0.0881	-90.3	<i>c</i>	1	-0.1011	-138.4
<i>a</i>	2	-0.0353	-35.7	<i>c</i>	2	-0.0162	-21.8
<i>a</i>	3	-0.0322	-32.6	<i>c</i>	3	-0.0071	-9.5
<i>a</i>	4	-0.0028	-2.8	<i>c</i>	4	-0.0044	-6.0
<i>a</i>	5	-0.0152	-15.8	<i>c</i>	5	0.0015	2.1
<i>b</i>	1	0.1953	149.8	<i>d</i>	1	0.3175	325.4
<i>b</i>	2	0.0484	35.0	<i>d</i>	2	0.0875	84.3
<i>b</i>	3	0.0195	14.0	<i>d</i>	3	0.0367	35.3
<i>b</i>	4	0.0143	10.3	<i>d</i>	4	0.0244	23.5
<i>b</i>	5	0.0112	8.5	<i>d</i>	5	0.0287	29.0
<i>R</i> ²		0.0423				0.1573	

Table 7-2: Results from vector autoregression using quote revisions and signed trade indicators from Anglo American PLC in March 2008.

Coefficient sums		
p	5	10
a	-0.1736	-0.1773
b	0.2887	0.3119
c	-0.1272	-0.1400
d	0.4948	0.5425
R^2		
r_t	0.0423	0.0428
x_t	0.1573	0.1588

Table 7-3

All the estimated coefficients are statistically significant at the 5% level. The R^2 value of 0.0423 for eq. (7.3) is low, but typical for analysis using high-frequency data (HASBROUCK, J., 1991). The negative signs of the a_i coefficients are evidence of reversal in quote revisions. The positive signs of the b_i and d_i coefficients show that prices rise as a response to buying, but trades do not immediately reverse direction. Signed trades actually show positive autocorrelation, which is inconsistent with microstructure models of inventory control. If inventory control considerations were dominant, the d_i coefficients would be negative, as market makers would raise prices in response to purchases, and

lower prices in response to sales – in order to replenish inventories (O'HARA, M., 1997). This would have an effect of trade reversal.

A possible explanation for this inconsistency with microstructure theory is that the traditional role of market makers has been replaced by electronic limit order books. Inventory considerations have been replaced by short term trade momentum caused by algorithms that divide trades into several orders, or use iceberg orders to the same effect.

The calculations were repeated using 10 lags and the results are shown Table 7-3. R^2 rises only slightly for each equation, but the sums of the coefficients change significantly. Table 11-1 in Appendix 11.5 shows detailed results of the estimated coefficients. All a and d coefficients are significant at the 5% level while the b and c coefficients are significant to lag 7 and 6 respectively.

To understand the short-term dynamics of the VAR model, we calculate $\alpha_m(v_{2,0})$ from eq. (4-11). It is the response of r_t to a time $t = 0$ trade innovation, $v_{2,t}$, of 1 unit (while $v_{1,t} = 0$) and can be written as $\alpha_m(v_{2,0}) = \sum_{i=1}^m r_i$. Figure 7-2 shows that the adjustment is rapid, after only 5 periods the majority of the response has taken place. In response to a trade, the quote midpoint has been revised higher by GBP 0.46 on average after twenty periods of transaction time. It is possible to interpret the response function $\alpha_m(v_{2,0})$ for eq. (7.4) as the likelihood that a trade event will be followed by another trade event after m periods. If $\alpha_m(v_{2,0}) > 0$ the sign of the trade is persistent, whereas $\alpha_m(v_{2,0}) < 0$ indicates trade reversal. In this case $\alpha_{20}(v_{2,0}) = 0.8634$, which implies that the

likelihood that a trade event will be followed by another trade with the same sign within 20 periods is 86.34% (TSAY, R. S., 2005). Note that the calculation of $\alpha_m(v_{2,0})$ involves the two-way dynamics between r_t and x_t , the initial trade innovation causes a quote revision which has an impact on future trades – that in turn impact future quotes and so on.

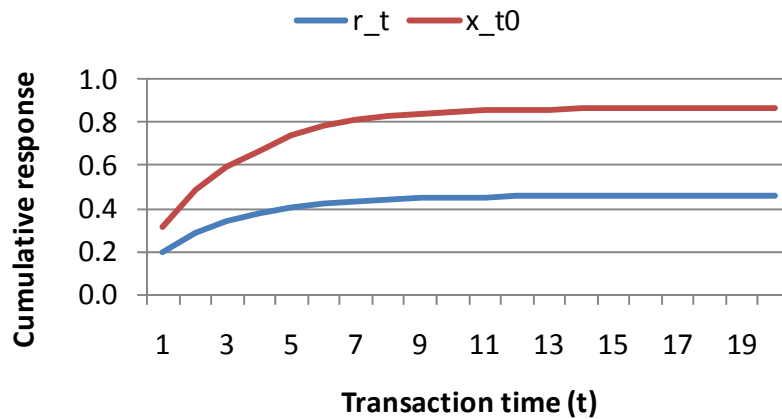


Figure 7-2: The response of r_t and x_t^0 through period 20 to a unit trade innovation at time $t = 0$.

It is possible that the response function $\alpha_m(v_{2,0})$ is affected more by larger trades than by small. To see if this is the case we estimate model (7.7)-(7.8) again using the signed volume x_t in place of x_t^0 .

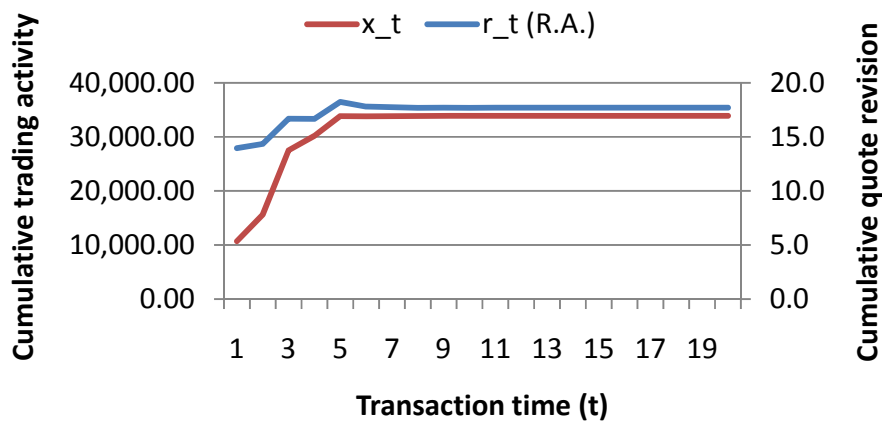


Figure 7-3: The response of r_t and x_t through period 20 to a purchase of 2 mln shares at time $t = 0$.

The above figure shows the impact of a time $t = 0$ purchase of 2 mln shares. A trade of this magnitude is on average followed by additional purchases of 33,900 shares and an upwards quote revision of GBp 17.7.

Subconclusion

The results from the vector autoregression analysis confirm those of Hasbrouck (1991), even though Hasbrouck's analysis was done with transaction data from 1989, 19 years older than the data used for this analysis. Noting the limited scope of the data sample here used, we may conclude that although markets have become more electronically driven over the past few decades, the basic dynamics between prices and trades remain largely unchanged.

8 ANALYSIS PART II: SPECULATIVE ALGORITHMS

Of the three types of speculative algorithms described in section 5.3, the first two will be illustrated using empirical data, namely a momentum strategy and a relative-value strategy. The goal is to illustrate how speculative algorithms can be constructed, and to determine whether they can make statistically significant positive profits out of sample. The process of designing speculative algorithms is arbitrary; there are many ways of approaching the problem, and many solutions. The same strategy may be implemented in several ways, of course with slight variations in results. Different variations of the same strategy will perform well under certain market conditions, and worse in others. The optimal strategy for a given security or market is therefore likely to evolve over time along with changing market conditions. The founder of a prominent algorithmic trading hedge fund explains the process of combining old strategies with new ones in a continuous process: “What you need to do is pile them up. You need to build a system that is layered and layered. And with each new idea, you have to determine, Is this really new, or is this somehow embedded in what we've done already? So you use statistical tests to determine that, yes, a new discovery is really a new discovery. Okay, now how does it fit in? What's the right weighting to put in? And finally you make an improvement. Then you layer in another one. And another one” (LUX, H., 2000).

8.1 Momentum strategies using moving averages

As described in section 5.3, momentum strategies are designed to exploit the trending behavior of markets. To understand the concept of a momentum strategy we first define a trend. An arithmetic random walk with a deterministic trend, or drift, may be written as

$$x_t = \mu + x_{t-1} + w_t = \mu t + \sum_{j=1}^t w_j \quad (8.1)$$

$$x_0 = 0 \quad t = 1, \dots, n \quad w_t \sim iidN(0,1)$$

where μ is the drift term and w_t is white noise.

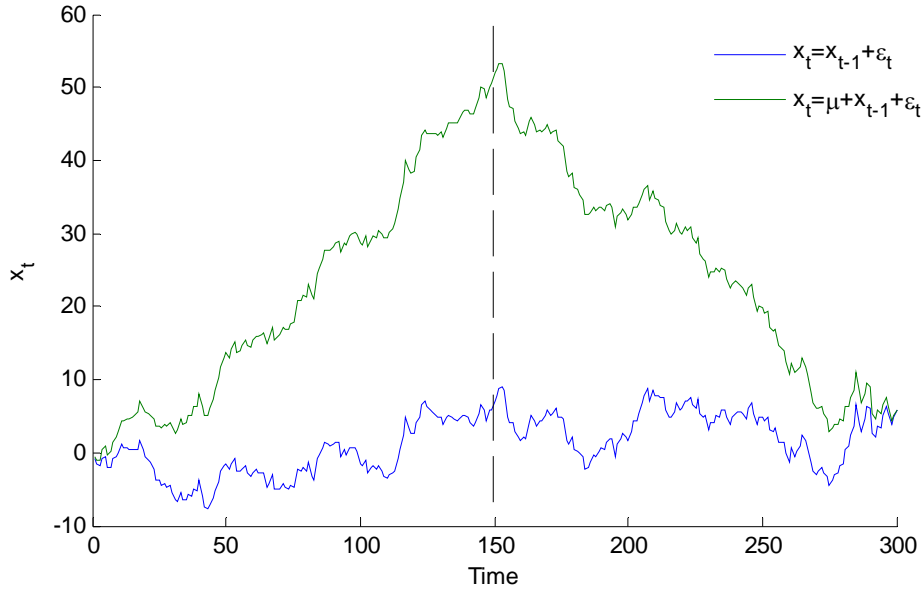


Figure 8-1: A random walk (RW) and random walk with drift. For the RW with drift,

$$\mu = \begin{cases} 0.3 & t = 1, \dots, 150 \\ -0.3 & t = 151, \dots, 300 \end{cases}$$

The standard model of stock prices mentioned in section 5.2.1 is based on geometric Brownian motion as

$$dS_t = \mu S_t dt + \sigma S_t dz \quad (8.2)$$

where dz denotes a Wiener process (HULL, J. C., 2006). The trend μ is now multiplied by the stock price S_t , but the effect of having a positive or negative trend is the same as in Figure 8-1. In (8.2) the trend μ is constant, but we may consider a model in which μ is a function of time t by writing

$$dS_t = \mu(t) S_t dt + \sigma S_t dz \quad (8.3)$$

This model may be approximated in discrete time by

$$\begin{aligned} \Delta S_t &= \mu(t) S_t \Delta t + \sigma S_t \Delta z \\ \Delta z &= \xi_t \sqrt{\Delta t} \quad \xi_t \sim iid N(0,1) \end{aligned} \quad (8.4)$$

It is impossible to predict when the drift parameter changes, but it is possible to observe whether it is positive, negative or zero over a given period of time. Therefore, if we assume that there is some persistence in the sign of μ , that is, μ is autocorrelated, we may position ourselves to take advantage of future positive or negative drift in the price. A possible specification for $\mu(t)$ is the AR(p) process

$$\mu(t) = \mu_t = \sum_{k=1}^p \mu_{t-k} + \varepsilon_t \quad (8.5)$$

where ε_t is a white noise error process.

One way of quantifying the existence and direction of a trend is to use exponentially weighted moving averages (EMAs)¹⁴. The exponential moving average \tilde{x}_t can be written as the infinite sum

$$\tilde{x}_t = \sum_{j=1}^{\infty} \lambda(1 - \lambda)^{j-1} x_{t-j+1} + w_t \quad (8.6)$$

where x_t is the underlying time series and w_t denotes a white noise error process. Depending on the smoothing parameter λ , less or more weight is given to new observations. λ may be defined as $\lambda = 2/(R + 1)$ where the constant R dictates the responsiveness of the EMA to new observations. Lower values of R give a more responsive EMA by increasing λ . This can be seen more clearly in the recursive representation¹⁵ given by

$$\tilde{x}_t = (1 - \lambda)\tilde{x}_{t-1} + \lambda x_t \quad \tilde{x}_0 = 0, \quad 0 < \lambda < 1 \quad (8.7)$$

(SHUMWAY, R. H. and Stoffer, D. S., 2006)¹⁶. Using two EMAs with different values of R , the more responsive called the leading EMA and the less responsive called the lagging EMA, a simple trading rule buys the security when the leading EMA is higher than the lagging EMA and vice versa. Figure 8-2 shows two EMAs with $R = \{10, 25\}$ superimposed on daily observations of the FTSE 100 index with the trading position on any given day shown below. The strategy is either long one unit of the index, or short one unit, i.e. the strategy is exposed to changes in market prices at all times. The values of R were chosen arbitrarily and the resulting profit and loss of the trading rule for a more extended period is shown in Figure 8-3. Trading costs are ignored and it is assumed that it is possible to trade the index at the closing price of each day.¹⁷

¹⁴ They are sometimes abbreviated EWMA.

¹⁵ See Appendix 11.6 for a derivation.

¹⁶ Note that the EMA equations in Shumway & Stoffer (2006), (3.132) and (3.133) have here been modified to correspond to more common versions by substituting $(1 - \lambda)$ for λ and updating the EMA at time t , \tilde{x}_t , using the contemporary new observation at time t , x_t .

¹⁷ This could be done using exchange traded funds (ETFs) or index futures. In this case the actual index values are used for illustrative purposes.

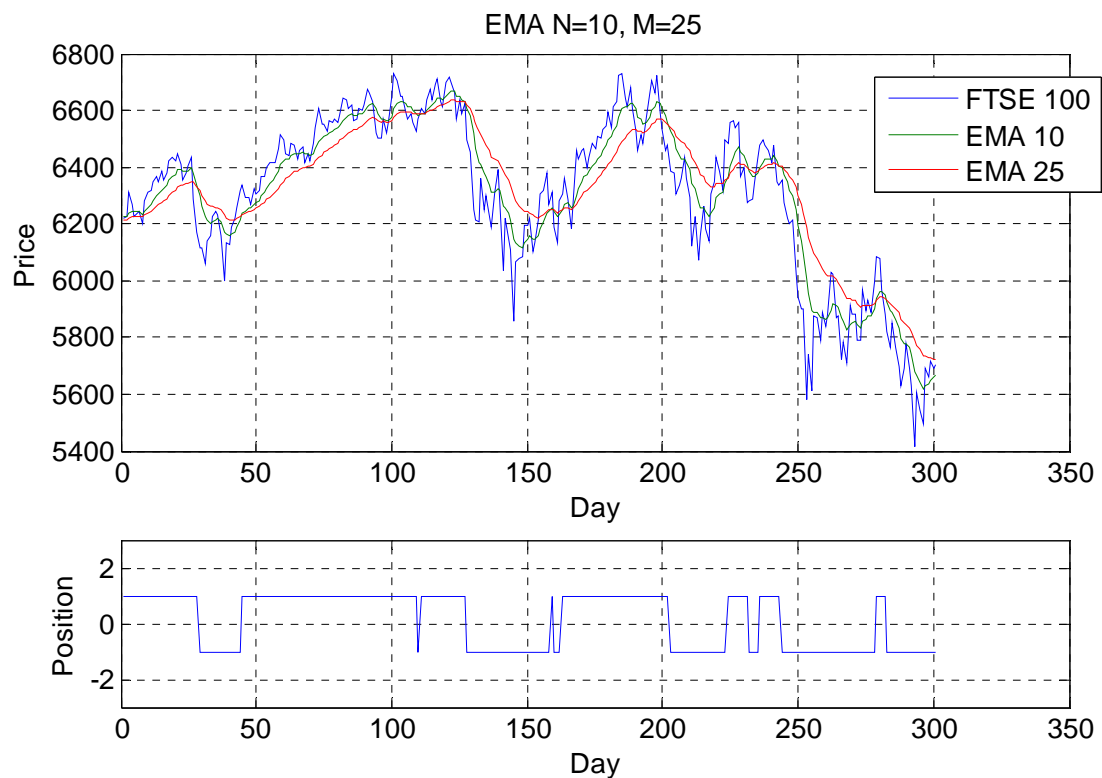


Figure 8-2: EMA trading rule applied to the FTSE 100 Index from January 19, 2007 to March 28, 2008.

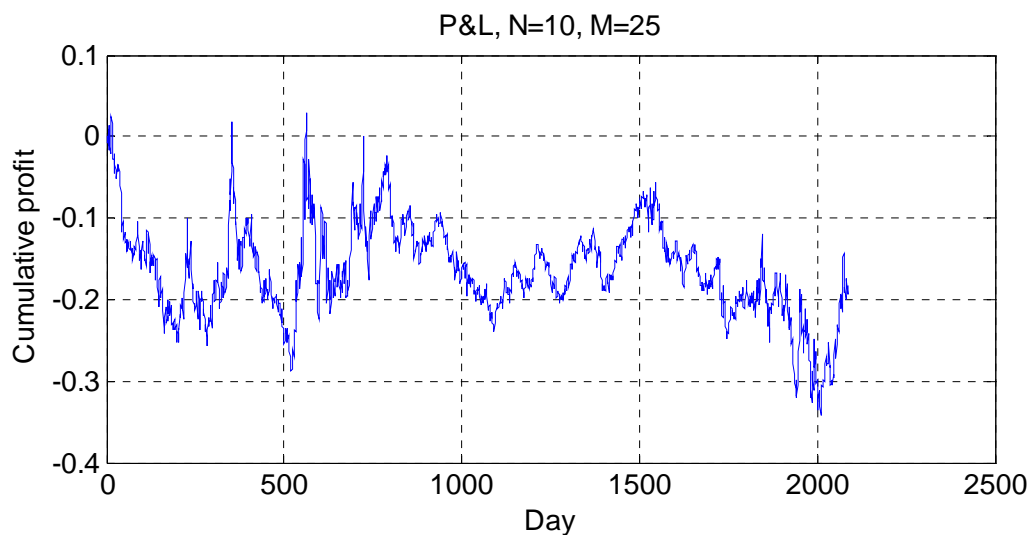


Figure 8-3: The cumulative profit and loss of the EMA strategy in Figure 8-2 from May 2, 2000 to August 1, 2008.

The return from the trading rule over this period is -16.68% (-2.19% annualized) with a Sharpe ratio¹⁸ of -0.12. To find which combination of leading ($R = N$) and lagging ($R = M$) EMAs provides the highest Sharpe ratio over the evaluation period, a back-test was run for a large number of combinations of N and M .

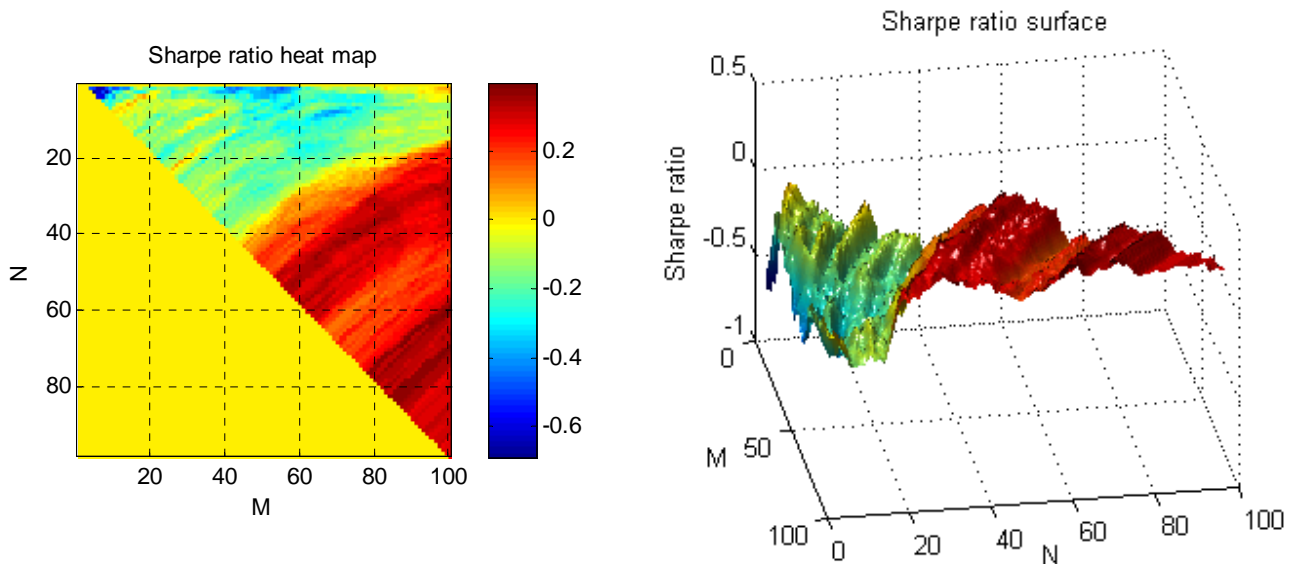


Figure 8-4: FTSE 100 back-test:
Sharpe ratio heat map and surface for EMA strategy with varying values of N and M .
 Source: Author's calculations, (KASSAM, A., 2008).

Figure 8-4 shows a heat map and surface plot of the Sharpe ratios obtained with various combinations of N and M . Clearly large values of N and M perform best, and the optimal solution is $R = \{61, 98\}$ with a Sharpe ratio of 0.3986 and a cumulative return of 82.43% or 7.55% annualized. The heat map is upper triangular because N must always be smaller than M . Figure 11-3 in Appendix 11.7 shows a graph of the FTSE 100 index with the optimal EMAs and the cumulative return superimposed.

So far the analysis has been based on daily observations of the FTSE 100 index. The index has trended strongly during this period, and this creates a bias towards momentum strategies with large values of R . Other assets show a different kind of trending behavior, or perhaps even mean-reverting behavior. This is often the case for interest rates such as the price of the German 10-year government bond or *Bund*. Applying the same procedure as above to daily observations of Bund futures gives optimal R -

¹⁸ The Sharpe ratio is here defined as $\sqrt{252}\mu/\sigma$ where μ is the average daily return and σ is the standard deviation of the daily returns. I.e. the risk-free rate is not subtracted from μ . This is also known as the *information ratio*.

values of $R = \{3,23\}$ which is significantly less than for the FTSE 100 index. The cumulative return is in this case 26.51% (7.05% annualized) with a Sharpe ratio of 1.4080. Figure 8-5 shows a heat map and surface plot of the Sharpe ratios obtained with various combinations of N and M for Bund futures prices sampled daily. The much higher Sharpe ratio shows that the Bund data is better suited to EMA-momentum strategies than the FTSE 100 index when sampled daily.

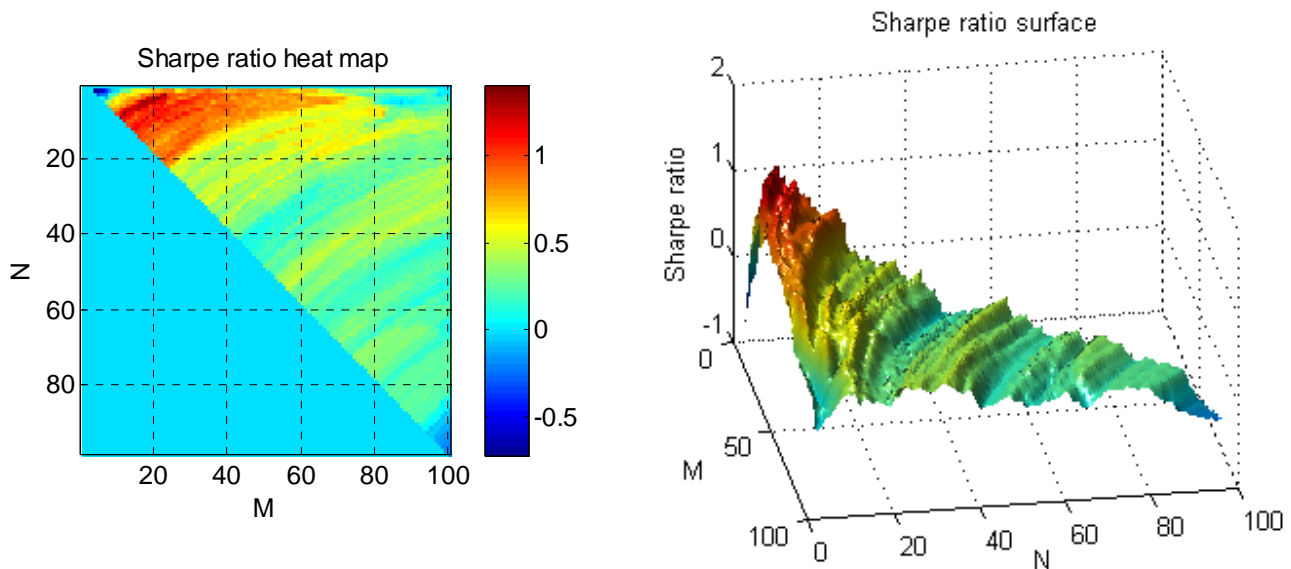


Figure 8-5: Bund back-test:
Sharpe ratio heat map and surface for EMA strategy with varying values of N and M .
Source: Author's calculations, (KASSAM, A., 2008).

8.1.1 EMA-momentum strategy with different sampling frequencies

It is interesting to consider the effect of different sampling frequencies on the performance of the EMA-momentum strategy. As described in section 4.3.1, the most detailed level of market information, namely tick data, is usually sampled in bins to show price movements in fixed time intervals. The security markets that generate the tick data trend over time, but the trends may go in different directions on different time scales. At a given point in time the Bund may be trending upward on a monthly basis, downward on a daily basis, upward on a 15-minute basis, and so on. Like the concept of collecting trade data in bins, the concept of a trend can be seen as a scaling phenomenon - it can be observed across the entire range of sampling frequencies.

Therefore, a strategy that is designed to exploit trending behavior, such as the EMA-momentum strategy, should be calibrated to find the most profitable sampling interval to operate in. This can be done by testing the strategy for different R -values as well as different sampling frequencies. In this case it is done using Bund futures data with sampling frequencies ranging from 1 minute to 11 hours (a full trading day). The Sharpe ratios for various combinations of the three parameters N , M and sampling frequency F is depicted in Figure 8-6. The data sample spans from September 15, 2002 to March 21, 2007.

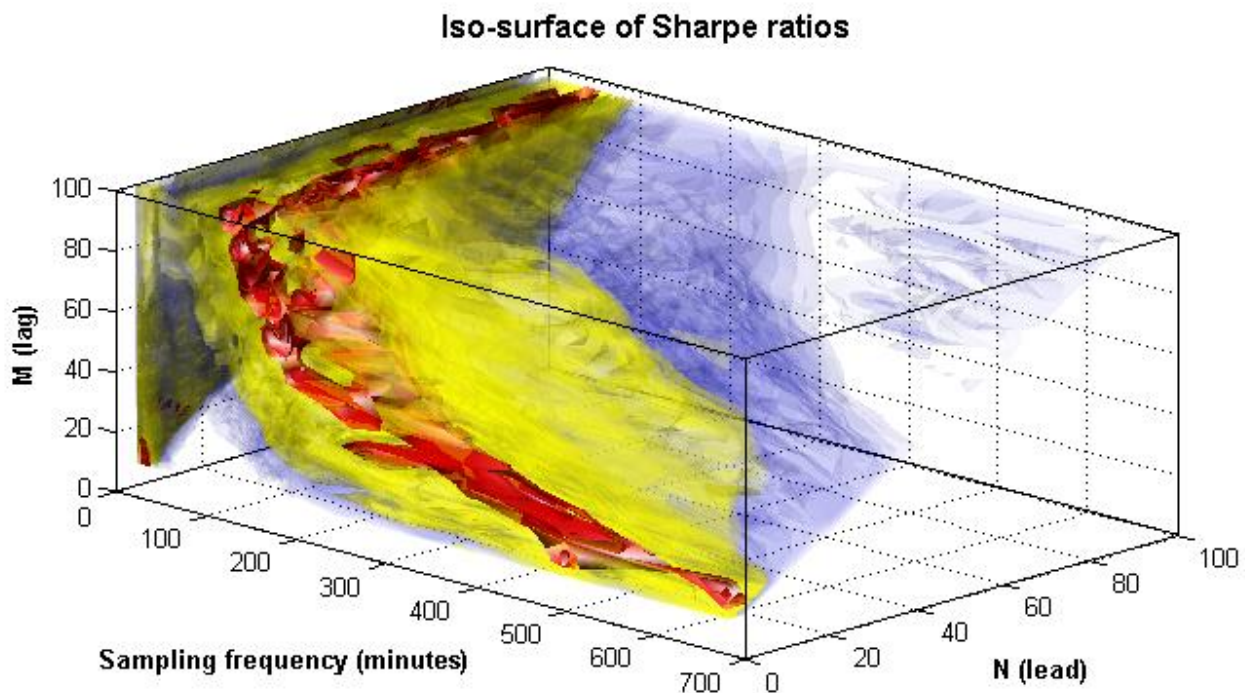


Figure 8-6: Iso-surface¹⁹ of Sharpe ratios for various sampling frequencies.
Legend for Sharpe ratios: Blue {0.0-0.7}; yellow {0.7-1.4}; red {1.4-2.1}.
Source: Author's calculations, (KASSAM, A., 2008).

The performance of the EMA-momentum strategy shows great variety across the range of sampling frequencies, and many combinations of the parameters yields Sharpe ratios above 1.4 (shown in red in Figure 8-6). Indeed for every sampling frequency there is a combination of R values for the lead and lag EMAs that yields a Sharpe ratio above 1.4. The highest Sharpe ratio obtained in the sample

¹⁹ An iso-surface is a 3D surface representation of points with equal values in a 3D data distribution. It is the 3D equivalent of a contour line.

was 2.05 and resulted from $R = \{N, M\} = \{8, 12\}$ with a sampling frequency of 5 minutes. The corresponding cumulative return was 29.46% or 5.8853% on an annualized basis, also the highest in the sample. This is a remarkably high number for a strategy with annual volatility of 4.93%. But due to the high sampling frequency, the strategy trades very frequently, namely 1,159 times per year on average. Assuming that transactions costs are 0.03%, this is equivalent to a loss of -34.77% per year, and wipes out any positive gains from trading. This shows that while there may be attractive algorithmic trading opportunities at high sampling frequencies, transaction costs quickly become a big issue. Figure 5-6 shows another iso-surface of Sharpe ratios now including transaction costs of 0.03% per trade.

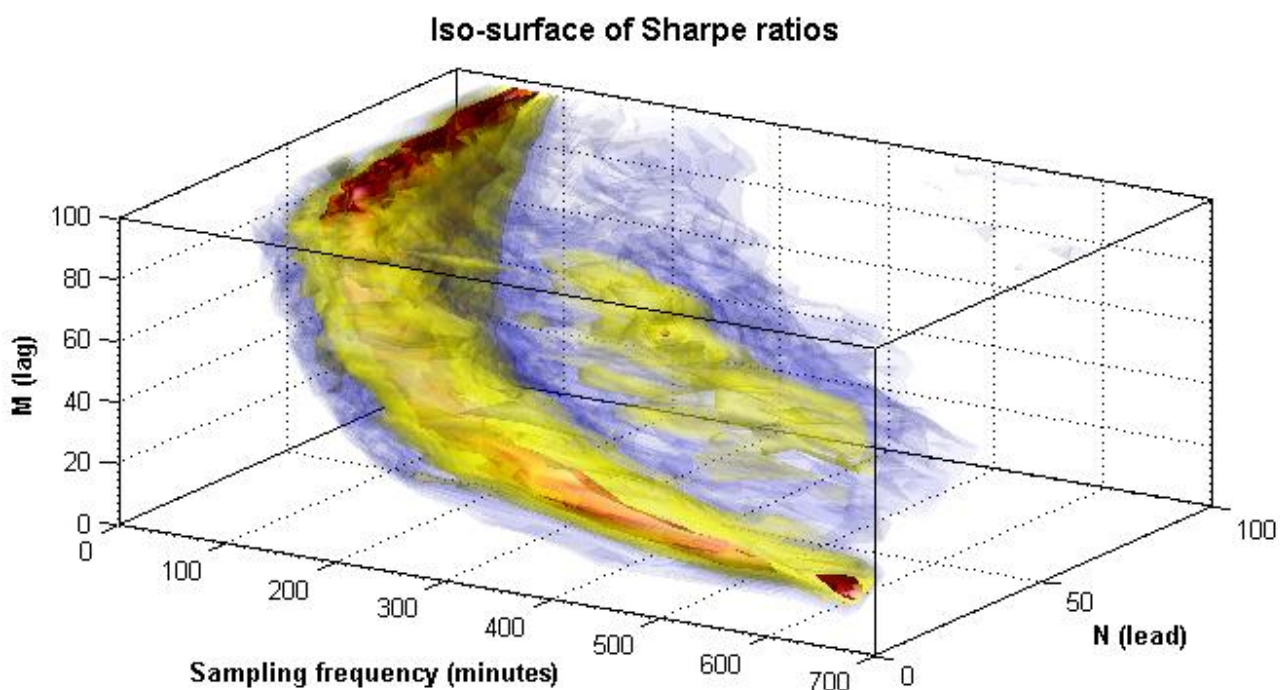


Figure 8-7: Iso-surface of Sharpe ratios for various sampling frequencies including transaction costs.
Legend for Sharpe ratios: Blue {0.0-0.4}; yellow {0.4-0.8}; red {0.8-1.2}.
Source: Author's calculations, (KASSAM, A., 2008).

Like in Figure 8-6 only positive Sharpe ratios are included, and now there are significantly less data points. The largest concentration of high Sharpe ratios is found at low sampling frequencies coupled with relatively high M -values. Other high Sharpe ratios are seen at very high sampling frequencies for

low values of N and M . The highest Sharpe ratio is 1.1484 and resulted from $R = \{N, M\} = \{10, 72\}$ with a sampling frequency of 120 minutes. This strategy also has the highest cumulative return which is 15.66% or 4.31% annualized with volatility of 4.88% per annum. It has very attractive properties, particularly the trade-off between risk and return, and its ability to make money in rising and falling markets. And since it is based on futures prices, it is easy to leverage (a typical margin requirement for Bund futures is 2-5% of the notional amount of one contract).

The investigation here has been done in-sample, in the sense that no 'training' period was used to determine which combination of the parameters N , M and F to use out-of-sample. Yet, the strategy is remarkably stable over time in different market environments for a large number of combinations of the parameters. A more thorough back-testing of the EMA-momentum strategy would require out-of-sample testing, however, and include a more realistic treatment of trade execution and transaction costs. Trade execution can be made more realistic by using historical bid/ask prices (rather than the 'close ask' as in this analysis), as those are the best estimates of prices that can actually be traded at a given point in time. Market depth and varying liquidity conditions over time should also be taken into account. The importance of this will depend on the desired amount of market exposure taken with the strategy. As an example, the depth of the market in Bund futures places limits on how many contracts can be sold or purchased at any given time without moving the bid/ask spread. Yet relative to other markets, such as the markets for individual stocks, the Bund futures market is highly liquid, and will therefore be suitable for the EMA-momentum strategy. But even in liquid markets, the EMA-momentum algorithm should be combined with an optimal execution algorithm that steps in once a trading signal has been created to minimize the implementation shortfall.

8.2 Univariate pairs trading

This section will present empirical results from a pairs trading back-test routine. In order to resemble an institutional setting, the trading algorithm will search for pairs in a large population of stocks and trade the pairs simultaneously as a portfolio. The results will show whether it has been possible to earn positive excess returns from pairs trading over an extended period of time.

The empirical analysis uses 75 stocks in the FTSE 100 index traded on London Stock Exchange as the population in which to search for pairs (the remaining 25 were not actively traded during the entire data period). The data set is 2085 daily observations of closing prices from May 2, 2000 to August 1, 2008.

As explained in section 5.3.4, trading signals are created when the difference between the normalized price of a stock in the data set and its pair exceeds a certain pre-specified barrier level b . The strategy uses a rolling window in which to ‘train’ the algorithm; that is to find a pair for each stock based on the minimum distance criteria. This rolling window may be for instance 252 trading days which is equivalent to a calendar year. For each trading period starting on the final day in the initial ‘training window’, the difference e_{it}^* in normalized prices is calculated for each pair. If the difference exceeds the barrier level b , then a trading signal is created for the first trading period. If d is positive, the stock being considered is trading more expensive than its pair, so a short position is taken in this particular stock and a long position is taken in its pair. If d is negative, the opposite happens, so the stock being considered is bought (long position) and its pair is sold (short position).

The pairs are reevaluated at fixed intervals during the data period. The interval may be anything from a few weeks to a year; a short interval will keep the pairs and positions up to date, but will incur high transaction costs, while a long interval risks having outdated positions in a changing market environment. Several different lengths of training windows (TW) and evaluation periods (EP) will be attempted.

The initiated positions are kept until one of two events:

1. Convergence between the stock and its pair is achieved, i.e. the difference in normalized prices goes to zero or below.
2. The pair has not converged after two periods *subsequent* to the evaluation period in which the pairs trade was opened. So the strategy has a memory of maximum three evaluation periods in the case where the pairs trade is opened on the first day in an evaluation period. Virtually no positions take longer than two periods to converge, so it is not an unreasonable limitation.

An example of a pairs trade is shown in Figure 8-8. The distance e_{it}^* is shown on the y-axis, the red lines denote the barrier levels $b = \pm 0.5637$, the target stock (tgt) is National Grid PLC and its pair is

Liberty International PLC. The time period is from October 13, 2004 to April 12, 2005. Five trades are made over the period for a profit of £120,600 with an exposure of £2mIn, which is approximately 12% annualized. The cointegration statistic of the pair is -4.3243 , which is well below the critical value of -1.96 . The squared distance in normalized price space, for this pair, $\sum_{t=1}^{125} (e_{it}^*)^2 = 19.78$ is close to the minimum of 18.27 (British Land Co PLC also with Liberty International PLC) for the period and can be compared to a mean of 78.15 for all pairs in the period.

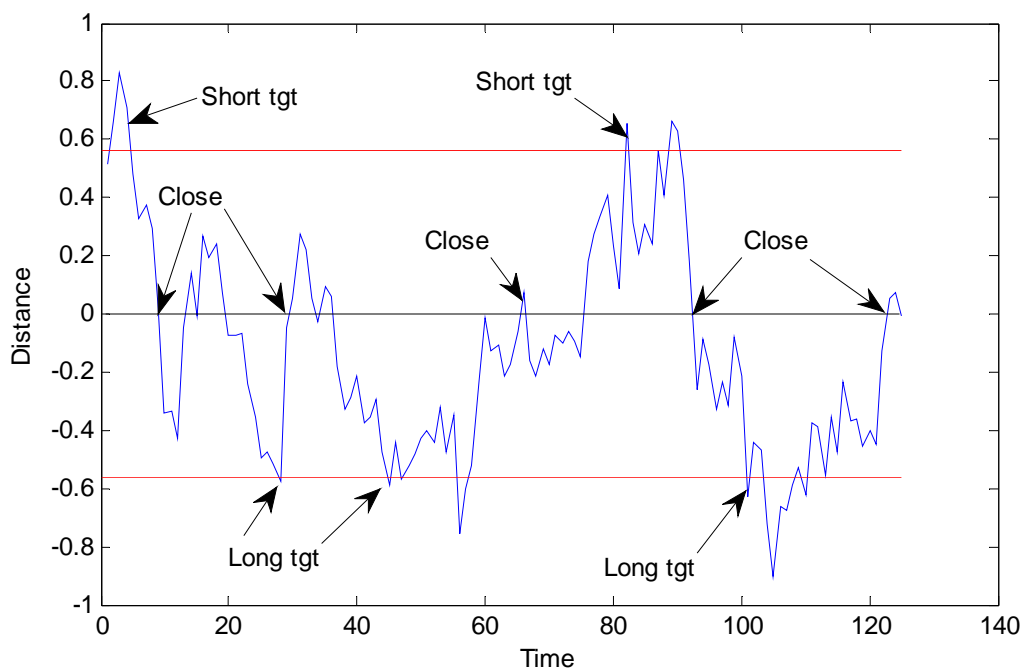


Figure 8-8: Distance between National Grid PLC (tgt) and Liberty International PLC.
Evaluation period number 8/15 using TW = 250 days and EP = 125 days.

It is worth noting that this is the only period of 15 in which these two stocks form a pair, and that National Grid PLC is the pair of 15 other stocks during the data period.

The back-test routine

The univariate pairs trading back-test routine is depicted in Figure 11-4 in Appendix 11.8, and a description of each step in the routine follows:

1. The first step is to preallocate memory to the various vectors and arrays that keep track of prices, positions and other variables over time.

2. The next step is to make initial calculations of daily returns for the stocks during the training window.
3. Then the main loop is initiated, which repeats for each day beginning on the last day of the initial training window. The first day is also the beginning of the first evaluation period, so the pair of each stock is identified using the minimum distance criteria. Duplicate pairs are found and removed. Individual stocks may be paired with more than one other stock, however.
4. The 'quality' of each pair is evaluated by either 1) calculating the correlation between historical price returns of the target stock and its pair or 2) calculating the augmented Dickey-Fuller coefficient to determine the degree of cointegration between the two. If the results are satisfactory, the given pair is considered 'active'.
5. For each pair $i = 1, \dots, n$ the barrier level b_i is calculated and compared to the price difference e_{it}^* . If the conditions in step 4 are fulfilled and $|e_{it}^*| > b_i$, then the appropriate position is opened in pair i . b_i is calculated as $b_i = std(e_{it}^*) * a_i$ for $a_i > 0$ (calculated using the rolling window).
6. Step 5 is repeated on each day until the beginning of the next evaluation period on which we start over with step 3. Pairs, barrier levels and positions from the previous periods are remembered in order to preserve open positions that haven't converged. The memory of the algorithm is two evaluation periods in addition to the one in which the position was opened.
7. At the beginning of each new day, all positions are updated according to market movements.
8. When the end of the data set is reached the positions are aggregated and results and risk key figures are calculated.
9. Finally, the results are reported.

Features of the implementation

Gross market exposure is measured as:

$$\text{Gross market exposure} = |\text{long positions}| + |\text{short positions}| \geq 0 \quad (8.8)$$

Net market exposure is measured as:

$$\text{Net market exposure} = \text{long positions} + \text{short positions} \quad (8.9)$$

- When a pair is opened, a position of £1 is taken in both the target stock (+1) and its pair (-1). So if a pair is opened on day 1, net market exposure will be £0. On each day the positions are adjusted in size according to market movements. For example, if the target stock rises by 10% on the following day, the position is now £1.1. If the pair stock is unchanged, net market exposure at the end of the day will be £0.1.
- The results are reported in the equivalent units. So for example, a cumulative return of £30 for the strategy implies that the profit is equal to 15 times what was initially invested in each individual pair.
- Transaction costs are fixed at 0.1% of the nominal amount invested when a position is opened, and again when it is closed. If the position has increased in magnitude, the cost of closing the position will be higher in absolute terms. The chosen value is arbitrary and includes bid-ask spread and any fixed trading fees. It may be biased to the downside, especially if the amount invested in each pair is large enough to move the bid-ask spread.

Managing capital

The performance of the pairs trading strategy may be reported in several different ways. In this analysis three methods will be used.

- Raw excess return
- Return to committed capital
- Return from a 'fully invested' strategy

The **raw excess return** is the profit generated by the strategy using the market exposure dictated by the strategy. It shows how successful the strategy has been in absolute terms (£ sterling) without taking into account the amount of capital invested in the strategy.

The **return to committed capital** compares the profit generated by the strategy to a prespecified initial investment that is considered large enough to sustain the market exposure taken by the strategy. The committed capital must be large enough to cover any margin requirements by a broker, and gives a fairly realistic idea of the returns that may be generated in an institutional setting. An investment bank would most likely measure committed capital in terms of regulatory capital, but it is hard to speculate in nominal terms what this may be, so using committed capital in absolute terms

provides an intuitive alternative. Committed capital does not earn the risk-free rate of interest in times of little or no market exposure.

The **fully invested** return measures performance under the assumption that capital allocated to the strategy is adjusted daily to match the gross market exposure of the strategy. It gives a less realistic view of performance than the committed capital strategy, as it is unrealistic that a hedge fund or trading desk is allocated capital on a one-day basis. But it allows us to compare the performance of pairs trading to the performance of holding the market portfolio with the same gross exposure.

Results

The pairs trading strategy was initially back-tested on the entire sample of price observations, with no upper limit on the number of possible pairs (in the population of 75 stocks). The following parameters were used in the initial back-test:

- Barrier parameter $a_i = 2$
- Minimum required cointegration test statistic (aDF) = 3
- Rolling windows of between 25 and 500 days (required to be $\geq 2 * EP$)²⁰
- Evaluation periods between 4 and 125 days

The best performing strategies had rolling windows between 50 and 150 days, and evaluation periods between 5 and 25 days. The optimal strategy had TW = 150 days and EP = 7 days. It earned £40.53 over the data period with an annualized information ratio of 1.75.²¹ The maximum drawdown using committed capital of £25 was 7.85% on a single day. See Appendix 11.8 for additional graphs of the results. Figure 8-9 shows the annual performance of the optimal strategy compared to a long position in the FTSE 100 index with *equivalent risk* defined as daily return volatility.

²⁰ The restriction is caused by programming issues. Strategies that didn't fulfill this criterion weren't used. They are found in the lower-left corner of the heat maps in Appendix 11.8.

²¹ The annualized mean return was £5.27 with a standard deviation of £3.01.

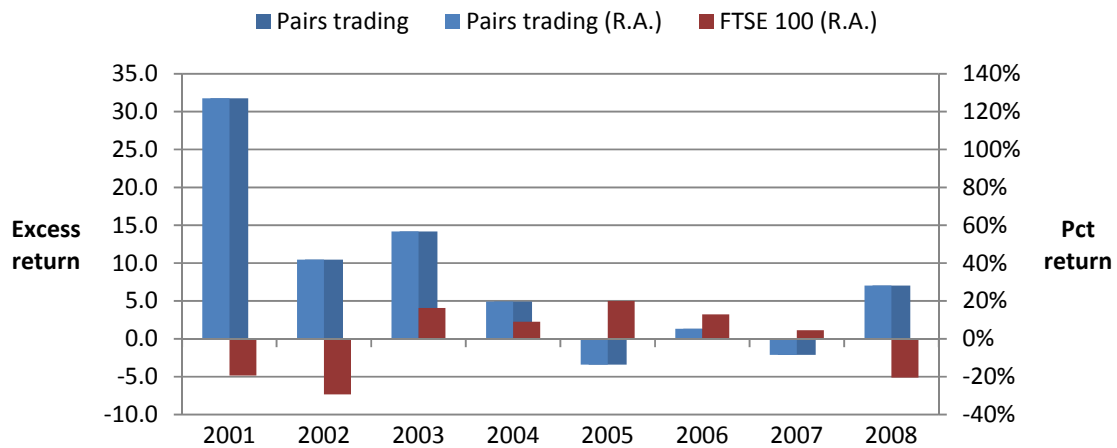


Figure 8-9: Annual return from strategy with TW = 150, EP = 7 and committed capital of £25. R.A. = right axis.

Looking at the two best strategies, we now try to vary the barrier parameter a_i and find the following:

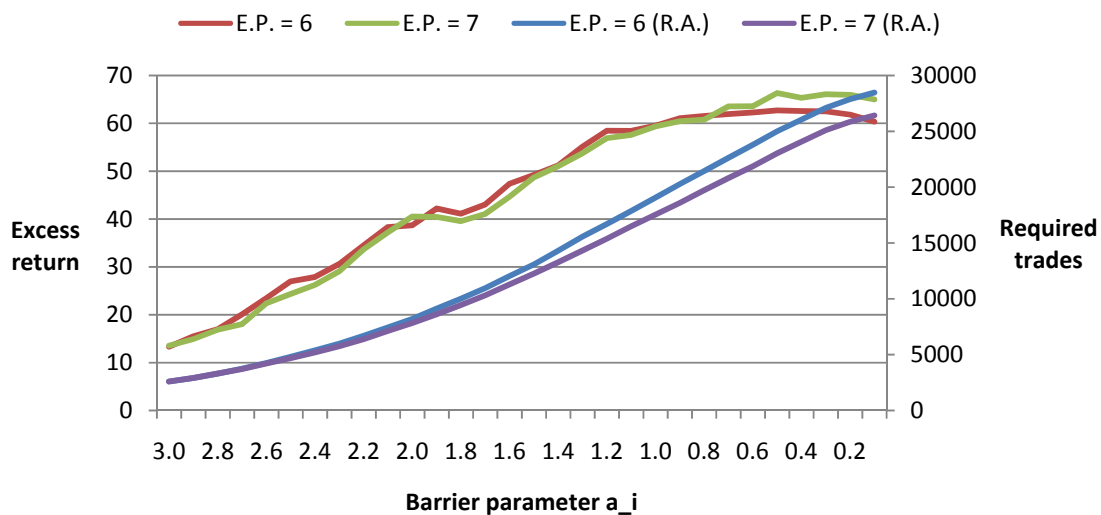


Figure 8-10: The effect of changing the barrier parameter for two different evaluation periods (E.P.). R.A. = right axis.

Disabling the cointegration screening

Figure 11-11 in Appendix 11.8 shows the effect of disabling cointegration screening for the strategy with TW = 150, EP = 7 and $a_i = 2$. The market exposure of the strategy rises as more pairs are traded, and so does excess return. But the annualized information ratio drops from 1.75 to 1.51 and maximum drawdown rises from £3.32 to £5.75, i.e. the tradeoff between risk and return has

deteriorated. Similar results were found for other combinations of window and evaluation period, which is evidence that explicit testing for cointegration improves univariate pairs trading.

Long / short positions and market neutrality

The total excess return of the optimal strategy, TW = 150 and EP = 7, is not evenly divided between long and short positions. Long positions earned £33.53 (82.7%) and short positions £7.00 (17.3%) of the total excess return. The result is common to all pairs trading strategies that were back-tested, that is, long positions consistently outperform short positions for different combinations of TW, EP and a_i . This shows that the specific pairs trading strategy tested here is a good stock-picker, while at the same time hedging its bets profitably. To test the degree of market neutrality alpha and beta with respect to the FTSE 100 index are calculated using the regression

$$R_{P,t} = \alpha + \beta(R_{M,t} - r_t) + \varepsilon_t, \quad t = 1, \dots, n. \quad (8.10)$$

where $R_{P,t}$ is the daily return of the pairs trading strategy, $R_{M,t}$ is the daily return of the FTSE 100 index, and r_t is the daily over-night risk-free rate of interest. The results are shown below

Regression results			
Parameter	Estimate	T-stat	Est. annual.
$\hat{\alpha}$	1.52E-04	2.70	0.0391
$\hat{\beta}$	5.42E-04	6.19	0.1464
$\hat{\rho}(R_{P,t}, R_{M,t})$	0.1394		
	Strategy	Market	
Sharpe ratio ²²	0.9706	0.0348	

Both parameters are statistically significant at the 1% level, so using some degree of beta exposure, the strategy produces annual alpha of 3.91%. Linear correlation with market returns is fairly low at 13.94%, while the risk-adjusted (volatility-adjusted) return of the strategy is much higher than the market.

The returns have skewness of 0.42 and kurtosis of 6.17. The average skewness of the returns of the FTSE 100 stocks in the sample is 0.15 and the kurtosis is 9.76. The equivalent figures for the pairs

²² The Sharpe ratio is here defined as $\sqrt{T_D}E(R_{i,t} - r_t)/\sigma_{i,t}$ where $\sqrt{T_D} = 252$ is the number of trading days per year.

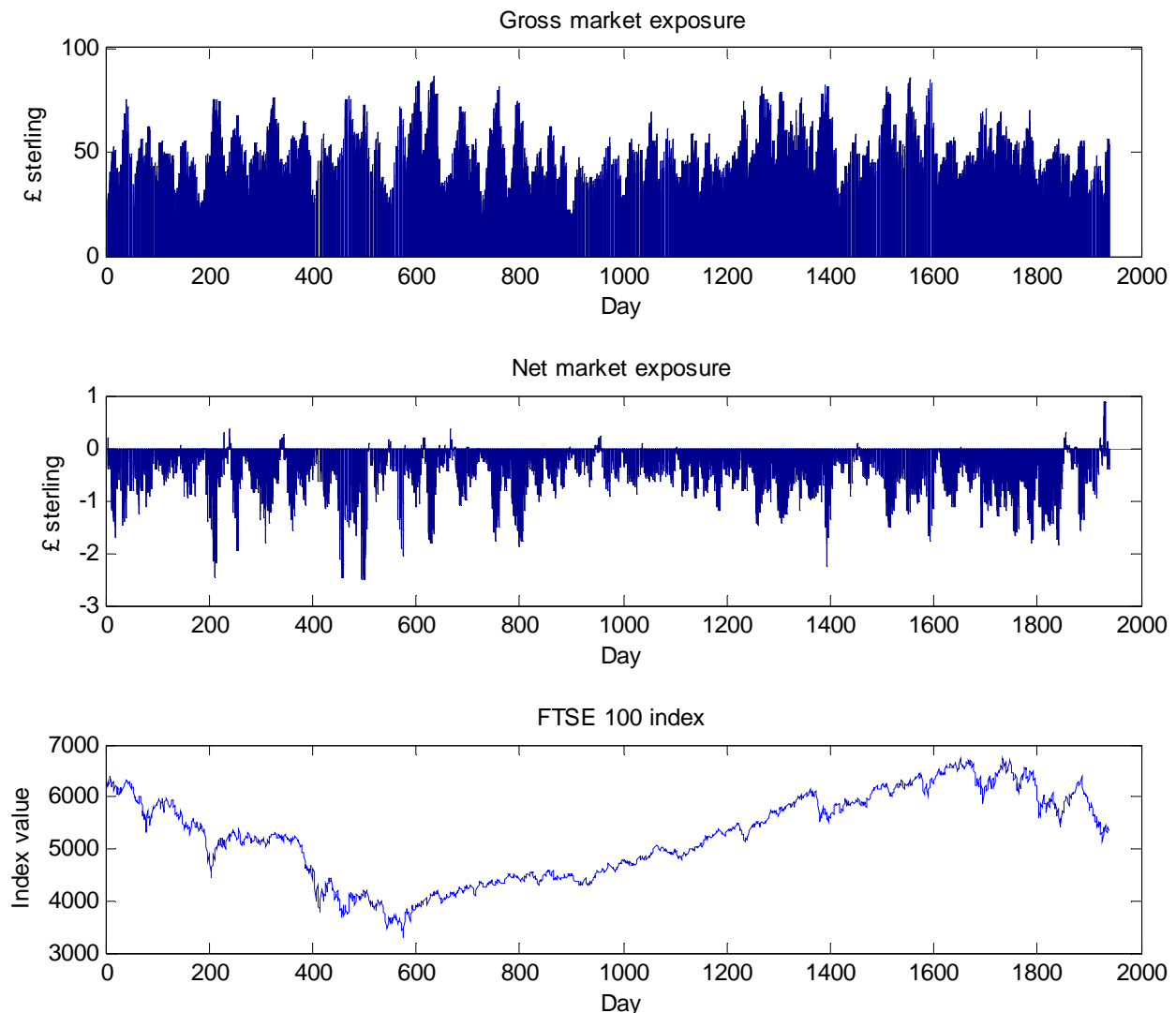
trading strategies using many different combinations of TW and EP are 0.19 and 5.68, which is not far from the individual stocks. The returns from the pairs trading strategy thus have a heavier tail than a normal distribution, and they consist of many slightly negative daily returns and few large positive daily returns.

Net market exposure

The net market exposure of each pair trade is zero on the day the trade is opened, but becomes different from zero as the value of the target stock and its pair changes. Most of the time this market exposure is trivial in magnitude for the aggregate portfolio, but it may occasionally spike due to large price changes in single stocks.

Interestingly, this net exposure is almost always negative (net short) – see Figure 8-11. The most likely explanation is that there is more positive than negative momentum in the stocks in the sample. Say that a hypothetical stock A (with pair B) rises significantly and its price diverges from B by a magnitude large enough to be shorted by the algorithm. If stock A keeps rising faster than stock B during the following days (or weeks) the short position in stock A will grow larger than the corresponding long position in stock B, and the pair will become net short the market.

The tendency of the portfolio as a whole to be net short the market is an indication that there is a positive momentum effect in the stocks in the sample (which is stronger than a possible negative momentum effect). This hypothesis could be tested by applying the EMA momentum strategy to the stocks in the sample and see if (any) profits are generated primarily by long or short positions. It is tempting to suggest that the negative net market exposure is a consequence of all stocks rising in general. But Figure 8-11 shows that there doesn't seem to be any clear connection between general market movements and the net market exposure of the strategy.



**Figure 8-11: Gross and net market exposure of pairs trading strategy with parameters:
TW = 150 days, EP = 5 days, cumulative return = £33.07.**

Improving the univariate pairs trading algorithm

There are several ways to improve the pairs trading algorithm. The algorithm used for this analysis is quite simple, and in practice many enhancements would need to be made. The most important of which are:

- The data set of daily observations used in this analysis does not include dividends. The implication is that every price change in the dataset caused by a dividend payment will result in an incorrect return calculation. For example, if a stock pays a dividend on January 5, the

share price will fall by an amount equal to the dividend per share, compared to the stock price on the previous trading day. Thus a negative return is recorded even though the actual return may have been zero or positive. The extent of this problem is determined by the number of stocks paying dividends and the size of those dividends. The bias in the results will be limited by the fact that the algorithm will sometimes be long and sometimes short dividend paying stocks.²³

- The use of bid-ask quotes and possibly historical limit order books rather than ‘last traded’ quotes. ‘Last traded’ quotes are an approximation of the actual price that may be traded at the end of each day. However, it is not unrealistic that the algorithm will be able to trade during the closing auction and obtain a price close to the ‘last traded’, with a bias towards worse prices due to the bid-ask bounce.
- The inclusion of short-selling costs to the analysis. The assumption of equal transaction costs for long and short positions should be modified to take into account the higher funding costs of short-selling, and possibly the existence of short-selling restrictions on some stocks.
- The use of stop loss orders to avoid holding pairs that show no signs of convergence. Instead of holding each pair until convergence or the passing of three evaluation periods, a stop loss order may be placed at a predetermined price spread. If the given pair keeps diverging instead of converging in price, it is closed down. Although stop loss orders are typically used by practitioners, there is no guarantee that they will improve the algorithm (NATH, P., 2003). How to place them optimally is also a challenge. Nath (2003) takes the approach of opening pair trades at the 25th and 75th percentiles of daily distances (e_{it}^*) and stops at the 5th and 95th percentiles, respectively. Trade convergence is defined by the median of e_{it}^* .

Sub-conclusion

Univariate pairs trading is capable of generating positive excess returns that are uncorrelated with the ‘long-only’ return of the population of stocks used. The strategy is highly sensitive to the parameters used. Changing the length of the ‘training window’, the time between reevaluation of pairs, or the barrier parameter makes a large impact on results. Nevertheless, trying many different combinations

²³ There is a risk, of course, that the algorithm itself is biased towards shorting dividend paying stocks, and going long non-dividend paying stocks.

of parameters an optimal strategy was found. In addition, it was shown that lower values of the barrier parameter a_i yields higher profits. This seems counterintuitive, as one would guess that opening trades at larger deviations from equilibrium values would yield better results. The conclusion is that despite the higher transaction costs, it is profitable to enter many quick trades betting on immediate reversals in price movement.

The minimum distance criterion is successful at finding profitable pairs. Screening each pair for a high cointegration value further ensures the quality of the pairs, and increases the information ratio of the strategy.

It is also found that a portfolio of many open pair trades tends to have a negative net market exposure (if positions aren't rebalanced on a daily basis). This suggests that the stocks in the sample have stronger *positive* return momentum than *negative*.

The performance of pairs trading in this particular data sample shows that after four good years from 2001 to 2004, pairs trading hit a rough patch from 2005 to 2007. Performance seems to have picked up again in 2008, maybe thanks to the higher price volatility created by the financial crisis of 2007-2008.

8.3 Multivariate pairs trading

In this section we will test the multivariate pairs trading methods presented in section 5.3.5 by means of two main approaches. One that makes use of state space methods and one that doesn't. They both extract a signal, M_{it}^* , from the M-pair portfolio of stock i at time t . The size of the M-pair portfolio is arbitrary, and is a parameter of the algorithm. The distance in 'price space' between target stock i and its signal is defined as

$$e_{it}^* = P_{it}^* - M_{it}^* \quad (8.11)$$

Based on this signal the algorithm trades in the same way as the univariate algorithm, so the main difference lies in the calculation of M_{it}^* . Figure 11-12 in Appendix 11.9 depicts the structure of the multivariate back-test routine.

M_{it}^* can be calculated by means of OLS as described in section 5.3.5:

$$M_{it}^* = \hat{\alpha} + \sum_{j=1}^n \hat{w}_j Q_{jt}^* + \epsilon_t \quad (8.12)$$

where ' $\hat{\cdot}$ ' denotes estimated parameters.

Alternatively a state space method may be estimated. The aim is to extract an underlying signal from the M-pair portfolio. We remember the observation and state equations (6.1) from section 6.2:

$$\begin{aligned} y_t &= Z_t \alpha_t + \varepsilon_t & \varepsilon_t &\sim N(0, H_t) & t &= 1, \dots, n \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t & \eta_t &\sim N(0, Q_t) \end{aligned} \quad (8.13)$$

With n stocks in our M-pair portfolio we write the observation equation as

$$\begin{pmatrix} Q_{1,t}^* \\ \vdots \\ Q_{M,t}^* \\ Q_{T,t}^* \end{pmatrix} = \begin{pmatrix} \varphi_{11} & \varphi_{12} \\ \vdots & \vdots \\ \varphi_{n1} & \varphi_{n2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} + \varepsilon_t \quad (8.14)$$

and the state equation as

$$\begin{pmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \end{pmatrix} = \begin{pmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} + \eta_t \quad (8.15)$$

where

$$H_t = \begin{pmatrix} h_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & h_{M+1} \end{pmatrix}, \quad Q_t = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix} \text{ and } R_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

$Q_{1,t}^*, \dots, Q_{M,t}^*$ are the normalized prices of the M stocks in the M-pair portfolio and $Q_{T,t}^*$ is the normalized price of the 'target' stock (T), i.e. the stock we are trading the M-pair portfolio against. The variable of interest is $\alpha_{2,t}$ which is the underlying signal 'driving' the observations in y_t . We will use $\alpha_{2,t}$ as the M_{it}^* for target stock i to calculate e_{it}^* .

In order to estimate the model based on the observed data $Q_{1,t}^*, \dots, Q_{n,t}^*, Q_{T,t}^*$ we must make an initial guess of the various parameters and the initial state mean α_0 and state covariance matrix Q_0 . The guess will be $\varphi_{11}, \dots, \varphi_{n2} = 0.5$; $\delta_1, \delta_2 = 1$; $h_1, \dots, h_n = q_1, q_2 = 2$; $\alpha_0 = 0$ and $Q_0 = I_n$ where I denotes the identity matrix. The parameters are estimated using maximum likelihood as described in section 6.2.

Results

It was not possible to include position ‘memory’ in the multivariate pairs trading routine. Therefore all positions are closed down at the end of each evaluation period. This has a negative impact on performance because each open trade has less time to converge towards equilibrium, i.e. $e_{it}^* = 0$. Initially an M-pair portfolio size of $M = 3$ stocks is used. The stocks are found using the minimum distance criterion described in section 5.3.5. Using the OLS based method of calculating M_{it}^* in equation (8.12) various combinations of TW and EP were tested using barrier parameter $a_i = 1.5$. The results are shown in Figure 11-13 in Appendix 11.9. The optimal strategy uses TW=200 and EP=40. It produces an excess return of £10.0 with an information ratio of 0.57. Transaction costs at £25.41 have become a concern. The EP is significantly longer than the optimal univariate strategy (EP=7), as a higher EP usually means less trading activity and lower transaction costs. Figure 8-12 shows the effect of varying the barrier parameter. A higher a_i results in less trading activity, lower transaction costs and a higher information ratio.

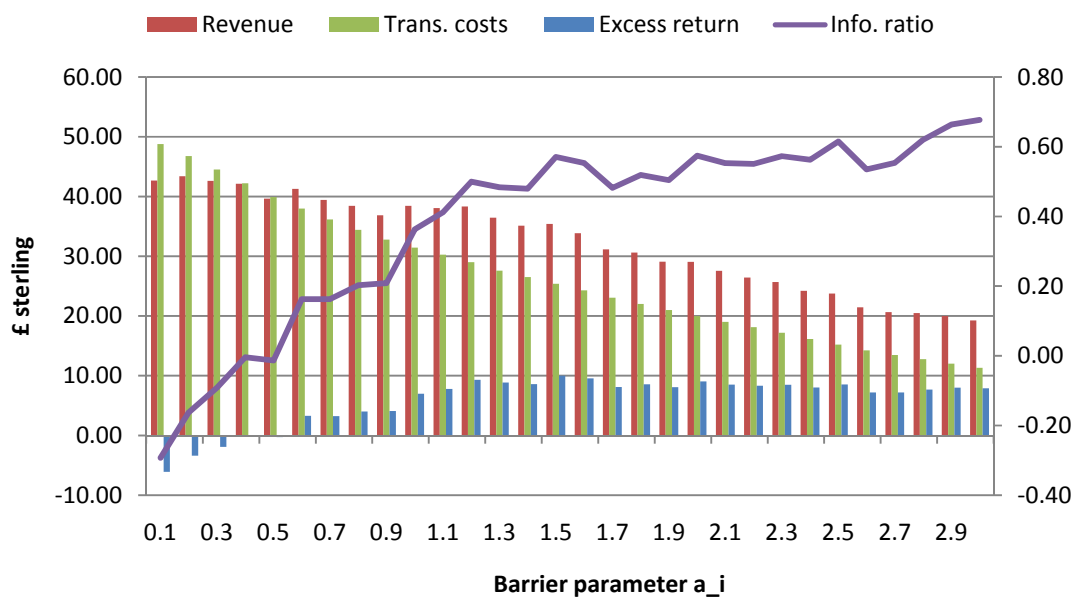


Figure 8-12: Results for various barrier parameter values. TW=200 and EP=40.

In an attempt to enhance the strategy we try to change the way the M stocks in the Q_{jt}^* portfolio are chosen. Instead of using the minimum distance criterion (MD), we use the cointegration-based (CI) method described in section 5.3.5. The results are shown in Figure 8-13.

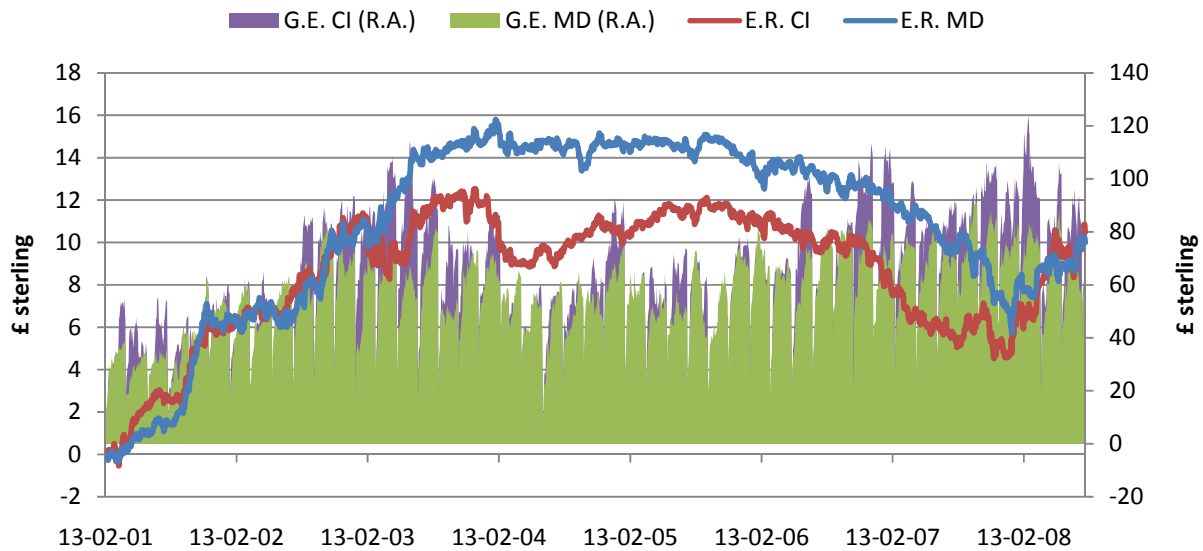


Figure 8-13: Results using two different approaches to create the Q_{jt}^* portfolio.
 G.E. = gross exposure, E.R. = excess return, CI = cointegration based method,
 MD = minimum distance, R.A. = right axis.

Method	CI	MD
Excess return (£)	10.49	10.00
Annual volatility (£)	2.75	2.34
Information ratio	0.51	0.57
Avg. market exposure (£)	64.14	56.47

The CI method takes more risk, but has a worse risk-return tradeoff as measured by the information ratio.

The jagged edges of the market exposure in the above figure reflect that all positions are closed down at the

end of each evaluation period.

We may also attempt to enhance the strategy by changing the number of stocks in the M -pair portfolio. Table 8-1 reports results. In the space of multivariate strategies, $M = 2$ dominates the other possibilities. But the univariate pairs trading strategy performs significantly better than the best multivariate strategy.²⁴

²⁴ The 'memory' feature of the univariate strategy has been disabled to be on an even footing with the multivariate strategy. I.e. positions are not carried into new evaluation periods.

M	1	2	3	4	5	6
Revenue (£)	32.69	31.85	35.41	37.94	36.02	37.68
Trans. costs (£)	12.92	18.26	25.41	32.70	38.74	45.55
Excess return (£)	19.77	13.60	10.00	5.24	-2.72	-7.87
Required trades	6,460	9,128	12,706	16,349	19,371	22,776
Information ratio	0.97	0.77	0.57	0.28	-0.15	-0.41
Max. drawdown (£)	-6.54	-8.14	-10.07	-12.84	-17.65	-20.01

Table 8-1: The effect of varying the number of stocks in the M-pair portfolio.
TW = 200, EP = 40 and $\alpha_i = 1.5$.

State space results		We now test the state space (SS) approach to see if it is an improvement of the above. Using the same parameters as above, namely TW = 200 and EP = 40, the results are not encouraging. The barrier parameter is lowered to $\alpha_i = 1$, but this doesn't help much. Although the strategy has positive revenue over the data sample, transaction costs are too high. The SS strategy trades significantly less than the OLS strategy, and the market exposure is unevenly distributed over time.
Revenue (£)	2.28	
Transaction costs (£)	7.08	
Excess return (£)	-4.81	
Annual volatility (£)	0.96	
Information ratio	-0.67	
Avg. market exposure (£)	17.92	
Required trades	3,542	
SS models estimated	1,026	

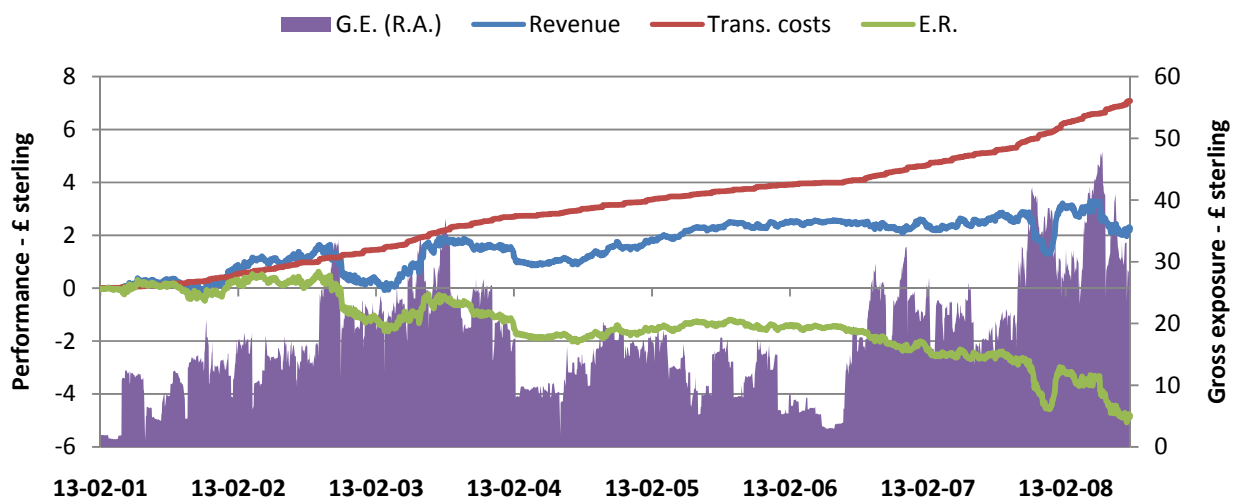


Figure 8-14: Results for state space strategy. TW = 200, EP = 40 and $\alpha_i = 1$.
G.E. = gross exposure, E.R. = excess return, R.A. = right axis

As an example, in the first evaluation period, stock $i = 16$ matched up with stocks $j = \{42, 64, 53\}$.²⁵

The estimated state space system with estimated parameters looks like

$$\begin{pmatrix} Q_{1,t}^* \\ Q_{2,t}^* \\ Q_{3,t}^* \\ Q_{T,t}^* \end{pmatrix} = \begin{pmatrix} -0.1400 & 0.4482 \\ 0.2804 & 0.3482 \\ 0.2674 & 0.2940 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} + \varepsilon_t \quad \begin{pmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \end{pmatrix} = \begin{pmatrix} 0.4712 & 0 \\ 0 & 0.6026 \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{pmatrix} + \eta_t$$

$$H_t = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix}, \quad Q_t = \begin{pmatrix} 1.6587 & 0 \\ 0 & 1.6587 \end{pmatrix}$$

The results are obtained after 2 iterations and 121 function evaluations, and the final likelihood value is 1075.52. Comparing the estimated parameters of the different models, there was wide dispersion in the estimated φ_{nk} . But the error variance h_t and q_t were most of the time close to 1 and 1.66 respectively. A possible reason for the bad earnings performance of the strategy is that the estimated signal $\alpha_{2,t}$ resembles the target stock too closely. This might be because the maximum likelihood function is very jagged and it is difficult for the BFGS algorithm to find the global minimum. In some cases estimation is not possible because individual matrices in the state space recursions come too close to singularity.²⁶

The distance e_{it}^* is very stable over time compared to univariate pairs trading, so there are relatively few trade signals. The average linear correlation between ΔP_{it}^* and ΔM_{it}^* is 94.23% which is a very high value, and considerably higher than the corresponding values for the other approaches to pairs trading.

Sub-conclusion

The multivariate pairs trading strategies tested in this section underperformed the univariate approach significantly. Each additional stock in the Q_{jt}^* portfolio doesn't contribute enough new information to the pairs trade to make up for the higher transaction costs.

Forming pairs using a cointegration-based approach does not improve performance, and neither does the use of state space methods. It is possible that the M-pair formation approach can be improved by calculating the distance e_{it}^* or degree of cointegration using all possible combinations of stocks in the

²⁵ Stock $i = 16$ is British Petroleum Plc which operates in the oil & gas industry. Arranged in order of increasing L^2 norm in normalized price space, stocks $j = \{42, 64, 53\}$ are Lonmin Plc (basic materials), Smith & Nephew Plc (healthcare products) and Reed Elsevier Plc (media).

²⁶ The back-test routine skips M-pairs with this problem by using a matrix reciprocal condition number estimate in the Kalman filter. This is done with the MATLAB function 'rcond'.

M-pair. The combination with the lowest distance, the highest degree of cointegration, or a weighted average of the two is then chosen. Note that for values of $M > 2$ it is a very time consuming procedure.

It is possible that the state space model in (8.13)-(8.14) can be improved by using a different specification. One possibility is to separate the stocks in the M-pair portfolio into different groups depending on their characteristics (either statistical or fundamental, such as industry). The observation equation of such a specification with $M = 4$ could take the form

$$\begin{pmatrix} Q_{1,t}^* \\ Q_{2,t}^* \\ Q_{3,t}^* \\ Q_{4,t}^* \\ Q_{T,t}^* \end{pmatrix} = \begin{pmatrix} \varphi_{11} & 0 & \varphi_{13} \\ \varphi_{21} & 0 & \varphi_{23} \\ 0 & \varphi_{32} & \varphi_{33} \\ 0 & \varphi_{42} & \varphi_{43} \\ 0 & 0 & \varphi_{53} \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{pmatrix} + \varepsilon_t \quad (8.16)$$

with state equation

$$\begin{pmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \end{pmatrix} = \begin{pmatrix} \delta_1 & 0 & 0 \\ 0 & \delta_2 & 0 \\ 0 & 0 & \delta_3 \end{pmatrix} \begin{pmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{pmatrix} + \eta_t \quad (8.17)$$

where

$$H_t = \begin{pmatrix} h_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & h_{M+1} \end{pmatrix}, \quad Q_t = \begin{pmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{pmatrix}.$$

The M_{it}^* signal would in this case be $\alpha_{3,t}$. This would leave open the question of how to find the $Q_{j,t}^*$ stocks $j = \{1,2,3,4\}$.

9 CONCLUSION

The subjects covered in this paper have been chosen to give the reader a broad introduction to market microstructure and trading algorithms. Both subjects are extensive and challenging, and tools from several fields within finance, economics and statistics have been applied. The analysis has touched on both theoretical and empirical aspects of market microstructure and trading algorithms.

The academic literature in market microstructure is rich, and the possibilities for empirical research are many. Section 4.3 showed how a simple microstructure model can be tested empirically using vector autoregression techniques. Several challenges arose due to the high-frequency nature of the empirical data, as the number of observations is very high. Nevertheless, the analysis was able to confirm the findings of Hasbrouck (1991). Quotes are revised in response to trading, and trading is done in response to changes in quotes, giving rise to a two-way dynamic relationship between the two events. While quote revisions were self-correcting, trade events had positive autocorrelation. An important feature of the model is that private (asymmetric) information is revealed as the ‘trade innovation’.

The following section on the optimal execution of portfolio transactions showed how a relatively simple model of intra-day trading can give much intuition on the challenges of trading large amounts of shares in a short period of time. Almgren and Chriss (2001) show that the problem of minimizing implementation shortfall can be solved as a quadratic optimization problem using a quadratic utility function. An important conclusion of the analysis is that the optimal execution strategy is static over time. I.e., due to market efficiency it is not possible to improve the execution process by forecasting prices. Even if it were possible to predict the drift of the price process, or if prices exhibit serial correlation, the benefits of including such information in an execution strategy are too few.

The volume-weighted average price can be used as an execution benchmark. Furthermore, using the analysis of McCulloch and Kazakov (2007) it was shown that if intra-day volume is modeled as a

doubly stochastic binomial point process, mean-variance analysis can be used to find an optimal VWAP trading strategy.

It has been shown that exponential moving averages can be used to capture drift in the price process of tradable instruments, known as price momentum. A comprehensive back-test procedure was carried out by varying the sensitivity of the EMAs and the price sampling frequency. The strategy was capable of generating positive excess returns with an annualized information ratio in excess of 1.

Using a minimum distance criterion to pair stocks from a large population, as suggested by Gatev et al (2006), univariate pairs trading also generated positive excess returns. The majority of the returns were earned in the beginning of the sample period, however, showing that the strategy is unstable over time. The use of explicit testing for cointegration between individual stocks was shown to improve the information ratio significantly.

Multivariate pairs trading was less successful. The higher transaction costs involved in trading one stock against a portfolio of stocks outweighed the benefits. Attempting to improve the procedure by choosing stocks in the M-pair portfolio using a cointegration measure was unsuccessful. The use of state space methods was not successful either, as the trade signal M_{it}^* created by the state space model resembled the target stock P_{it}^* too much. Finally, an alternative specification of the state space model was suggested.

Suggestions for further research

- Much interesting analysis can be carried out using historical tick data, for example, to see how the results from the VAR framework differ between stocks in different markets and over time. Another possibility is to back-test an optimal execution algorithm on historical intra-day data.
- The back-testing of speculative algorithms was carried out using daily 'last traded' price quotes. The realism of the analysis can be greatly improved by using intra-day bid and ask quotes.
- In theory, pairs trading should be possible using high-frequency data in transaction time. State space methods would be particularly suitable to such a strategy, as it is possible to treat non-trading events as missing observations.

10 SOURCES

- ALEXANDER, C. and A. DIMITRIU. 2002. The Cointegration Alpha: Enhanced Index Tracking and Long-Short Equity Market Neutral Strategies. *ISMA Centre, University of Reading, UK*.
- ALMGREN, R. 2008. Execution Costs. *Encyclopedia of Quantitative Finance*.
- ALMGREN, R. and N. CHRISS. 2001. Optimal Execution of Portfolio Transactions. *Journal of Risk*. **3**, pp.5-39.
- ALMGREN, R., C. THUM, E. HAUPTMANN, and H. LI. 2005. Equity Market Impact. *RISK*. **18**, pp.57-62.
- BOSTRÖM, D. 2007. *Market Neutral Strategies*. Kaupthing Bank. Stockholm.
- BOX, G., G. JENKINS, and G. REINSEL. 1994. *Time Series Analysis, Forecasting and Control 3rd edition*. San Francisco: Holden-Day.
- CAMPBELL, J. Y., A. W. LO, and A. C. MACKINLAY. 1997. *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University Press.
- DB. 2008. *Deutsche Bank autobahn*. [online]. Available from World Wide Web: <<http://www.autobahn.db.com/>>
- DURBIN, J. and S. KOOPMAN. 2001. *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.
- ENGLE, R. F. and C. GRANGER. 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*. **55**(2), pp.251-276.
- ENGLE, R. and J. RUSSELL. 1998. The Autoregressive Conditional Duration Model. *Econometrica*. **66**, pp.1127-1163.
- FAMA, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*. **25**, pp.383-417.
- FAMA, E. F. 1991. Efficient Capital Markets: II. *The Journal of Finance*. **46**(5), pp.1575-1617.
- GATEV, E., W. N. GOETZMANN, and K. G. ROUWENHORST. 2006. Pairs Trading: Performance of a Relative Value Arbitrage Rule. *Yale ICF Working Paper No. 08-03*.
- GUJARATI, D. N. 2003. *Basic Econometrics (Fourth Edition)*. New York: McGraw-Hill/Irvin.

- HASBROUCK, J. 2007. *Empirical Market Microstructure*. New York: Oxford University Press.
- HASBROUCK, J. 1991. Measuring the Information Content of Stock Trades. *Journal of Finance*. **XLVI**(1), pp.179-207.
- HULL, J. C. 2006. *Options, Futures and Other Derivatives 6th Edition*. Upper Saddle River, New Jersey: Prentice Hall.
- INGERSOLL, J. 1987. *Theory of Financial Decision-Making*. New Jersey: Rowman and Littlefield.
- KASSAM, A. 2008. *The MathWorks - Recorded Webinar: Algorithmic Trading with MATLAB for Financial Applications*. [online]. [Accessed 25 September 2008]. Available from World Wide Web: <<http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=18854&objectType=file>>
- KIM, K. 2007. *Electronic and Algorithmic Trading Technology*. Burlington, MA: Elsevier.
- KONISHI, H. 2002. Optimal slice of a VWAP trade. *Journal of Financial Markets*. **5**, pp.197-221.
- LUX, H. 2000. *Interview with James Simons, founder of Renaissance Technologies Corp.* [online]. [Accessed 28 June 2008]. Available from World Wide Web: <<http://chinese-school.netfirms.com/abacus-hedge-funds-Jim-Simons.html>>
- MCCULLOCH, J. 2007. Relative Volume as a Doubly Stochastic Binomial Point Process. *Quantitative Finance*. **7**, pp.55-62.
- MCCULLOCH, J. and V. KAZAKOV. 2007. Optimal VWAP Trading Strategy and Relative Volume. *Quantitative Finance Research Centre, Research Paper Series 201*. University of Technology, Sydney.
- NATH, P. 2003. High Frequency Pairs Trading with U.S. Treasury Securities. *London Business School*.
- O'HARA, M. 1997. *Market Microstructure Theory*. Malden, MA: Blackwell Publishing.
- PERLIN, M. S. 2007a. Evaluation of a Pairs Trading Strategy at the Brazilian Financial Markets. *ICMA/Reading University Working Paper*.
- PERLIN, M. S. 2007b. M of a kind: A Multivariate Approach at Pairs Trading. *ICMA/Reading University Working Paper*.

SHUMWAY, R. H. and D. S. STOFFER. 2006. *Time Series Analysis and Its Applications (Second Edition)*. New York: Springer.

TEITELBAUM, R. 2007. *Simons at Renaissance Cracks Code, Doubling Assets (Update1)*. [online]. [Accessed 13 June 2008]. Available from World Wide Web:
<<http://www.bloomberg.com/apps/news?pid=20601213&sid=aq33M3X795vQ&refer=home>>

TRAGSTRUP, L. 2008. Interview with the author on 30 December 2008. Copenhagen.

TSAY, R. S. 2005. *Analysis of Financial Time Series*. Hoboken, New Jersey: John Wiley & Sons.

TURQUOISE. 2008. *Turquoise Homepage*. [online]. Available from World Wide Web:
<<http://www.tradeturquoise.com/>>

WELCH, G. and G. BISHOP. 2006. An Introduction to the Kalman Filter. *Department of Computer Science - University of North Carolina at Chapel Hill*.

WIKIPEDIA. 2008c. *Erlang (programming language)*. [online]. Available from World Wide Web:
<[http://en.wikipedia.org/wiki/Erlang_\(programming_language\)](http://en.wikipedia.org/wiki/Erlang_(programming_language))>

WIKIPEDIA. 2008a. *Financial Information eXchange*. [online]. Available from World Wide Web:
<http://en.wikipedia.org/wiki/Fix_protocol>

WIKIPEDIA. 2008b. *Markets In Financial Instruments Directive*. [online]. Available from World Wide Web: <<http://en.wikipedia.org/wiki/MiFID>>

11 APPENDIX

11.1 Event types in the FTSE 100 tick data

Supplement to section 7.1:

The data set is obtained from Reuters Datascope and the following information is from the Reuters Datascope documentation.

Each line in the dataset will be called an event. There may be more than one event at a given point in time. Some events in the data set have one or more 'qualifiers' that classify the event. There are five types of qualifier: Generic, price, secondary price, irregular trade and exchange specific. The qualifiers are coupled with a 'TAG' that gives specific information about the qualifier. For example, the most typical qualifier is A[LSE], where 'A[]' is the qualifier and 'LSE' is its tag.

The two main event types of interest are trades and quotes, as they are the only events used for calculations. To prepare the data set for calculations it must be 'cleaned' from unwanted events. The qualifiers were used to identify those events, and make the process easy (possible). The main events of interest are trade events. Their qualifiers are:

- A[]: Automatic trade generated by the system through automatic execution.
- O[]: Non protected portfolio. As opposed to protected portfolio events that are not reported.
- E[]: Error correction.
- L[]: Late reported trade.
- N[]: Overnight execution.
- n[]: SEATS non risk trade
- Others

The vast majority of trades are automatic (A[]). Non protected portfolio trades (O[]) are often traded at price levels far from the most recent quote midpoint at the time they are reported. Therefore they were excluded from the calculations if the traded price was more than 1% from the previous quote.

Other important events are:

- OB[]: Intra-day Opening order book price derived from order driven trade.
- CLS[]: Market closed

- AUC[]: Auction call periods. They announce the beginning of opening auctions in the morning and closing auctions in the afternoon.

Automatically generated generic events are (tagged with 'GEN'):

- Other OB[] order book events such as 'High' and 'Low' intra-day price.
- OB VWAP[]: The order book volume weighted average price. This event is reported every time a trade is executed from the consolidated order book.

11.2 Volume and open interest of Bund and Euribor futures

Supplement to section 5.3.1:

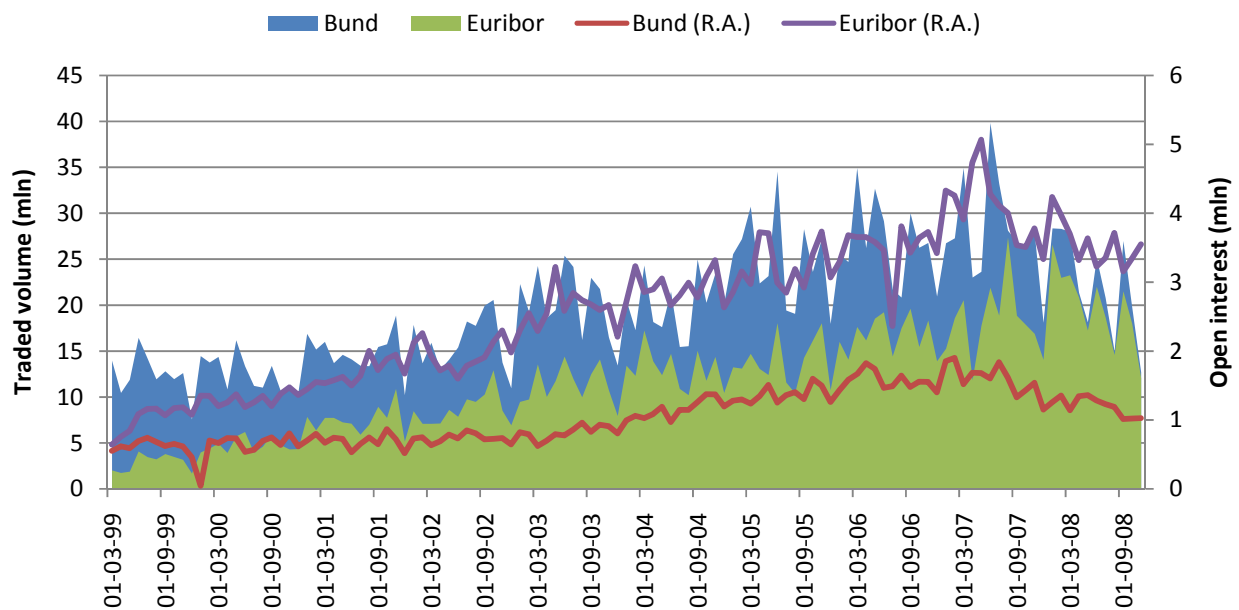


Figure 11-1: Monthly traded volume and open interest of Bund and Euribor futures on Eurex.
R.A. = right axis. Source: Bloomberg.

11.3 Additional co-integration tests

Supplement to section 5.3.3:

The Durbin-Watson test for cointegration uses the d -statistic obtained from the regression

$$y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t \quad (11.1)$$

The d -statistic is defined as $d = \sum_{t=1}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2 / \sum_{t=1}^{t=n} (\hat{u}_t^2)$. The null hypothesis is that $d = 0$ which implies that \hat{u}_t has a unit root and that y_t and x_t are *not* cointegrated. This is because $d \approx 2(1 - \hat{\rho})$ and $\hat{\rho}$ is defined as $\hat{\rho} = 1 + \delta$ in the regression $\Delta y_t = \delta y_{t-1} + u_t$. So if $\hat{\rho}$ is about 1 then d is about 0. The critical values for the d -statistic are 0.511, 0.386 and 0.322 at the 1, 5 and 10% level respectively. If the observed d -statistic is below 0.511 it may then be concluded at the 1% level that $d = 0$, \hat{u}_t has a unit root, and that y_t and x_t are *not* cointegrated (GUJARATI, D. N., 2003).

11.4 Derivation of Kalman filter recursions and illustration

Supplement to section 6.4:

The following is adapted from Durbin & Koopman (2001):

The derivation requires only some elementary properties of multivariate regression theory:

Suppose that x , y and z are random vectors of arbitrary orders that are jointly normally distributed with means μ_p and covariance matrices $\Sigma_{pq} = E[(p - \mu_p)(q - \mu_q)']$ for $p, q = x, y$ and z with $\mu_z = 0$ and $\Sigma_{yz} = 0$. The symbols x, y, z, p and q are employed in this lemma for convenience and have no relation to other parts of the paper.

Lemma 11.1

$$\begin{aligned} E(x|y, z) &= E(x|y) + \Sigma_{xz}\Sigma_{zz}^{-1}z \\ \text{Var}(x|y, z) &= \text{Var}(x|y) - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma'_{xz} \end{aligned}$$

Proof: See Durbin & Koopman (2001).

The linear Gaussian state space model is written as

$$\begin{aligned} y_t &= Z_t\alpha_t + \epsilon_t & \epsilon_t &\sim N(0, H_t) \\ \alpha_{t+1} &= T_t\alpha_t + R_t\eta_t & \eta_t &\sim N(0, Q_t) \quad t = 1, \dots, n. \\ & & \alpha_t &\sim N(a_1, P_1) \end{aligned} \tag{11.2}$$

Let Y_{t-1} denote the set of past observations y_1, \dots, y_{t-1} . The objective is to obtain the conditional distribution of α_{t+1} given Y_t for $t = 1, \dots, n$. Because all distributions are normal, starting at $t = 1$ and building up the distributions of α_t and Y_t recursively, it is easy to show that

$p(y_t|\alpha_1, \dots, \alpha_t, Y_{t-1}) = p(y_t|\alpha_t)$ and $p(\alpha_{t+1}|\alpha_1, \dots, \alpha_t, Y_t) = p(\alpha_{t+1}|\alpha_t)$. Here we derive the Kalman filter in (11.2) for the case where the initial state α_1 is $N(a_1, P_1)$ where a_1 and P_1 are known.

Our objective is to obtain the conditional distribution of α_{t+1} given Y_t for $t = 1, \dots, n$ where $Y_t = \{y_1, \dots, y_n\}$.

Since $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$, we have

$$\begin{aligned} \alpha_{t+1} &= E(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_tE(\alpha_t|Y_t) \end{aligned} \tag{11.3}$$

$$\begin{aligned}
P_{t+1} &= \text{Var}(T_t \alpha_t + R_t \eta_t | Y_t) \\
&= T_t \text{Var}(\alpha_t | Y_t) T_t' + R_t Q_t R_t'
\end{aligned} \tag{11.4}$$

for $t = 1, \dots, n$. Let $v_t = y_t - E(y_t | Y_{t-1}) = y_t - E(Z_t \alpha_t + \varepsilon_t | Y_{t-1}) = y_t - Z_t a_t$.

Then v_t is the one step forecast error of y_t given Y_{t-1} . When Y_{t-1} and v_t are fixed then Y_t is fixed and vice versa. Thus $E(\alpha_t | Y_t) = E(\alpha_t | Y_{t-1}, v_t)$. Using Lemma (11.1) we have

$$\begin{aligned}
E(\alpha_t | Y_t) &= E(\alpha_t | Y_{t-1}, v_t) \\
&= E(\alpha_t | Y_t) + \text{Cov}(\alpha_t, v_t) [\text{Var}(v_t)]^{-1} v_t \\
&= a_t + M_t F_t^{-1}
\end{aligned} \tag{11.5}$$

where $M_t = \text{Cov}(\alpha_t, v_t)$, $F_t = \text{Var}(v_t)$ and $E(\alpha_t | Y_{t-1}) = a_t$ by definition of a_t . Here,

$$\begin{aligned}
M_t &= \text{Cov}(\alpha_t, v_t) = E[E\{\alpha_t(Z_t \alpha_t + \varepsilon_t - Z_t a_t)' | Y_{t-1}\}] \\
&= E[E\{\alpha_t(\alpha_t - a_t)' Z_t' | Y_{t-1}\}] = P_t Z_t', \text{ and} \\
F_t &= \text{Var}(Z_t \alpha_t + \varepsilon_t - Z_t a_t) = Z_t P_t Z_t' + H_t.
\end{aligned}$$

We assume that F_t is nonsingular. Substituting in (11.3) and (11.4) we get

$$\begin{aligned}
a_{t+1} &= T_t a_t + T_t M_t F_t^{-1} v_t = T_t a_t + K_t v_t, \quad t = 1, \dots, n, \\
&\text{with } K_t = T_t M_t F_t^{-1} = T_t P_t Z_t' F_t^{-1}.
\end{aligned} \tag{11.6}$$

We observe that a_{t+1} has been obtained as a linear function of the previous value a_t and v_t , the forecast error of y_t given Y_{t-1} . Using Lemma (11.1) again we have

$$\begin{aligned}
\text{Var}(\alpha_t | Y_t) &= \text{Var}(\alpha_t | Y_{t-1}, v_t) \\
&= \text{Var}(\alpha_t | Y_{t-1}) - \text{Cov}(\alpha_t, v_t) [\text{Var}(v_t)]^{-1} \text{Cov}(\alpha_t, v_t)' \\
&= P_t - M_t F_t^{-1} M_t' = P_t - P_t Z_t' F_t^{-1} Z_t P_t.
\end{aligned}$$

Substituting in (11.4) gives

$$P_{t+1} = T_t P_t L_t' + R_t Q_t R_t', \quad t = 1, \dots, n, \tag{11.7}$$

with $L_t = T_t - K_t Z_t'$. The recursions (11.6) and (11.7) constitute the Kalman filter for model (11.2) (DURBIN, J. and Koopman, S., 2001).

Illustration of the Kalman filter recursion process

Based on their initial estimates the state estimate and error covariance a_t and P_t are first projected into the future, and then corrected using the 'Kalman gain'. A '-' superscript denotes an estimate that has been projected, but not yet corrected.

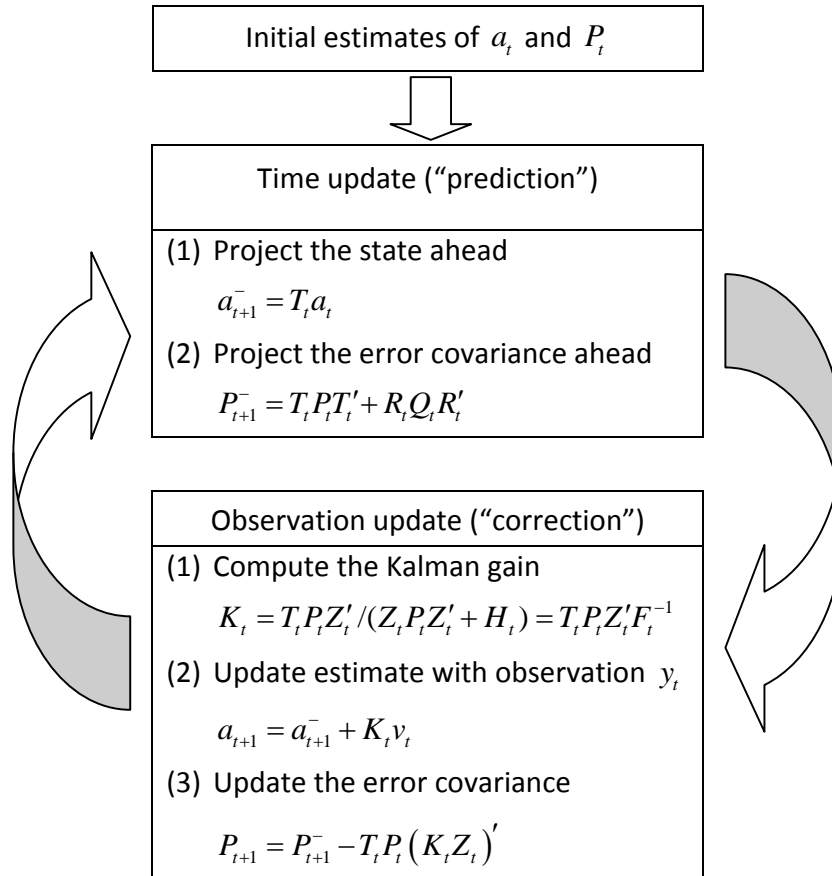


Figure 11-2: The Kalman filter recursion.
(WELCH, G. and Bishop, G., 2006)

11.5 Empirical analysis of tick data

Supplement to section 7.3

	Lag	Coeff.	T-stat		Lag	Coeff.	T-stat
<i>a</i>	0	0.0004	0.8	<i>c</i>	0	-0.0009	-2.1
<i>a</i>	1	-0.0881	-90.3	<i>c</i>	1	-0.1016	-139.2
<i>a</i>	2	-0.0359	-36.3	<i>c</i>	2	-0.0171	-23.1
<i>a</i>	3	-0.0325	-32.8	<i>c</i>	3	-0.0083	-11.3
<i>a</i>	4	-0.0040	-4.0	<i>c</i>	4	-0.0065	-8.7
<i>a</i>	5	-0.0174	-17.5	<i>c</i>	5	-0.0034	-4.6
<i>a</i>	6	0.0065	6.6	<i>c</i>	6	-0.0022	-3.0
<i>a</i>	7	-0.0041	-4.1	<i>c</i>	7	-0.0009	-1.1
<i>a</i>	8	0.0042	4.2	<i>c</i>	8	-0.0010	-1.3
<i>a</i>	9	-0.0102	-10.3	<i>c</i>	9	-0.0006	-0.8
<i>a</i>	10	0.0041	4.3	<i>c</i>	10	0.0016	2.2
<i>b</i>	1	0.1943	148.9	<i>d</i>	1	0.3155	323.2
<i>b</i>	2	0.0476	34.3	<i>d</i>	2	0.0859	82.8
<i>b</i>	3	0.0186	13.3	<i>d</i>	3	0.0348	33.5
<i>b</i>	4	0.0124	8.9	<i>d</i>	4	0.0211	20.3
<i>b</i>	5	0.0064	4.6	<i>d</i>	5	0.0194	18.7
<i>b</i>	6	0.0125	9.0	<i>d</i>	6	0.0139	13.3
<i>b</i>	7	0.0049	3.5	<i>d</i>	7	0.0127	12.2
<i>b</i>	8	0.0006	0.5	<i>d</i>	8	0.0102	9.8
<i>b</i>	9	0.0046	3.3	<i>d</i>	9	0.0127	12.2
<i>b</i>	10	0.0100	7.5	<i>d</i>	10	0.0163	16.5
R ²		0.0428				0.1588	

Table 11-1: Output from VAR analysis of Anglo American PLC data of March 2008.

11.6 Derivation of the exponential moving average

Supplement to section 8.1:

To show that

$$\tilde{x}_t = \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j+1} + w_t$$

has the recursive representation $\tilde{x}_t = (1-\lambda)\tilde{x}_{t-1} + \lambda x_t$ for $\tilde{x}_0 = 0$ and $0 < \lambda < 1$ we write

$$\begin{aligned}\tilde{x}_t - \tilde{x}_{t-1} &= \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j+1} - \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j} \\ &= \lambda x_t (1-\lambda)^0 + \sum_{j=2}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j+1} - \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j} \\ &= \lambda x_t + (1-\lambda) \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j} - \sum_{j=1}^{\infty} \lambda(1-\lambda)^{j-1} x_{t-j} \\ \text{So } \tilde{x}_t &= (1-\lambda)\tilde{x}_{t-1} + \lambda x_t.\end{aligned}$$

11.7 EMA momentum strategies

Supplement to section 8.1:

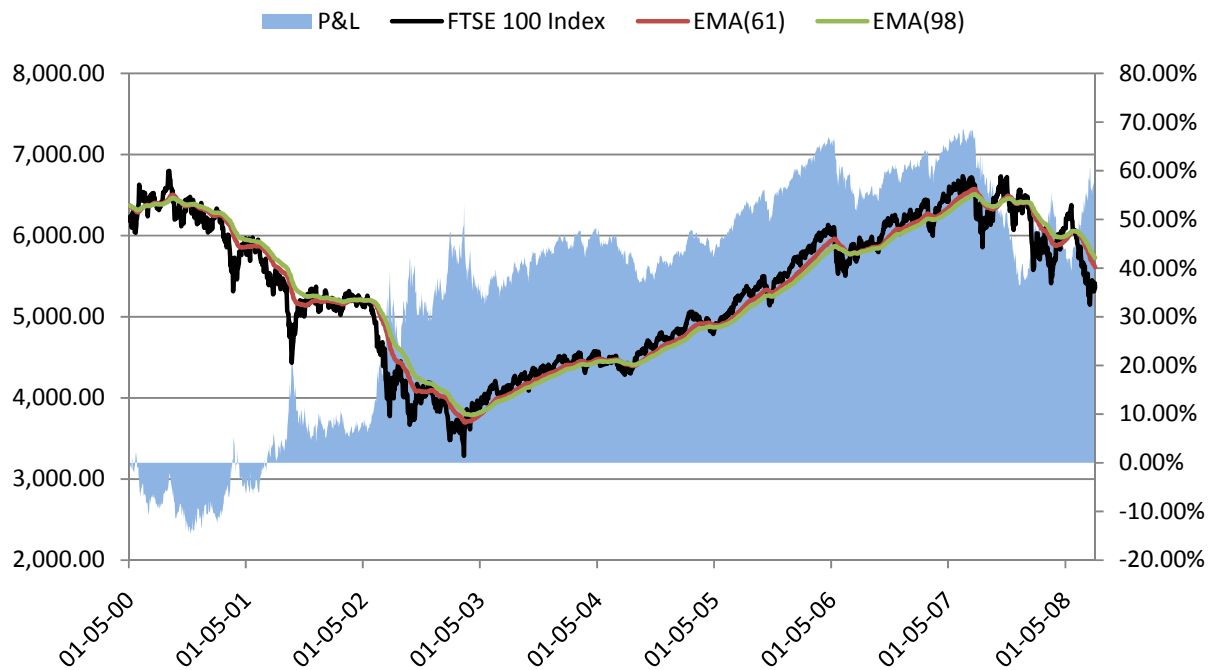


Figure 11-3: The FTSE 100 Index with an EMA(61), EMA(98), and the cumulative return of the strategy superimposed.

11.8 Univariate pairs trading

Supplement to section 8.2:

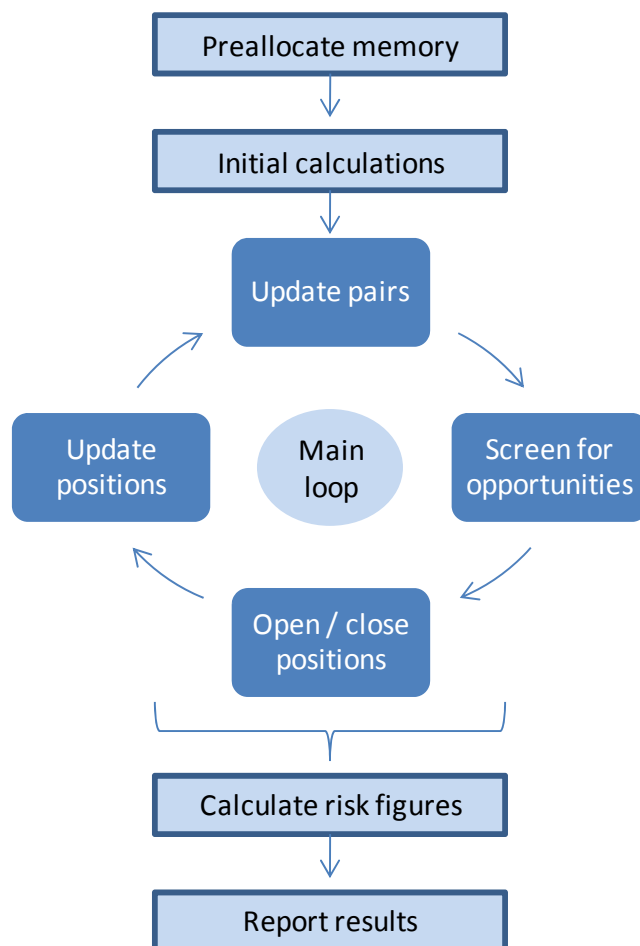


Figure 11-4: Univariate pairs trading back-test routine.

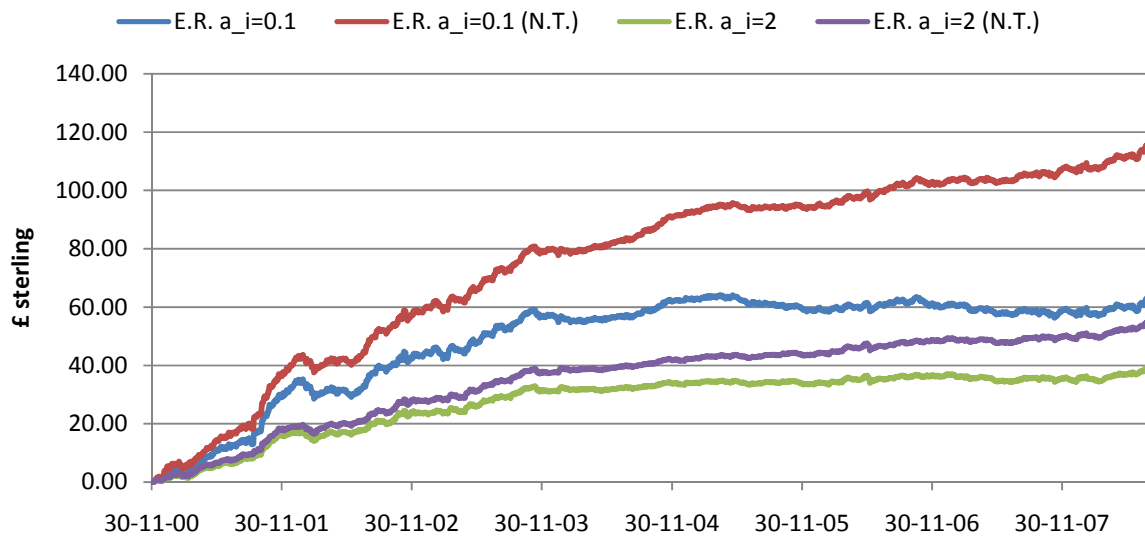


Figure 11-5: Excess return and excess return without transaction costs (N.T.) of pairs trading strategies with TW = 150, EP = 7, and $a_i = 0.1$ or 2.

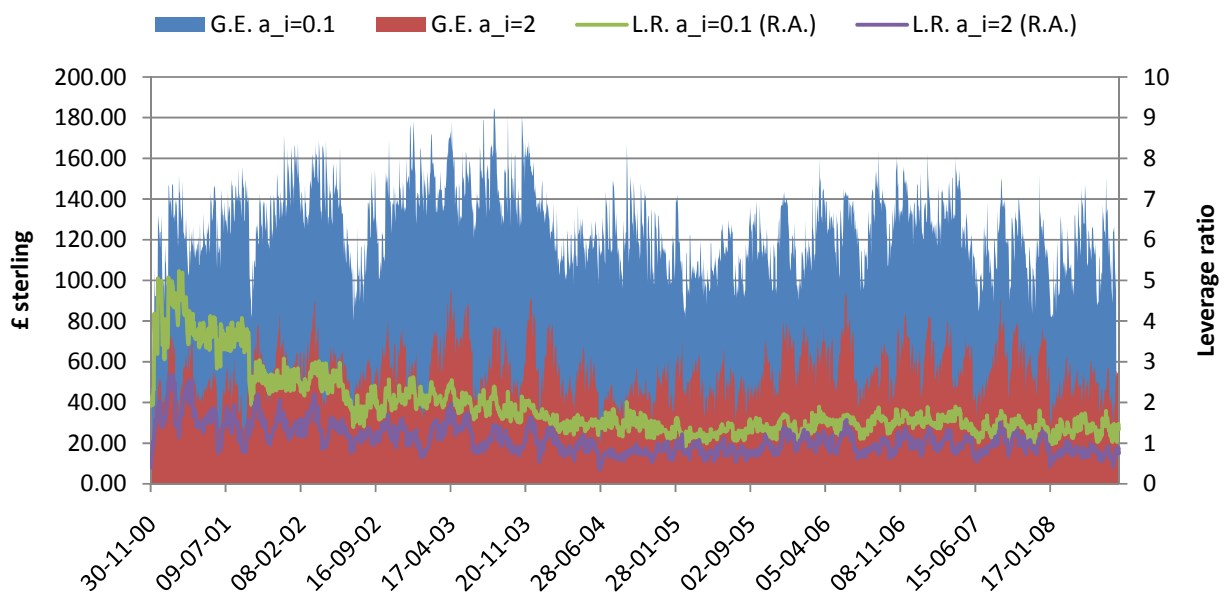


Figure 11-6: Gross market exposure (G.E.) and leverage ratio (L.R.) using committed capital of the same two strategies, $a_i = 0.1$ or 2.

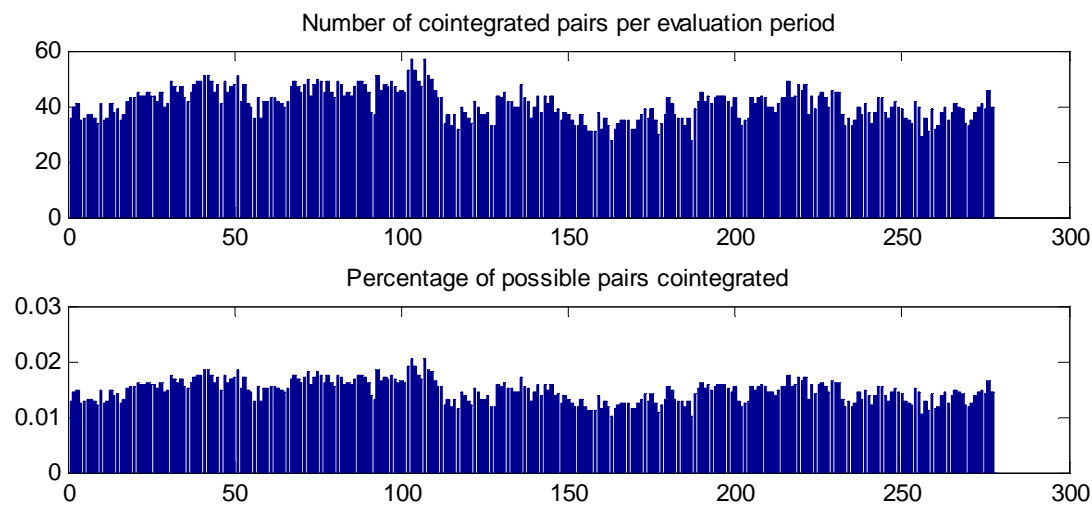


Figure 11-7: The stock population is large enough to create a sufficient number of cointegrated pairs in each evaluation period (TW = 150, EP = 7).

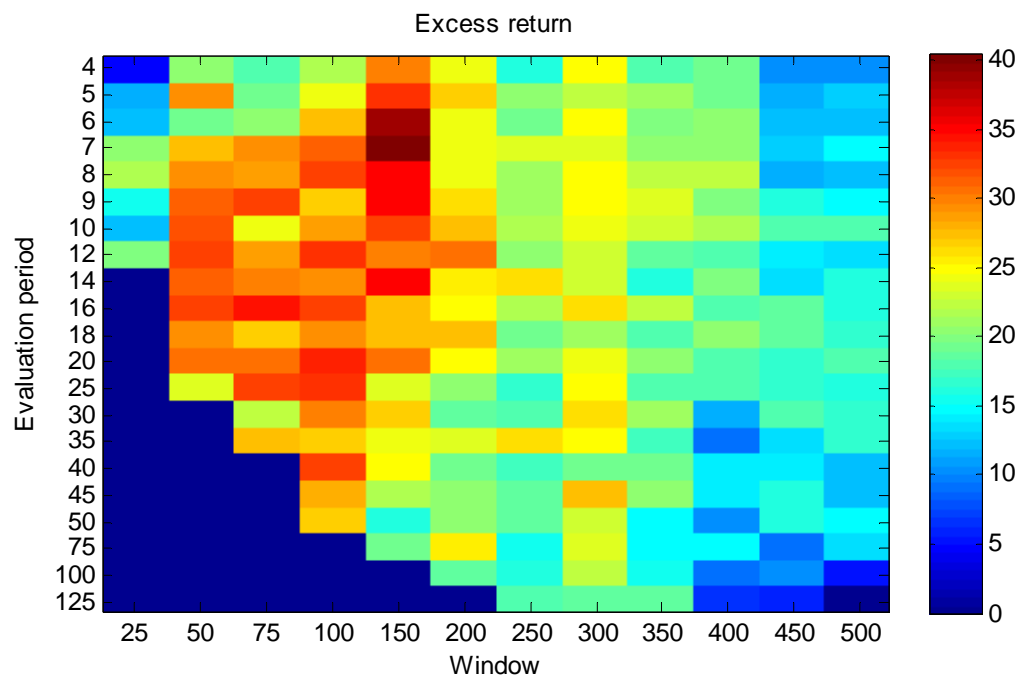


Figure 11-8: Heat map of excess returns for various combinations of window and evaluation period lengths.

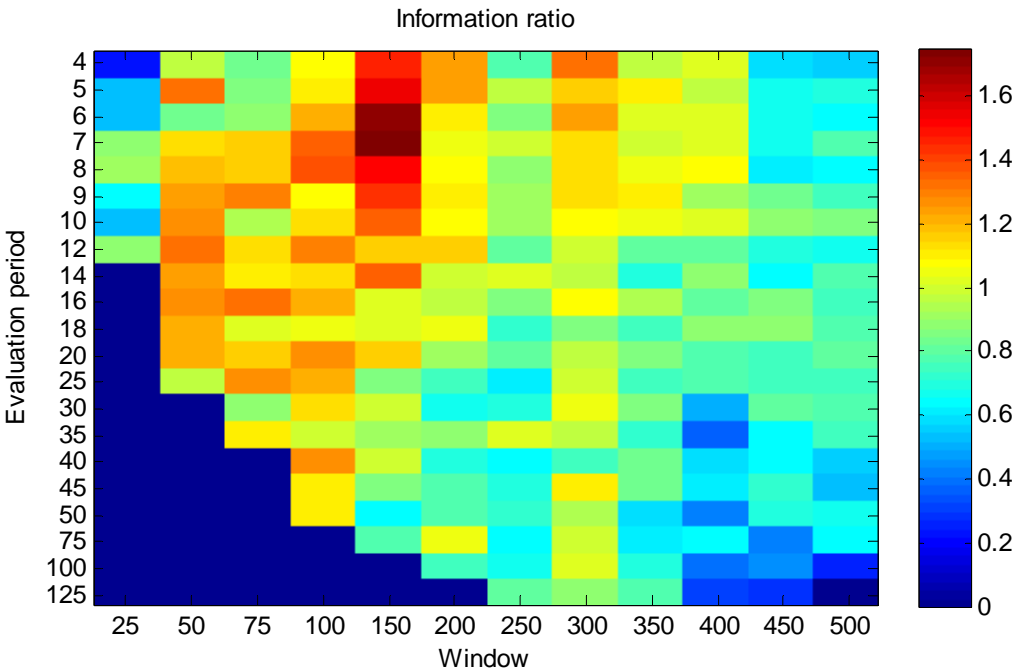


Figure 11-9: Heat map of annualized information ratios.

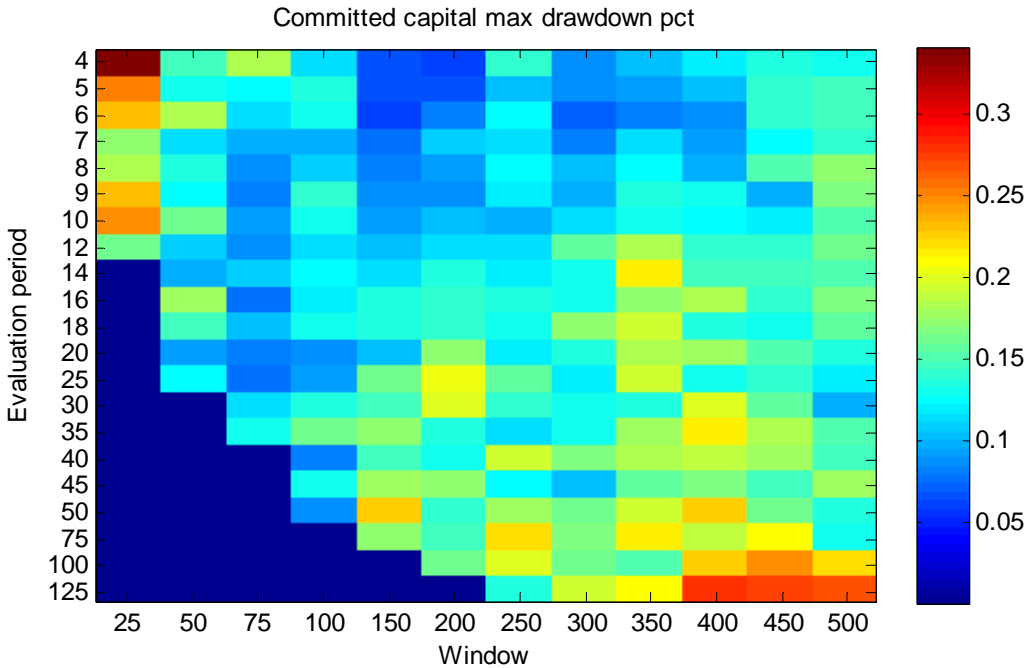


Figure 11-10: Heat map of maximum drawdown of committed capital.
The data is shown in absolute values.

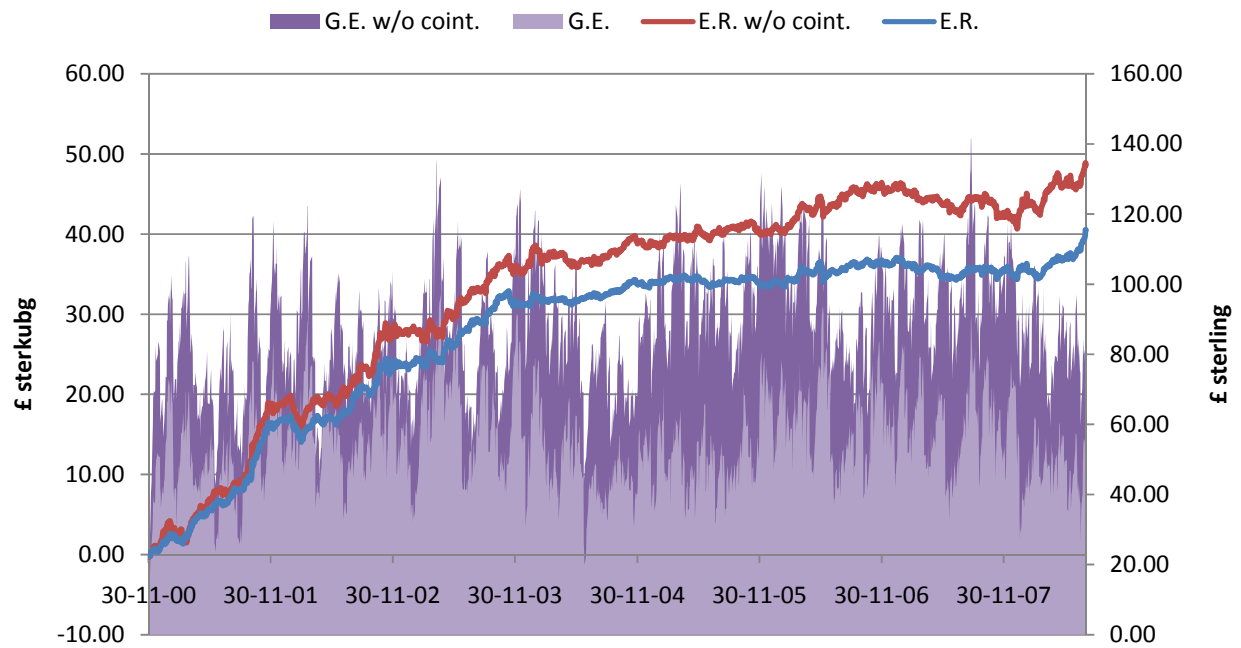


Figure 11-11: Gross exposure (G.E.) and excess return (E.R.) with or without cointegration screening.

11.9 Multivariate pairs trading

Supplement to section 8.3:

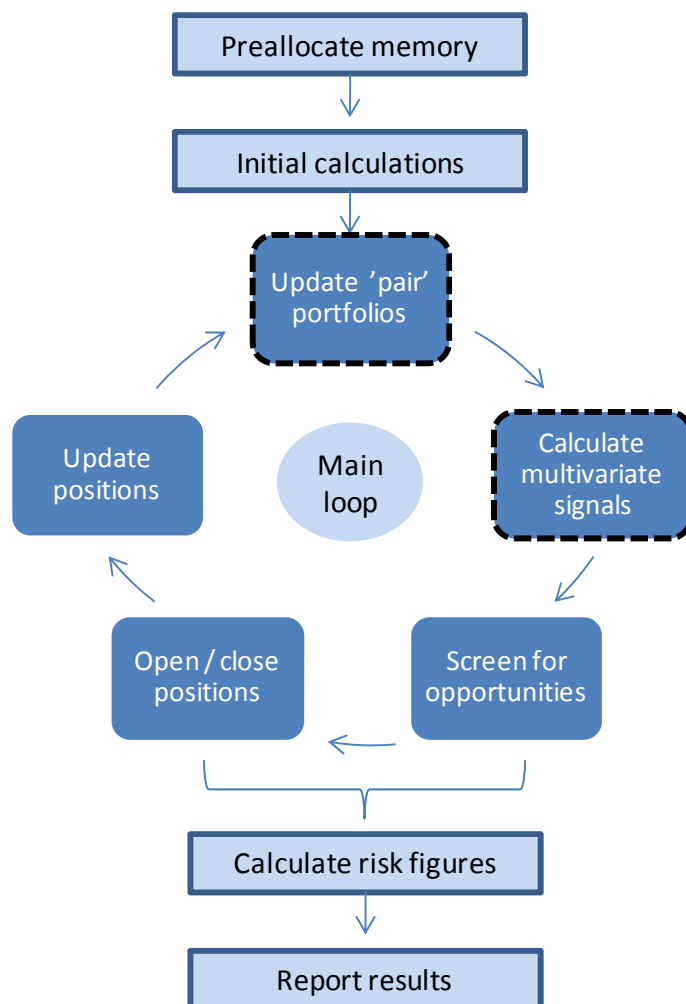


Figure 11-12: Multivariate pairs trading back-test routine.

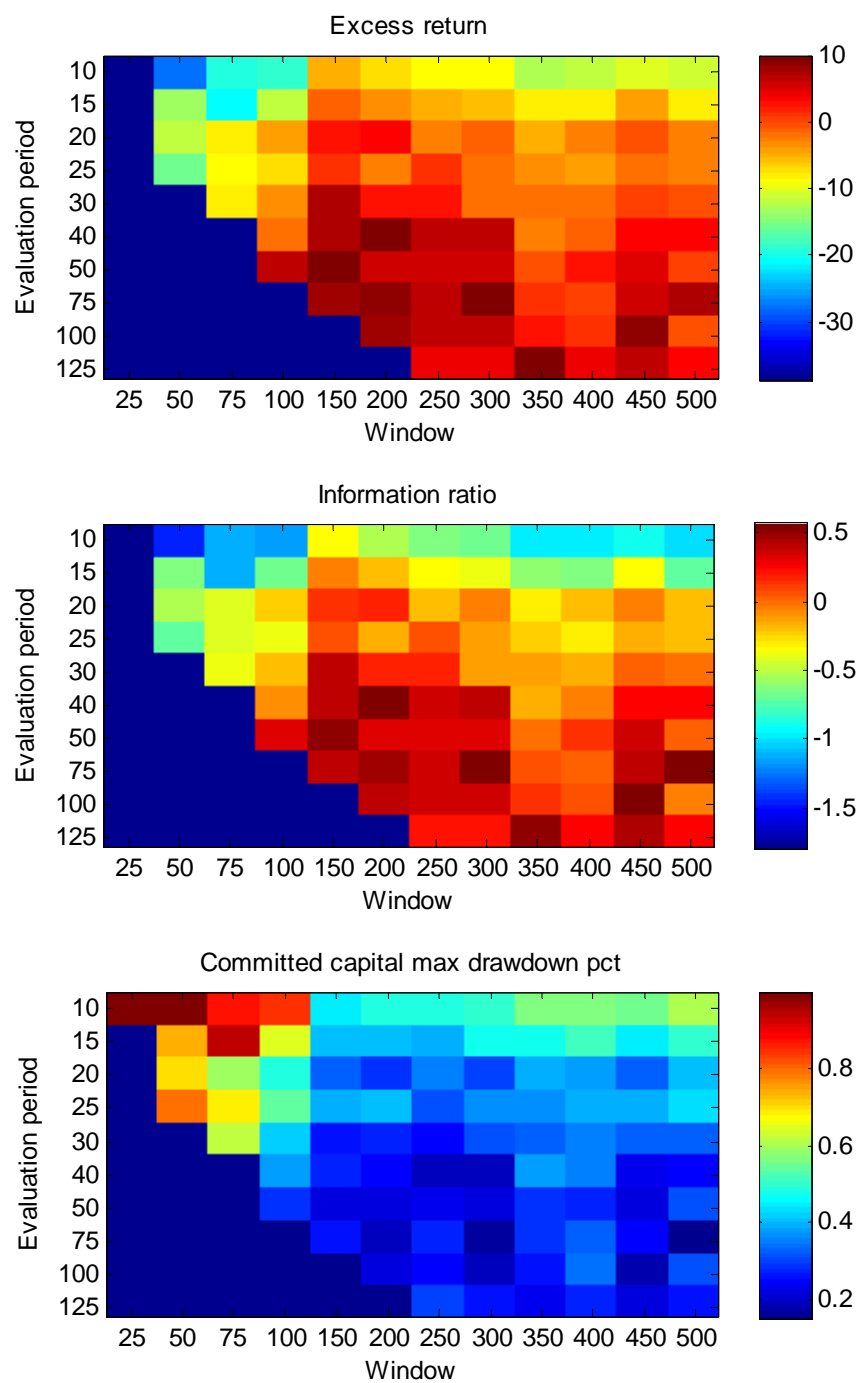


Figure 11-13: Back test using OLS method to find M_{it}^* .