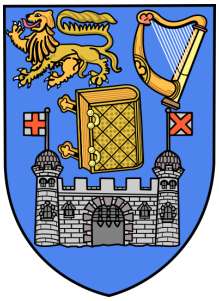
Sentiment Analysis and Time Series Analytics An Analysis of the Banking Crisis in the UK and Ireland



Damian Griffith Bourke

School of Computer Science and Statistics

University of Dublin, Trinity College

A thesis submitted for the degree of

MSc Computer Science

May, 2015

I, Damian Griffith Bourke, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

_____ May 21, 2015

Damian Griffith Bourke

Summary

This work performs time series analytics and sentiment analysis, specifically vector autoregression on return and sentiment data for several major banks in the United Kingdom and Ireland as well as the ISEQ and FTSE100 indices with a view towards determining if it is possible to detect sentiment before a breakdown in a specific sector of the economy. Time series analytics consists of examining stylised facts, summary statistics and correlations for each bank in order to perform a comparison of counterparts, both banks and markets.

Analysis indicates evidence against the Efficient Market Hypothesis and Mean Reversion in times of market crisis, in line with previous research. Estimates obtained indicate the presence and influence of sentiment in the movement of returns however due to the tenuous nature of the data underpinning the model there is not sufficient evidence available to reject the null hypothesis.

Abstract

Sentiment analysis is the process of using a computer to examine a piece of writing in order to discover the writer's opinions or feelings about something. It is useful for monitoring media in order to automatically determine the mood of consumers, this data can then be used by traders, marketers, company leaders and academics. The use of sentiment analysis has accelerated in recent years with the development of large-scale computational infrastructure that can analyse large unstructured textual data sets. This body of work looks at the effectiveness of sentiment analysis in the context of the most recent global financial crisis in 2007/2008. It will focus on the Irish and British banks most affected by the event as well as two major market indices and research will be performed in order to determine if it would have been possible to detect the influence of sentiment. Analytical methods are used to build up a picture of the various banks and indices and these stylised facts allow for better understanding and direct comparisons to be made.

Acknowledgements

I would like to thank my supervisor, Professor Khurshid Ahmad, for his time, comments, remarks and engagement through the learning process of this dissertation. I would also like to thank Dr. Carl Vogel for his advice and discussions on research methods. Finally I would like to thank friends and family for their support.

"And I rose
In rainy autumn
And walked abroad in a shower of all my days."

- Dylan Thomas, *Poem In October*

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter will begin with a short introduction to the problem domain. After this an overview of the research undertaken will be given with particular attention payed to the scope and objectives. From here, an overview of the Methodology employed during this research will be given and finally the chapter will conclude with a short outline of the remaining structure of the work.

1.1 Background

With the advent of the World Wide Web dissemination of information is now more rapid than ever. Billions of people are connected through a wide variety of mediums and not only consume but also readily produce content as well. Within the financial markets the ebb and flow of information can take many forms and the question of it's affect on the market is an interesting one.

Recent research has found that the information about financial instruments, for example stocks and shares, currencies and commodities, available in news and blogs increasingly appear to have an affect on the price of these instruments. More specifically it has been observed that the 'sentiment' within fact or opinion pieces can cause a change in the price of instruments.

1.2 Research Overview

The first objective of this research is to obtain and compute statistical features for the returns of British and Irish banks encompassing the period 1999-2010. Following on from this the second objective is to obtain sentiment data for the relevant banks and determine if sentiment in news pieces had any detectable influence on the value of their returns encompassing the period 1999-2010. Specific consideration will be given to the time period surrounding the 2008 Financial Crisis, 2008 - 2010. Two major indices, one from the Irish stock exchange and the other from the London stock exchanges shall also be examined.

1.3 Methodology Overview

The approach to be taken will consist of calculating specific stylised facts to allow for juxtaposition and differentiation of return data. Following on from this vectored auto-regression will be performed on returns and then later on returns combined with sentiment with a view towards determining if any influence can be detected on returns, on the part of sentiment.

1.4 Dissertation Structure

Chapter 2 will look at recent literature in this area of research with a focus on the concepts of sentiment and returns. Chapter 3 will introduce and discuss the methodology employed in completing the work with a specific emphasis on a variety of statistical facts to be employed, sources of data, instrumentation utilised and limitations considerations and difficulties. Chapter 4 consists of the research findings including stylised facts and correlations, regressions and visual observations. Following on from this Chapter 5 will present a discussion of the results obtained. Finally Chapter 6 will present a conclusion and recommendations for future research.

Chapter 2

Literature Review

2.1 Introduction

This chapter will discuss the background and motivation behind this research, established knowledge and ideas on this topic will be communicated with emphasis placed on journal published research in this field. Following on from this the concepts of arbitrage and sentiment analysis will be further engaged with. There will be a section on stylised facts and summary statistics to assist in the understanding and analysis of results. Finally there will be a section on hypothesis and approach before a summary conclusion.

2.2 Background and Motivation

The existence of the financial markets is important. Markets align supply and demand and allow people to trade in a wide variety of fungible items. Despite recent negative performance the financial industry, the continued functioning of the markets is a necessity for the existence and advancement of modern human society. In order for the exchange of goods and services to take place buyers and sellers must have an understanding of the value of such assets. The markets are used by buyers and sellers for price discovery for the purposes of exchange of goods and services. This price discovery process is the method by which the prices of assets are set through the interactions of buyers and sellers. Prices, for

the most part, are said to be governed by the laws of supply and demand, an economic model of price determination in a market: If demand falls or supply rises there will be more to go around and so there will be a decrease in the price in order for sales or exchanges to continue. Conversely if demand rises or supply falls there will be less to go around and so there will be an increase in the price.

What determines supply and demand is governed by a number of elements including production costs, number of suppliers, expectations about future prices, income, current trends, the number of customers and the prices of related goods and services. In the case of many financial products, in order to determine their value, investors perform financial analysis. Financial analysis can be broken down into fundamental analysis, technical analysis and behavioural analysis.

The two main types of financial analysis are fundamental and technical. Fundamental analysts looks at the underlying elements of a company: the balance sheet, directors, relative place in the market and competition in the market in an attempt to determine the value of a company (or currency, commodity, asset etc. as appropriate). Contrasting this, technical analysis looks at the historical movement of prices as well as the volume traded over the same period. The legitimacy of both these forms of analysis is disputed by the efficient-market hypothesis which states that market prices are inherently unpredictable, however, during times of regular normal market operation fundamental and technical analysis appear to suffice.

The other type of analysis, behavioural analysis, at the effects of psychological, social and emotional factors on economic events. It allows us to analyse the effect of human behaviour e.g. herd behaviour, cyclical patterns and repetition of historical events as a result of basic human behavioural patterns. Behavioural analysis is based on the idea that traders do not always act rationally. They discount correct information at times and believe incorrect information at others. These type of traders are known as noise traders because they often get their information or trade not based on facts but based on feeling and rumour.

Over long periods of time fundamental and technical analysis can help to ensure a return on an investment however in the short term the movements in the market every second, minute, hour, appear to be a random walk. While supply and demand are the ultimate governing factors, what role does sentiment in the markets play in the movement of prices? Another difficulty faced is that our traditional financial models such as the EMH have difficulty explaining financial crises. Many studies have documented long-term historical phenomena in markets that contradicts standard theory and these events are rarely predicted by large numbers of economists but usually by a small group. These cycles of boom and bust occur repeatedly throughout history but every time they arise the people involved are convinced that "this time is different". The 2007/08 Financial Crisis was not predicted by most experts yet has had a significant effect on the global economy and specifically the economies of Europe including the UK and Ireland.

While fluctuations in investor sentiment have bee regularly mentioned as an explanatory factor of financial crises, in the past they have been rarely analysed (De Long and Shleifer, 1991; Shiller, 2000). Controlling for known patterns however, recent research has found the predictability of returns using news content is concentrated in recessions (Garcia 2013). Additionally investor sentiment plays a role in volatility and in one study generated return predictability consistent with corrections of overreaction (Baker, Wurgler, Yuan 2012).

Tetlock (2007) quantitatively measures the interactions between media and market and finds that high media pessimism predicts downward pressure on prices followed by a reversion to fundamentals, and unusually high or low pessimism predicts high trading volume. Baker, Wurgler and Yuan (2009) determine that sentiment, both global and local, is a statistically and economically significant contrarian predictor of market returns while Schmeling (2009) finds that on average, sentiment negatively forecasts aggregate stock market returns across countries. As a result if market crisis are preceded by some change in sentiment, can it be seen in a subsequent forensic analysis?

Zouaoui et al. (2011) tested the impact of investor sentiment on international capital markets by studying its ability to predict the occurrence of stock market crises. They suggested stock market crises should be preceded by periods of rising investor euphoria. Therefore, it should be expected that periods characterised by excessive investor optimism are followed by market crises. Looking at the stock market crisis of 15 European countries and the U.S they concluded that investor sentiment increases the probability of the occurrence of stock market crises within a one-year horizon. In addition, the impact of sentiment is more pronounced on countries that are culturally prone to a herd-like behaviour and less institutionalised.

This is entering the realm of behavioural finance but it is important to remember that when we use sentiment we cannot be biased. When news comes out, even in the online world, it has to be typed up and then read. This means that often the first sign of the effect of a piece of news on a market is a change in the volume traded on that market as per Tetlock (2007). Traders will usually act rationally, however at times when the normal market conditions break down, when fundamental and technical analysis no longer suffice, we must turn to another form of analysis in order to attempt to understand movement of the markets.

For this research we will be concentrating on performing aspects of technical analysis on time series as well as delving into sentiment analysis. Pooling data is known to increase the ability of statistical tests providing better estimates (Ang and Bekaert, 2007). However due to cross-correlations between banks there will be fewer independent observations per period.

2.3 Arbitrage and Sentiment analysis

Arbitrage is defined as the simultaneous buying and selling of securities, currency, or commodities in different markets or in derivative forms in order to take advantage of differing prices for the same asset. An arbitrageur is a person who engages in arbitrage. In the event that buyers and sellers overvalue or undervalue the market together, the amount of trading taking place will increase as an op-

portunity for arbitrage is presented. Buying in one market and selling in another happens as close to simultaneously as possible in order to avoid risk.

Noise traders, also known as naïeve traders, are the people most influenced by news. They may be unfamiliar with a market or simply do not trade as a full time job and are therefore more easily swayed. Arbitrageurs are more often professional traders, they are not so easily swayed by news rumour or opinion, they perform fundamental and technical analysis when looking at their markets. Noise traders will be taken advantage of by arbitrageurs who understand the true value of a tradable asset. Arbitrage therefore has the effect of causing prices to converge, the speed of this convergence is used as a measure of market efficiency. This existence of arbitrageurs reduces the instances of price differentiation in commodities and also results in different currencies moving towards parity.
The following table illustrates the result when arbitrageurs meet naïeve traders:

		Naïve Traders	
Arbitrageurs		Overvalue	Undervalue
	Overvalue	Boom Market	Profit
	Undervalue	Profit	Bust Market

Figure 2.1: Market Functioning

A recently developed form of analysis is sentiment analysis. Sentiment analysis is the examination of structured or unstructured bodies of text with a view towards determining the subjective mood or feeling of the content, the sentiment. Looking at a set of textual data which may be a news article opinion piece, blog posts or other piece of writing, we must decide whether this set is positive or negative. This sentiment may prove influential and is worthy of analysis.

When the market is operating normally there is an equilibrium, a steady speed to the market. While this is the case fundamental and technical analysis should suffice for the purposes of pricing and long term investment. However in times of turbulence, in the event of a sudden change the balance in the market can suddenly change. Normal market forces may fail and traditional methods of analysis

prove useless. During these times news can have effects on market prices. Events such as an analyst opinion changes, a dividend payment, insider trading, a court settlement; can cause investor sentiment to change and valuations along with them. Additionally, during periods of market boom and bust when arbitrageurs meet noise traders and they move together, the ability of news and the media to influence the movement of the market comes into force.

Fundamental and technical analysis rely on so called rational behaviour; namely that over-priced or under-priced assets will be returned to their true valuation through their interaction in the market. In addition, because fundamentals are priced into the value of a stock, currency or commodity at a specific point in time, their influence on the day-to-day ebb and flow of the markets is limited. Sentiment analysis therefore is related to unexpected fluctuations and is primarily used as an emergency measure. A change in volume traded is often regarded as the first sign of sentiment as volume is usually impacted first, this is because it takes time for a piece of news to come in, be translated, created, sent out, discovered, taken in and processed. The belief is that large deviations of the expected value are caused by sentiment. However sentiment analysis will only work when movement depends on the irrationality of people. It works when the normal rules of the market are failing. At best, it allows us to explain 20bps (basis points: 1 percent of 1 percent) of movement but when it is able to explain this, it is the only thing that will. We focus on negative sentiment because people are more positive and the existence of negative is more significant.

2.4 Summary

Price is primarily governed by supply and demand with a number of forms of financial analysis being used as part of price discovery. Human behaviour has an influence on the movements of the markets, though many investors are rational there are those who are not. As a result investor sentiment plays a role in the movement of prices though research suggests the predictability of returns is concentrated in recessions. Sentiment analysis is used to quantitatively evaluate the role of sentiment. The existence of arbitrageurs mean that prices do not stray from their correct value for long but over-corrections and under-corrections may occur.

Chapter 3

Methodology

3.1 Introduction

This chapter describes the methodology used to collect and process the research data. Objectives and Rationale provide a statement of the research objectives and goals as well as a rationale for the approach taken. The section on Data Sources describes the sourcing of research data whilst Instrumentation discusses each tool used to source and process data. Target Institutions provides a background of the chosen organisations, broken down by country of origin.

3.2 Objectives and Rationale

The objective of this research is not to predict the future but to perform a forensic analysis of econometric data post the 2008 Financial Crisis, to look at one economic sector specifically, to analyse this sector of the financial markets and to determine if we can detect the influence of sentiment in the movement of prices. As a computer scientist, this research will focus on the numeric data available and all estimations will be given relative to this. The approach will be one of obtaining econometric time series data and using statistical methods to analyse it. Finally, sentiment data will be obtained and processed with a view towards discovering whether its influence can be detected. For this work, a specific economic sector has been chosen which has had a large impact on the markets, the

banking sector. Econometric and text data will be sourced for a specific time period. This period will encompass the financial crisis and allow us to make inferences and estimates.

There are a number of considerations when sourcing data; firstly, sentiment data from before 1999 is quite limited and therefore no data will be sought from before this year. Secondly, the underlying currency of the euro area was put in January 1999, as a result econometric time series data from before this year will not be used in statistical calculations or observations.

3.3 Hypotheses and Approach

For this dissertation I will be looking at the recent financial crisis and specifically at the British and Irish banks most heavily involved, in order to determine if sentiment analysis played any role in their fall. I will be taking data relating to the banks most affected by the crash: Standard Chartered, Northern Rock, HBOS, Barclays, RBS Group, Lloyds, HSBC, Anglo Irish Bank, Bank of Ireland, Allied Irish Bank, Irish Life and Permanent as well as the FTSE 100 and the ISEQ index.

I will gather a number of time series relating to the above financial institutions and measures. Visual examples of these have were presented previously. With these data sets I will be able to perform technical analysis which will return some empirical statistical regularities. This will be tied in to sentiment analysis of the same period and correlations will be sought. The null hypotheses H_0 is that it is not possible to detect sentiment before a breakdown in a specific sector of the economy. The alternative hypotheses H_1 expresses the effect that we wish to find evidence for, namely that it is possible to detect sentiment before a breakdown in a specific sector of the economy. The null hypotheses has been asserted and this body of research will seek any evidence that might allow for its rejection. The approach taken will follow a series of steps.

Fundamental technical and sentiment analysis will be performed on each of the institutions of interest. After collecting the relevant data on shares the prices and volume traded will be studied and relevant stylised facts, as well as returns,

will be calculated. Once the data is collected and relevant stylised facts calculated we will have the necessary statistical evidence about the movements of prices and changing of volume to begin the process of sentiment analysis, a test of significance will be designed to assess the strength of the evidence against the null hypothesis in an attempt to answer this question.

3.4 Returns vs Prices

Returns are profits or losses made upon investments and the average return over a single day is a very small figure. Previous research from as far back as Working (1934) and Kendall (1953) has shown that there is almost no correlation between returns for different days. In contrast the correlation of prices from day-to-day is exceptionally high. This is intuitively understood as the value of the price today will nearly fully depend on the value of the price yesterday. Prices and returns can be easily visualised in graph format and graphs of price and return for the banks and indices in this work will be included in Chapter 4. Fundamental analysis sets the price at a specific point in time. Day-to-day fluctuations are therefore affected by other factors. While prices are a commonly observed and easily understood figure they are affected by inflation. In contrast returns are not and the log of the return is a useful quantity as it is used to calculate the rate of return. On average, observations shows that returns in the market move towards 0 due to the previously mentioned interplay between naieve traders and arbitrageurs and so returns are the basis for this analysis. The formula for calculating returns is as follows where R is return, P is price and t is time:

$$R = \log(\frac{P_t}{P_t - 1})$$

3.5 Stylised Facts and Summary Statistics

An important part of financial research is the analysis of relevant financial data, this has led to the characterisation of some empirical statistical regularities, known as stylised facts. The term stylised facts was introduced by the economist Nicholas Kaldor in the context of a debate on economic growth theory in 1961.

Kaldor argued that any construction of a theory should begin with a summary of the relevant facts. These regularities are statistical properties of daily returns that have been documented for the mean, variance, distribution and auto-correlation of returns across time and across markets and they are as follows. Firstly, the correlation between returns for different days is low. Secondly the distribution of returns is not normal and finally the correlation between the size of returns on nearby days is positive and statistically significant. In addition to stylised facts there are summary and descriptive statistics that can assist in any analysis being undertaken. These summary and descriptive statistics are as follows.

3.5.1 Mean

The mean is a measure of central tendency otherwise known as the average. The Efficient Market Hypothesis says that in an efficient market the mean should always reverse upon itself, going from positive to negative. Clustering is when the mean continually moves in the same direction over time and is a sign that the market is not functioning correctly

3.5.2 Standard Deviation

The Standard Deviation is a measure of the distance from the mean. A value of 0 indicates that data points tend to be very close to the mean while a high value indicates that points are spread out over a wide range. For any normally distributed data set 68.27%, 95.45% and 99.73% of the values lie within one, two and three standard deviations of the mean. Even for non-normally distributed data at least 98% of the data should fall within three standard deviations. The standard deviation changes over time and within econometrics and finance this is used as a measure of volatility. Volatility indicates the tendency of stock prices to rise and fall quickly. A stock whose returns differ significantly from the mean return over a short period will have a high standard deviation.

3.5.3 Skew and Kurtosis

The skewness of a data set measures the symmetry/asymmetry of a distribution. If the value for skewness is less than 0 then a distribution is skewed to the left and is said to be negatively skewed. If the value for skewness is greater than 0 then a distribution is skewed to the right and is said to be positively skewed. If the value is 0 exactly then a distribution is a normal distribution and is symmetrical.

Kurtosis is a measure which tells us how tall of flat a distribution is. A value of 3 indicates that a distribution matches the normal distribution in terms of height. A value of less than 3 indicates that a distribution is flatter than a normal distribution. Conversely a value greater than 3 indicates that a distribution has a very high peak.

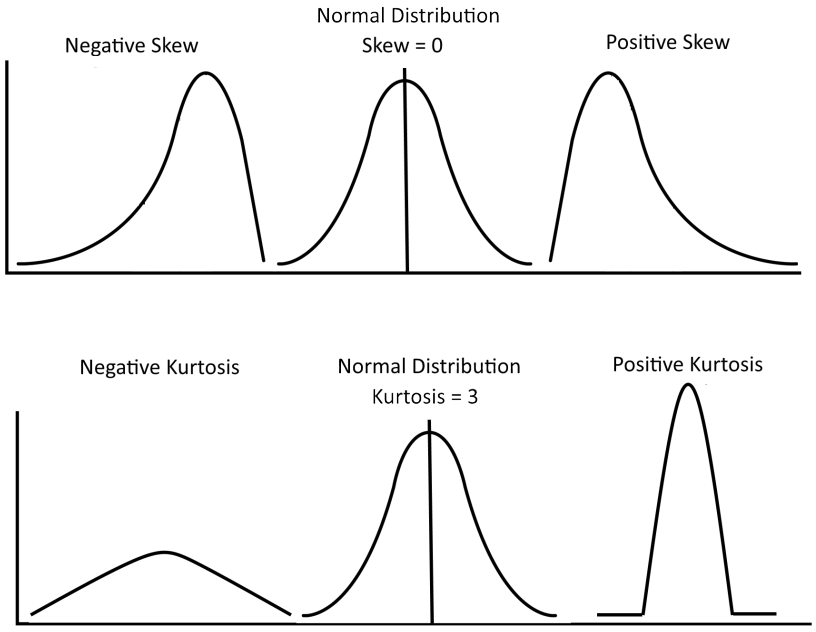


Figure 3.1: Skew and Kurtosis Distribution Examples

3.5.4 Z-Score

A z-score or standard score is a measurement of the relationship of a variable to the mean. A result of 0 indicates that the variable being scored is the same as the mean. A positive score is above the mean while a negative score is below and the score is indicative of how many standard deviations away form the mean a variable is. As with standard deviation for any normally distributed data set 68.27%, 95.45% and 99.73% of the values lie within one, two and three standard deviations of the mean. The z-score of a variable x is as follows:

$$z = \frac{x - \mu}{\delta}$$

Where x is the variable under test, μ is the mean and δ the standard deviation.

3.5.5 Correlation

The correlation of elements in a data set indicates how closely related these elements are. A high value indicates that data points are very correlated, a low value indicates that they are not. For example when we correlate prices on the market we find that the price today is heavily determined by the price yesterday and this is indicated by a high correlation. Correlation does not imply causation though as there can be other factors involved.

3.5.6 Additional Elements

Minimum and Maximum: These show the smallest and largest value in a data set respectively. Data points may or may not be outliers within the bounds of a normal distribution.

Mode and Median: The mode tells us the most common or frequent value in a data set while the median tells us the middle value of an ordered data set, the value separating the upper half from the lower half. Mode and median are two ways of summarising typical values associated with members of a data set.

Variance and Range: Variance is a measure of how spread out a set of data points are. A value of 0 indicates that all values in a data set are the same. A small variance indicates that all points are close to the mean and a high variance shows that data is spread out from the mean. Variance is always non-negative and it is used as a measure of risk. The range of a data set is the size of the interval which contains all the data and is indicative of the dispersion of a data set.

R^2 **and Adjusted** R^2: After a linear model has been fitted to a data set using regression analysis it is necessary to determine how well the model fits the data, R^2 is a measure of this. A model is deemed to be good if the differences between the predicted values and the actual values observed is small. In the results presented below, 0 indicated that the model explains none of the variability of data while 1 indicates a model has perfect fit. It is worth noting that any research attempting to predict human behaviour typically has R^2 values lower than 50%. Adjusted R^2 is a modified version of R^2 calculated by taking into account the number of predictors in a model. Adjusted R^2 increases if the addition of a variable improves the model more than chance would predict and decreases otherwise. Generally R^2 and Adjusted R^2 are defined as:

$$R^2 \equiv 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \qquad \bar{R} = 1 - (1 - R^2)(\frac{n-1}{n-p-1})$$

Where p is the total number of regressors in the model, n is the sample size, $\bar{y}$ is the mean of the observed data, y_i is the n values in the data set and f_i is the predicted values in the data set.

Durbin-Watson: The Durbin-Watson test is used to detect the presence of a relationship between values separated by a time lag (auto-correlation) in the prediction errors (residuals) from a regression analysis. The result of the test will always lie in the range 0 - 4 however if it is equal to 2 it indicates no auto-correlation, if it is substantially less than 2 there is evidence of a positive correlation meaning that a positive error for one observation increases the changes of a positive error for another. If it is greater than 2 there is evidence of negative

correlation which implies that a positive error increases the chance of a negative error and a negative error increases the chance of a positive. It is worth noting that the test underestimates auto-correlation for auto-regressive moving average models and is not applicable when lagged dependant variables are included as explanatory variables. It is defined as the following where T is the sample size, t is a time and e_t is the residual at that time:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_t - 1)^2}{\sum_{t=2}^{T} e_t^2}$$

p-value and t-ratio: The p-value is used in hypothesis testing in order to determine statistical significance. A result is said to be statistically significant if it allows for the rejection of the null hypothesis. If $p > 0.05$ then a result is also not statistically significant, however if $p < 0.01$ then a result is highly significant. The t-ratio is another value which can be examined to determine statistical significance. It is calculated by dividing the coefficient in question by its standard error. T-ratios greater than 1.96 in absolute value suggest that a coefficient is statistically significant at a 95% confidence interval.

F Statistic: An F test determines if a group of variables are jointly significant. If this is the case it is likely that any results did not happen by chance. It is used in combination with the p-value and t-ratio when deciding whether to support or reject the null hypothesis. If $F < 1$ a result is not statistically significant.

3.6 Data Sources

Two forms of data were required for this work. The first was time series econometric data showing prices for the period leading up to and immediately following the financial crisis. The second was text data upon which sentiment analysis could

be performed. A number of systems were used to source and analyse this data. For econometric time series, the primary source was Datastream with additional checks being performed with Quandl and Yahoo Finance. For sentiment data, the primary source was the Financial Times Archives, accessed through Lexisnexis.

3.6.1 Datastream

Datastream is a database provided by the information firm Thomson Reuters that, amongst many other things, contains financial data for equities, indices, currencies and fundamentals, for a large number of markets around the world. Much of this data comes in the form of time series and data coverage is quite strong. The bulk of the time series data used in this research was obtained via Datastream Professional, the proprietary software used to access the database.

3.6.2 Quandl and Yahoo Finance

Quandl is a numerical data search engine providing access to a large number of financial and economic data sets. Yahoo Finance performs a similar role providing financial news, market data and commentary. These two systems were used as a comparison against Datastream in order to ascertain that the maximum amount of data points were being collected and that these points were reliable and valid.

3.6.3 Lexisnexis and the Financial Times Archives

Lexisnexis is a business and legal news data repository. It was used to source structured text data on which Sentiment analysis could be performed. Text data could be returned in multiple formats including HTML, PDF and Word however the simplest text formatting was chosen given the time constraints of the project and to allow for reduced memory usage during computation. To ensure quality and relevancy, all text data pulled was taken from the Financial Times Archive and the Financial Times London Archive. These cover a period from 1982 for newspapers and from 2006 for online content and are well respected reliable sources of data. Relevant text data is acquired by imputing search terms into the LexisNexis search engine, this then searches the relevant chosen source

for text data mentioning this term or terms. For this project the search terms utilised were the names of each of the banks and indices of interest. These terms were chosen in an attempt to ensure maximum reliability and validity.

3.7 Instrumentation

A description of each instrument used to collect and process data is provided as well as a rationale for each instrument used. For sentiment data, all processing was done with a tool called Rocksteady. Subsequent time series data produced as well as econometric time series data obtained were processed through Excel and with GRETL.

3.7.1 Rocksteady

Rocksteady is an affect analysis engine developed at Trinity College Dublin. Written in Java, Rocksteady takes in structured or unstructured text and through the use of general and domain specific affect dictionaries measures the emotional content of this text. Text is acquired through local storage or via RSS feeds; it can be time stamped or a time period can be selected. Sentiment analysis and term relevance is performed and a time series is generated which can then be analysed, compared and contrasted. For this work, structured time stamped news text was obtained and analysed by Rocksteady. The resulting time series produced were then able to be analysed with respect to the econometric time series obtained on the relevant institutions. Rocksteady is written in Java and was developed by a team led by Professor Khurshid Ahmad. Using it meant ready access to good support and a good understanding of the system and its workings was provided.

3.7.2 GRETL

The GNU Regression, Econometrics and Time-series Library (GRETL) is an open source statistical package primarily used to apply mathematics and statistics to economic data as part of performing quantitative analysis. GRETL is written in C and has been in development since early 2000. It supports several statistical estimators, time series models and econometric tests and the open source nature

of the project makes it appealing for academic research. GRETL was used to analyse econometric time series data, specifically stock market returns, as well as sentiment time series data provided by Rocksteady. Specifically, GRETL was used to perform Vector Autoregression which will be discussed later. The package was chosen because of its open source nature and ease of use as well as its ability to inter-operate with time series produced by Rocksteady.

3.7.3 Excel

Excel is a spreadsheet application developed by Microsoft. Time series data was held and stored in Excel format for processing and analysis. Excel provides a number of useful computational abilities including the ability to manually work with data to produce correlations and derive useful descriptive and summary statistics. Time series produced by Rocksteady and pulled from Datasteam were in the same standard Excel format and Excel could save data to a format efficiently consumed by GRETL.

3.8 Target Banks

A number of banking and financial services companies across Britain and Ireland were chosen as the focal point of this research. They were chosen based on organisation size, scope of business and the affect the financial crisis had on them. The following section will look briefly at each of these institutions in turn, divided up across the two countries. It is worth noting that Irish banks play a much more dominant role in the Irish market and specifically the ISEQ Index in comparison to their British counterparts. Thus the influence on the ISEQ by the Irish banks will be greater. While some Irish banks have an international presence, the preponderance of British banks are large scale multinational financial organisations with a global presence that far exceeds their Irish contemporaries.

3.8.1 Britain

The Big Four is a phrase used to describe the four largest banking institutions in a country. In Britain, the Big Four are currently Barclays, RBS, Lloyds and HSBC.

These were taken as a starting point for all research conducted and to them were added Northern Rock and HBOS as both were heavily affected by the financial crisis of 2008. Finally, Standard Chartered was added as this bank emerged from the crisis relatively unscathed. The British government's reaction to financial crisis consisted of a bank rescue package deployed in late 2008, designed to shore up investor confidence in the markets and to stabilise the banking system in Britain. The initial credit crisis meant that organisations were no longer able to borrow money to fund themselves on a day-to-day basis. This plan provided for a range of loans and guarantees as well as a series of direct investments in the banks themselves, which would provide the banks with funding. This, it was hoped would carry them through the crisis while the government guarantee would encourage stability. The British plan was quickly followed by the rest of Europe and also the United States which modified its previous plan to take after the British. This unified action had a positive affect on the markets, though the share prices of affected banks were not stabilised. A second rescue package followed in the form of a banking act in early 2009, this was designed to increase bank lending and took place after the bailouts of a number of large British banks.

Standard Chartered

Despite being a British multinational bank and financial services company, Standard Chartered conducts no retail banking in Britain. Much of the bank's profits come from Asia, the Middle East and Africa. The bank has a primary listing on the London Stock Exchange however, and is listed on the FTSE 100. Standard Chartered managed to escape the 2008 financial crisis relatively unharmed and even posted a profit during this period.

Northern Rock

Northern Rock was a British bank formed when the former Northern Rock Building Society was floated on the London Stock Exchange in 1997. During the financial crisis the highly leveraged positions taken by the bank meant that it was one of the first to run into difficulties. it was forced to approach the Bank of England for a loan to replace money market funding, which had dried up. This caused a

run on the bank, the first in 150 years, and having failed to find a buyer the bank was nationalised in early 2008. At this point, it was suspended from the London Stock Exchange having already been removed from the FTSE 100 in late 2007. The difficulties faced by Northern Rock would prove to be an early indicator of what would befall other banks.

HBOS

HBOS is a banking and insurance company formed in 2001 from the merger of Halifax Bank and the Bank of Scotland. Following failures of management and regulatory oversights, the bank,s position became untenable upon the onset of the financial crisis and it became necessary to join the emergency recapitalisation plan provided by the UK government. Having created a bank of comparable size to the other Big Four UK banks, it then became a subsidiary of Lloyds Banking Group, being acquired in 2009.

Barclays

Barclays is a large British multinational banking and financial services company. It has a primary listing on the London Stock Exchange and is included in the FTSE 100 Index. During the 2008 Financial Crisis, Barclays purchased the investment banking and trading divisions of Lehman Brothers, a United States investment bank that had filed for bankruptcy however Barclays themselves, however, were hit by the crisis making a loss during the period but sought to raise necessary capital privately rather than seek a cash injection from the British Government. The Barclays' sentiment data obtained was the most comprehensive out of all the banks chosen.

RBS Group

The Royal Bank of Scotland Group owns The Royal bank of Scotland, Ulster Bank and NatWest. The group operates in a number of countries and provides retail banking facilities throughout the UK and Ireland. Before the financial crisis, RBS Group was briefly the largest bank in the world. During the Financial Crisis it was forced to seek funding from the British government, which then took

a majority shareholding in the bank. This was increased at a later point, however, the banking group retains its listing on the London Stock Exchange and remains one of the largest companies on the market. The RBS loss posted in the fourth quarter of 2008 is the largest annual loss in UK history at 24.1 billion.

Lloyds

Lloyds bank is a British retail bank and traditionally considered part of the Big Four. Previously called Lloyds TSB, the bank was renamed following the acquisition of HBOS by Lloyds TSB Group which was renamed to Lloyds Banking Group. The group is listed on the London Stock Exchange and forms part of the FTSE 100. In response to the financial crisis the UK government announced a bank rescue package consisting of a recapitalisation fund. Through this the government bought a combination of shares in banks affected by the ongoing crisis. After the recapitalisation's and Lloyds' acquisition of HBOS, the Government owned over 40% of the Lloyds Banking Group. Following the discovery that HBOS losses were greater than predicted, Lloyds' share price on the London Stock Exchange fell and took several other banks with it. The UK government took a majority stake in the bank.

HSBC

Hong Kong Shanghai Banking Corporation, commonly known as HSBC, is a British multinational banking and financial services company and the third largest in the world by assets as of 2014. HSBC has a dual primary listing on both the Hong Kong Stock Exchange and the London Stock Exchange with secondary listings on a number of others. With its large market capitalisation it is one of the foremost companies on the FTSE 100. Although it was at the centre of a storm the wider HSBC group performed well during the 2008 financial crisis and indeed was one of the strongest banks to emerge, requiring no financial assistance from the British government by bringing in necessary capital from its overseas business.

FTSE 100

The Financial Times Stock Exchange Index 100 is a benchmark index comprising the one hundred largest companies on the London Stock Exchange by market capitalisation. Since its foundation in January 1984, the largest crashes of the index have been following the dot-com bubble burst in 2000 and the 2008 financial crisis. Following this fall the index has since recovered to previous levels and continues to ascend.

3.8.2 Ireland

In Ireland the Big Four are Allied Irish Bank, Bank of Ireland, Ulster Bank and until recently Danske Bank; however due to its withdrawal of personal banking services it is being replaced by either KBC Bank or Permanent TSB. Because of their primarily foreign composition both Danske Bank and KBC Bank have been excluded from this research.

The amount of sentiment on Irish banks that can be retrieved is limited compared to corresponding British banks. This is because the primary source, The Financial Times, is a British news source and thus the majority of mentions of Irish organisations such as the ones mentioned below are mostly limited to more newsworthy events. This is reflected in the count of news articles retrieved. As a result of the exposure of Britain to Irish banks as well as the international operations conducted by the largest banks there is enough data to conduct research. It is worth noting, however, that the bulk of international newsworthy events, as a result of their nature and the nature of the news source, will be heavier on both positive and negative sentiment compared to local news.

Anglo Irish Bank

Anglo Irish Bank was a large Irish commercial bank which was mortally wounded by the decline in the Irish property market in 2008. Following the downturn, the bank posted the largest loss in Irish corporate history and as a result of its exposure and untenable position, was nationalised by the Irish government. The bank was de-listed from the Irish and London stock exchanges upon nationalisation in late December 2008.

Bank of Ireland

Bank of Ireland is the oldest and largest banking institution in the Republic of Ireland. While never acting as the Central Bank of Ireland, like the Bank of England or Bank of Scotland, it has performed much of the functionality of a Central Bank at various times. It provides a broad range of financial services and operations both in Ireland and abroad. Post the 2008 Irish banking crisis, the institution was left with no choice but to take part in the Irish bank recapitalisation scheme however it was not taken over by the government and its position was tenable, when compared to the other major Irish banks.

Allied Irish Bank

Allied Irish Bank was formed in 1966 from the merger of Provincial Bank of Ireland, the Royal Bank of Ireland, and the Munster & Leinster Bank. This was done in order to tackle the fragmented Irish banking industry of the time and led to the creation of one of the largest banks in the Irish Republic. The bank offers a range of both personal and commercial services. Following the 2008 Irish banking crisis, the institution was forced to accept a bailout from the Irish government. As part of this bank recapitalisation scheme, the government took almost full control of the bank and its shares were de-listed from both the Irish Stock Exchange and the London Stock Exchange.

Irish Life & Permanent

Irish Life and Permanent was an Irish financial services and personal banking company formed from the 1999 merger of Irish Life Assurance and the Irish Permanent Building Society. Following the 2008 banking crisis, the company was split with part of the organisation being sold to the Irish state and the remainder coming under majority government control and now trading under the name Permanent TSB.

ISEQ Index

The ISEQ Overall Index is a benchmark stock market index comprising all stocks, except for investment funds and companies registered in the UK, that trade on the Irish markets through the Irish Stock Exchange. The index is useful for comparing the performance of the market and companies on the market over time. Another index, the ISEQ 20, includes the top 20 companies on the market.

3.9 Approach

First, stylised facts will be calculated for both the return time series and sentiment time series. These estimates will be noted and analysed, they will allow for comparisons and observations to be made. Following this GRETL will be used to perform vector autoregression on return and then return with sentiment data. These results will then be noted, interpreted, analysed and discussed. Lastly a conclusion will be made.

3.9.1 Vector Autoregression

Vector Autoregression is most easily understood as a linear model in which each variable has an equation which explains changes in the value of that variable based on its own past values (called lags) and the current and past values of other variables in the model.

VAR is used as the success of the model can be easily judged. The full equation for calculating VAR used in this work is as follows, where r is returns, α and β represents the coefficients of 5 lags of return, s represents the coefficients of sentiment and ϵ the error term.

$$\mathrm{r}_{\,t} = \alpha_0 + \alpha_1 \,\mathrm{r}_{\,t-1} + \alpha_2 \,\mathrm{r}_{\,t-1} + \beta_1 \,\mathrm{s}_{\,t} + \beta_2 \,\mathrm{s}_{\,t-1} + \epsilon_t$$

VAR estimates are based on endogenous (internal) variables and exogenous (external) variables. In this work, endogenous variables would include price and

return, which are internal to the model and the exogenous variable we are interested in is sentiment which is hypothesised to come from the outside and influence the system. GRETL allows for the calculation of a VAR model which takes into account both endogenous and exogenous variables. This will help us determine what impact, if any, sentiment has on returns.

Return will be regressed against return, sentiment will then be added into the modelling and regressed against as an exogenous variable. A constant is included as an independent variable, this allows the trend line to have a y-intercept. The constant is ignored and never interpreted but without it the trend line would go through the origin which could negatively affect results.

3.9.2 Correlation

Among the first steps in the analysis of many time series is the calculation of correlation coefficients. Coefficients are estimates of the impact of one unit change of an independent variable on a dependant variable. There are several correlation coefficients but the one most commonly used is the *Pearson product-moment correlation coefficient*. Noting whether the coefficient is positive or negative is of use as this tells us whether the dependent variable is positively or negatively affected by the independent variable. A result in the range 1 to -1 where 1 implies a full positive correlation, -1 a negative correlation and 0 no correlation. Statistical significance is important as this indicates to us whether the results were by chance or not by chance. The p-value indicates significance and is based on the t-ratio which itself is found by dividing the coefficient by the standard error. These results will allow us to determine if there is any statistical relationship between two variables or two sets of data.

Correlation amongst returns can be seen during periods of high market volatility and it has been theorised that this is indicative that the markets are not functioning correctly (efficiently). Correlation has also been attributed to breaks in the distribution of returns or market contagion during periods of unrest. Following on from the discussion of returns vs prices in the previous chapter, taking

returns over prices reduces the effects of multicollinearity in the model which is where two or more predictor variables are inter-correlated. This will reduce the size of the standard errors and increase the precision of the estimate of any one variables impact on a dependant variable while controlling for others.

3.10 Limitations Considerations and Difficulties

There are a number of limitations with the methodology as well as considerations and difficulties to be aware of. The econometric data set may include the addition of holidays. Standard practice is to omit all holidays and non-trading days from data sets and so, if present, these days will be removed. As previously mentioned, sentiment data from before 1999 is unreliable and the amount of data that can be obtained decreases the further back in time you go so it is necessary to focus on the more recent past.

Sentiment analysis engines traditionally have a number of difficulties when analysing text data. The process of expressing one's meaning by using language that normally signifies the opposite, known as irony, is something text analysis engines often struggle with. Correct processing of this requires an understanding of the greater context of a piece. Sarcasm, the use of irony to mock or convey contempt, could very well be used in a piece of discussion text such as an opinion piece and this is something to be aware of. An additional difficulty is words with multiple meanings and where the semantic orientation of phrases may have an impact on the sentiment. There is no direct solution to this but the Rocksteady system contains a number of affect dictionaries which classify terms in a broad manner. In addition, sourcing data from a reliable news organisation like the Financial Times, where the standard of writing will be more formal should help to be of benefit.

Articles or other pieces of writing may be about multiple topics and misclassifications of sentiment for a specific article may occur in this event. One approach to mitigating this is to aggregate as many articles as possible to minimise this risk. In general it is important to get a good representative sample of articles

and from a reliable source. Choosing the Financial Times was the most logical choice as this is a large well respected organisation with good archiving practices. Whilst a good sample is important, it is also important to handle duplicate articles as the presence of these can skew results.

When aggregating articles together reduces the risk of misclassification it may also increases the risk of encountering duplicate articles. Duplicate articles will espouse the same sentiment as each other and have the effect of strengthening the sentiment detected in the same direction as the original. Sourcing sentiment data from one primary source, the Financial Times, reduces the chances of this occurring as the newspaper is unlikely to publish the same or similar articles more than once although this does not completely remove the possibility.

The search term employed to obtain articles on a relevant topic is of great importance. In this work the name of the bank will be used to search for and obtain sentiment data for each specific bank. This will then be aggregated together and used as the sentiment for each index. The possible presence of multiple topics in any given article and the difficulty in determining if an article is truly about a specific bank are limiting factors however the quality of the data source and the use of market indices will be of assistance.

Historically bank shares are usually stable and do not experience much volatility. If they collapse they are usually the first or the last in the markets to do so. Although most large Irish institutions have an international presence and trade between Britain and Ireland is strong, Irish banks are mentioned less and therefore usually when things are "newsworthy". In addition, the market capitalisation and international presence of British banks is substantially greater than that of Irish banks and as a result a greater number of British banks will be targeted.

Chapter 4

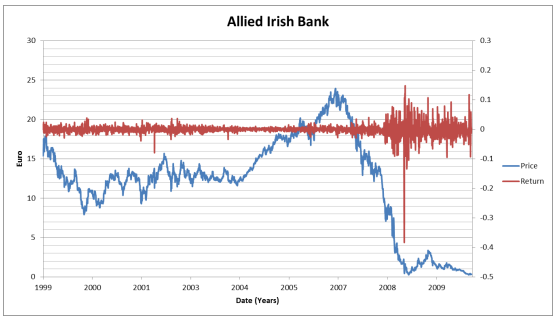
Research Findings

4.1 Introduction

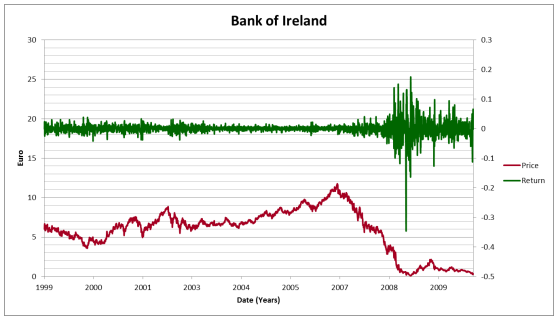
This chapter presents the findings of the research. Tables, figures of results and charts will be used to illustrate patterns in the data presented here. The chapter is organised and broken down into defined sections and is restricted to the presentation and analysis of collected data, without drawing conclusions or comparing results. Appendices contain additional data that is relevant but whose inclusion does not add significantly to the discussion.

4.2 Returns and Prices

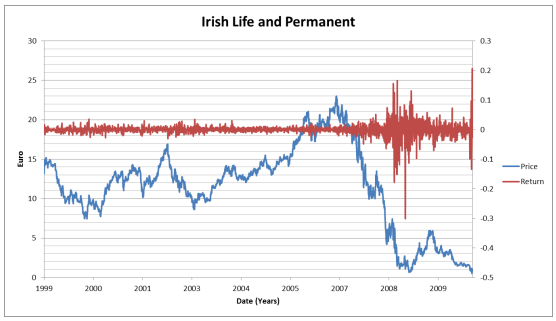
Following on from Returns vs Prices in Chapter 3, prices and returns can be easily visualised in graph format. A number of graphs of price and return for the banks and indices are included below. Price trends show the effect of inflation however the most prominent feature that can be observed surrounds the 2008 financial crisis. Increased volatility can be seen and a distinction observed between those banks which fared well and those which did not.



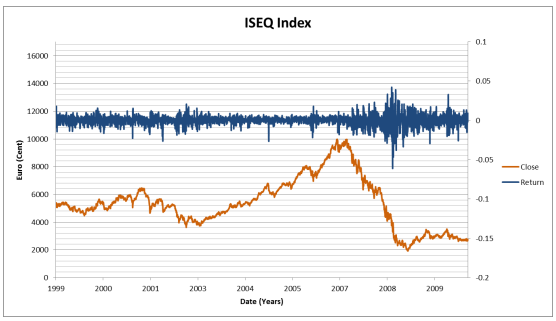
(a) AIB



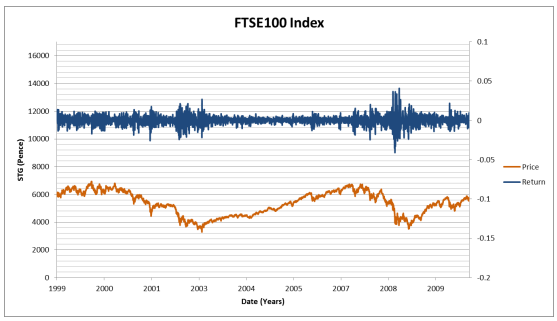
(b) BoI



(c) ILP

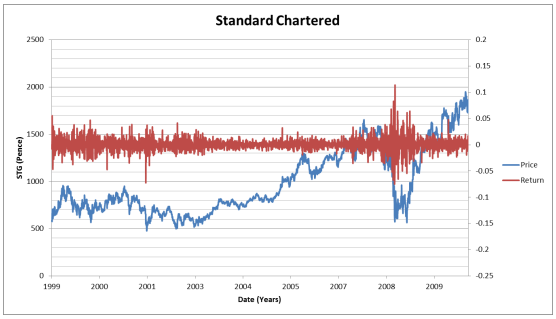


(d) ISEQ

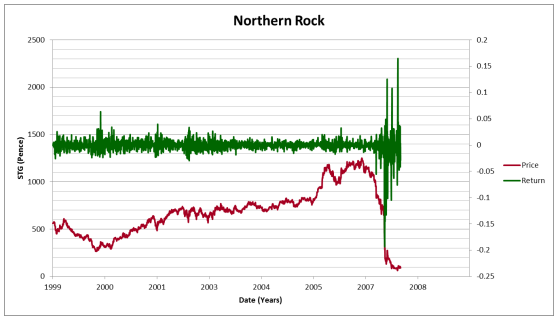


(e) FTSE

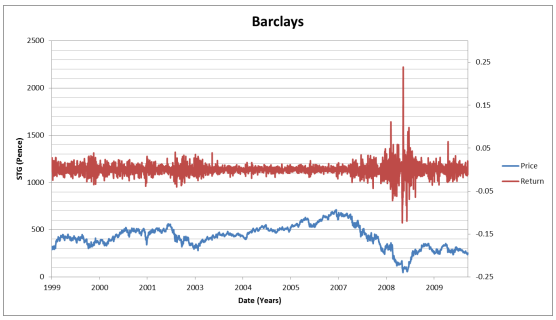
Figure 4.1: Charting price and return for each bank as well as the ISEQ and FTSE100 Indices, continued on the next page.



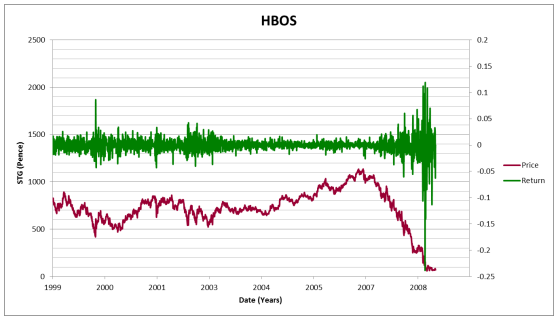
(a) Standard Chartered



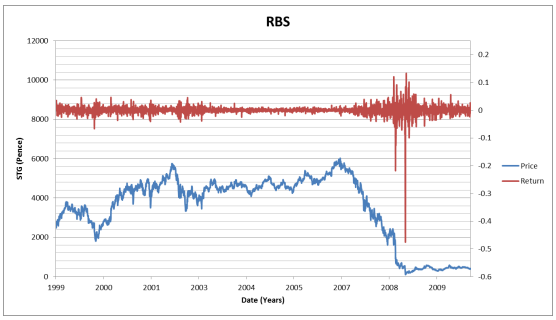
(b) Northern Rock



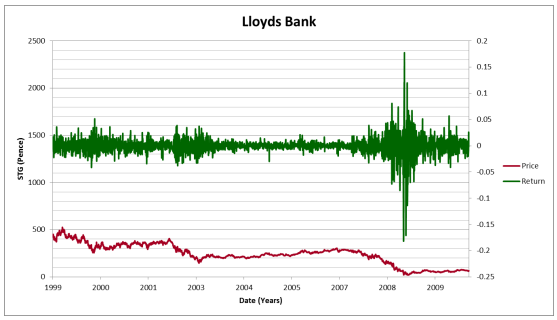
(c) Barclays



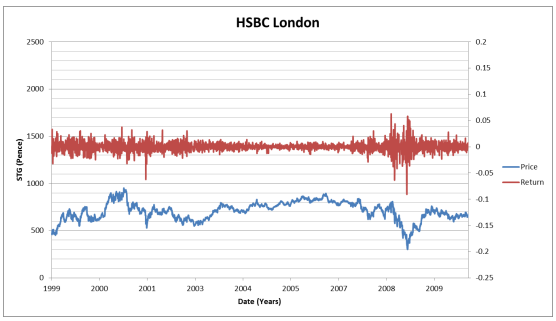
(d) HBOS



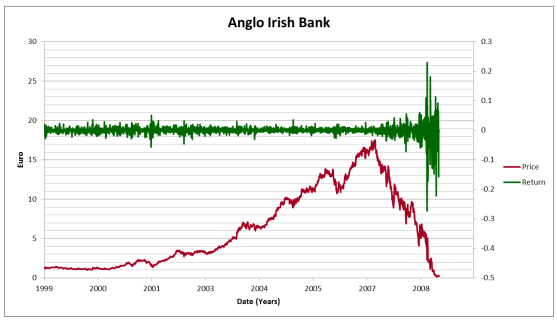
(e) RBS



(f) Lloyds



(g) HSBC



(h) Anglo

4.3 Stylised Facts

A number of statistical properties of the data collected are presented below. These properties are well understood and expected to be found in all relevant data sets. As a result they are useful for analysing, comparing and contrasting across different banks and across time. Statistical properties as well as a discussion of analytical methods are mentioned in the literature review.

Holidays where no trading takes place have been removed as per standard practice. The data fetched has all been taken from the same time period however because of differences in the operation of the markets between Britain and Ireland, for example different bank holidays, the starting and finishing point of the data and therefore the base sample size differ slightly. In addition, owning to the early-failure of some of the banks their time series have less samples than the surviving banks, this can be seen in the tables below. In general, price and sentiment data has been fetched starting January 1st 1999 and finishing December 3rd 2010 covering a period of nine years before the crisis and two years after. These dates were chosen because, as previously noted, sentiment data sets from before this time are both small in size and unreliable.

Sentiment data has been retrieved for the largest most heavily affected Irish banks - Allied Irish Bank, Bank of Ireland and Anglo Irish Bank. Data has also been retrieved for Standard Chartered, Northern Rock, HBOS, Barclays and RBS. Some charts in this chapter will demonstrate the structure of some of this sentiment data while computed stylised facts will be listed below and observations made. The results will be discussed in Chapter 5. The following tables include stylised facts as related to returns. Observations will be made based on these estimations and following on from this tables of stylised facts related to returns will be listed and observations made on these. The regression analysis of a smaller subset of the banks will be examined looking at return and return vs sentiment; and finally, sentiment will be modelled as an exogenous variable with the FTSE 100 and ISEQ indices as these include all the major banks.

4.3.1 Stylised Facts for Returns

Our estimates show that the mean of all the banks in Ireland and the majority of the banks in Britain is negative. The FTSE100 and ISEQ also have negative values indicating that the market is being affected although their values are among the smallest. The greater effect on the ISEQ is indicative of the makeup of the ISEQ index relative to the FTSE100.

The values demonstrate that over the period of time observed, the majority of the banks studied made a loss, in line with the markets. However it is interesting to note that Standard Chartered and HSBC both have positive values which reflects their stronger positions and indeed these banks proved more stable.

Profits from the banks are low or nonexistent however it can be observed that the numbers for Ireland are lower than the numbers for the UK. The noticeable exceptions in Britain are HBOS and Northern Rock both of which were bailed out and no longer exist. It can be seen that aggregate numbers demonstrate the reality of the situation. Negative values for skew indicate that all banks and indeed the markets in general are making losses. The kurtosis for all of the banks is outside of the normal expected value, in a number of cases by a considerable amount.

Fluctuations in the market are demonstrated through the minimum and maximum, performing a Z-score calculation we can see that movements at this range are far outside what would be expected from a normal distribution. Looking at the values for Range which is the distance from minimum to maximum, our estimates demonstrate that the Irish banking sector is less mature than the British.

	$\mu * 10^4$	$\delta * 10^4$	**Skew**	**Kurtosis**	**Min**	**Max**	**Z-Score (Min)**	**Z-Score (Max)**	**Range**	**Median**	**Mode**	**Sample Variance**	**Sample Size**
Standard Chartered	1.64	1.16	0.31	9	-0.075	0.114	-6.44	9.80	0.189	0	0	0.00013	3015
Northern Rock	-3.45	1.29	-2.33	65	-0.190	0.165	-14.74	12.82	0.355	-0.00025	0	0.00017	2307
HBOS	-4.22	1.44	-1.93	43	-0.233	0.118	-16.11	8.19	0.351	-0.00029	0	0.00021	2539
Barclays	-0.24	1.38	1.46	40	-0.124	0.239	-9.01	17.33	0.363	-0.00021	0	0.00019	3015
RBS	-2.58	1.66	-8.38	239	-0.476	0.132	-28.67	8.00	0.608	0	0	0.00028	3015
Lloyds	-2.67	1.43	-0.95	35	-0.180	0.177	-12.60	12.43	0.357	-0.00021	0	0.00020	3015
HSBC	0.49	0.83	-0.28	12	-0.090	0.063	-10.84	7.51	0.153	0	0	0.00007	3015
FTSE100	-0.03	0.57	-0.13	6	-0.040	0.041	-7.05	7.14	0.081	0.00017	N/A	0.00003	3015
Average	**-1.38**	**1.22**	**-1.53**	**56**	**-0.176**	**0.131**	**-13.18**	**10.40**	**0.307**	**-0.00010**	**0**	**0.00016**	
Anglo	-3.04	1.78	-2.59	61	-0.270	0.223	-15.13	12.52	0.492	0	0	0.00032	2527
BoI	-4.26	1.92	-1.52	48	-0.344	0.171	-17.93	8.91	0.515	0	0	0.00037	3003
AIB	-5.56	1.83	-3.01	74	-0.383	0.143	-20.94	7.84	0.526	0	0	0.00033	3003
ILP	-3.74	1.80	-1.58	49	-0.301	0.203	-16.69	11.29	0.504	0	0	0.00032	3003
ISEQ	-0.89	0.65	-0.58	8	-0.061	0.042	-9.35	6.54	0.103	0.00026	N/A	0.00004	3003
Average	**-3.50**	**1.60**	**-1.86**	**48**	**-0.272**	**0.156**	**-16.01**	**9.42**	**0.428**	**0.00005**	**0**	**0.00028**	

Table 4.1: Stylised Facts for Returns

4.3.2 Lagged Correlation for Returns

We can observe that prices are heavily correlated, in contrast the returns are significantly less correlated. This is in line with previous research. For Standard Chartered we can observe that it is anti-correlated for all five lags while Northern Rock is heavily positively correlated for four lags before reverting on the fifth.

A number of the banks in Britain inclusing HBOS, Barclays, RBS and Lloyds show elements of mean reversion however HBOS shows strong negative correlation for the second and third lags. HSBC and the FTSE100 meanwhile are mostly negatively correlated however both do show an element of reversion on the fourth lag.

In Ireland AIB and ILP are strongly positively correlated on the first lag however revert and become negatively correlated for the four remaining lags. Bank of Ireland shows consistent positive correlation before reverting for the fourth and fifth lags. Anglo Irish Bank moves in line with the ISEQ index starting positive and then reverting.

Estimates show that the first lag is fully positively correlated in Ireland and mostly positively correlated in Britain. The fourth lag is almost universally positively correlated in Britain however the opposite is true in Ireland which experiences mostly negative correlation. In contrast the fifth lag is almost universally negatively correlated in both Britain and Ireland.

In general the Irish market is more correlated than the British market however this is heavily influenced by AIB and ILP in Ireland and Standard Chartered in Britain. Mean reversion does take place with Barclays, Lloyds and to a lesser extent RBS, Bank of Ireland and the ISEQ Index.

	Correlation (Price)	**Correlation (Return)**	**L1**	**L2**	**L3**	**L4**	**L5**
Standard Cht	99.7%	-1.8%	-1.8%	-5.6%	-5.1%	-0.2%	-2.7%
Northern Rck	99.9%	13.9%	13.9%	2.6%	6.2%	3.5%	-3.8%
HBOS	99.7%	12.0%	12.0%	-21.1%	-11.9%	4.5%	9.0%
Barclays	99.7%	5.2%	5.2%	2.5%	-2.5%	1.2%	-6.5%
RBS	99.9%	11.3%	11.3%	-5.2%	6.0%	3.1%	-6.1%
Lloyds	99.9%	7.5%	7.5%	0.9%	-2.1%	1.5%	-9.2%
HSBC	99.2%	-1.6%	-1.6%	-6.0%	-3.2%	2.0%	-3.3%
FTSE100	99.7%	-5.5%	-5.5%	-5.2%	-7.8%	7.2%	-5.7%
Average	**99.7%**	**5.1%**	**5.1%**	**-4.6%**	**-2.6%**	**2.8%**	**-3.5%**
Anglo	99.9%	2.7%	2.7%	-5.9%	-5.4%	3.7%	-0.9%
BoI	99.9%	7.5%	7.5%	3.8%	4.6%	-4.2%	-6.6%
AIB	99.9%	21.8%	21.8%	-0.3%	-0.1%	-7.1%	-13.4%
ILP	99.9%	13.1%	13.1%	-0.3%	-5.0%	-7.5%	-11.0%
ISEQ	99.9%	5.0%	5.0%	-1.5%	-1.0%	0.6%	-2.0%
Average	**99.9%**	**10.0%**	**10.0%**	**-0.9%**	**-1.4%**	**-2.9%**	**-6.8%**

Table 4.2: Correlation and Lagged Correlation Table for Returns

4.3.3 Return and Clustering Histograms

Return histograms were calculated by taking the minimum and maximum value from each data set and computing a series of buckets between them. Each return value in the data set was then placed in a bucket, this gives us a distribution of our returns. A z-score for each bucket was then computed.

The clustering histogram looks at how often the return values move in the same direction, either positively or negatively. It is calculated by creating a series of buckets and placing each value in the data set into a bucket. Both the return and clustering histograms were calculated in Excel, the clustering histogram was calculated in excel using the following formula:

$$
\begin{aligned}
&= IF(\\
&\quad OR(\\
&\qquad AND(SIGN(x) = 0, SIGN(x-1) = 1), \\
&\qquad AND(SIGN(x) = 1, SIGN(x-1) = 0), \\
&\qquad IF(SIGN(x) = SIGN(x-1), TRUE, FALSE) \\
&\quad), \\
&\quad y + 1, 1 \\
&)
\end{aligned}
$$

The Efficient Market Hypothesis says that clustering should not occur in any significant manner. Evidence of clustering indicates that the markets are not functioning correctly. Below are histograms for the FTSE100 and the ISEQ Index. The histograms for all the banks are contained within the Appendices.

It can be seen from both tables below that there are a number of occasions where returns move in the same direction for an extended period of time. This is far outside what would be expected from a normal distribution and indeed these tables indicate a long tailed distribution with a number of return values outside the range predicted by standard models.

Table 4.3: FTSE100 Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.04023987	1	-7.161734262	0	0	0.0%
-0.03523987	1	-6.271780964	1	1570	50.5%
-0.03023987	1	-5.381827666	2	796	25.6%
-0.02523987	2	-4.491874368	3	384	12.3%
-0.02023987	15	-3.60192107	4	169	5.4%
-0.01523987	17	-2.711967772	5	92	3.0%
-0.01023987	81	-1.822014474	6	44	1.4%
-0.00523987	269	-0.932061176	7	26	0.8%
-0.00023987	1001	-0.042107878	8	12	0.4%
0.00476013	1296	0.847845421	9	8	0.3%
0.00976013	317	1.737798719	10	4	0.1%
0.01476013	79	2.627752017	11	3	0.1%
0.01976013	17	3.517705315	12	1	0.0%
0.02476013	8	4.407658613	13	1	0.0%
0.02976013	2	5.297611911	14	1	0.0%
0.03476013	2	6.187565209	15	0	0.0%
0.03976013	1	7.077518507	16	0	0.0%
0.04476013	1	7.967471805	17	0	0.0%

Table 4.4: ISEQ Index Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.060642994	1	-9.37652	0	0	0.0%
-0.055642994	0	-8.60229	1	1540	46.8%
-0.050642994	0	-7.82807	2	801	24.3%
-0.045642994	0	-7.05385	3	428	13.0%
-0.040642994	1	-6.27963	4	229	7.0%
-0.035642994	1	-5.5054	5	129	3.9%
-0.030642994	3	-4.73118	6	66	2.0%
-0.025642994	9	-3.95696	7	41	1.2%
-0.020642994	10	-3.18274	8	21	0.6%
-0.015642994	47	-2.40851	9	11	0.3%
-0.010642994	87	-1.63429	10	7	0.2%
-0.005642994	261	-0.86007	11	7	0.2%
-0.000642994	937	-0.08585	12	3	0.1%
0.004357006	1352	0.688377	13	3	0.1%
0.009357006	426	1.4626	14	1	0.0%
0.014357006	96	2.236823	15	1	0.0%
0.019357006	37	3.011045	16	1	0.0%
0.024357006	12	3.785268	17	1	0.0%
0.029357006	5	4.55949	18	0	0.0%
0.034357006	3	5.333713	19	0	0.0%
0.039357006	1	6.107936	20	0	0.0%
0.044357006	1	6.882158	21	0	0.0%

4.3.4 Stylised Facts for Sentiment

Looking at the following table we can observe that the mean is mostly negative but in all cases is extremely small. The standard deviation is extremely close to one which together with the mean, results in the z-score calculation for the minimum and maximum remaining unchanged for two places of decimal.

The value for z-score of minimum are approximately what would be expected from a normal distribution however the values for the z-score of maximum are outside the expected range, though not by as much as returns. Estimates show that sentiment is skewed slightly positive however the values here are small and the kurtosis is near normal. The sample sizes for the British banks are significantly higher than the Irish banks however the values from the Irish banks are all in line with their British counterparts. The minimum and maximum values and the consequent range values together with the skew and kurtosis values appear to indicate a random walk with an approximate normal distribution.

It can be observed that the range values for the failed banks are generally smaller than the same values for successful banks. In addition the kurtosis figures are generally lower. The exception to this is HBOS which has a greater range than Standard Chartered and who's kurtosis is closer to a normal distribution than all other banks except RBS.

	$\mu * 10^4$	$\delta * 10^4$	**Skew**	**Kurtosis**	**Min**	**Max**	**Z-Score (Min)**	**Z-Score (Max)**	**Range**	**Median**	**Mode**	**Sample Variance**	**Sample Size**
Standard Chartered	2.88E-07	10002.49	0.57	1.17	-2.75	4.36	-2.75	4.36	7.11	-0.051	-1.88	1.00050	2012
Northern Rock	-5.89E-08	10004.21	0.02	0.53	-2.59	3.80	-2.59	3.80	6.39	0.047	-2.59	1.00084	1188
HBOS	7.94E-08	10002.21	0.51	2.11	-2.37	6.02	-2.37	6.02	8.39	-0.032	-2.37	1.00044	2266
Barclays	-2.73E-08	10001.36	0.04	1.56	-3.29	6.00	-3.29	6.00	9.29	-0.001	-3.29	1.00027	3665
RBS	-5.68E-08	10004.06	0.35	2.13	-3.88	5.49	-3.88	5.49	9.37	-0.031	-0.40	1.00081	1233
Average	**4.49E-08**	**10002.87**	**0.30**	**1.50**	**-2.98**	**5.13**	**-2.98**	**5.13**	**8.11**	**-0.01**	**-2.11**	**1.00057**	
Anglo Irish Bank	-5.96E-08	10009.96	0.29	0.43	-2.16	3.45	-2.15	3.45	5.61	-0.033	-2.16	1.00199	503
Bank of Ireland	-1.38E-07	10006.88	0.44	1.84	-2.11	6.01	-2.11	6.00	8.12	0.017	-2.11	1.00138	727
Allied Irish Bank	-3.86E-08	10006.43	0.53	1.35	-2.06	5.58	-2.06	5.58	7.64	-0.063	-2.06	1.00129	778
Average	**3.77E-09**	**10004.50**	**0.34**	**1.40**	**-2.11**	**5.01**	**-2.11**	**5.01**	**7.12**	**-0.03**	**-2.11**	**1.00155**	

Table 4.5: Stylised Facts - Sentiment

4.3.5 Lagged Correlation for Sentiment

From the following table it can be seen that correlation is fully positive for the time period taken. It is heavier for those banks which suffered the worst damage including Northern Rock, HBOS and Anglo Irish Bank and in addition is generally heavier for Irish banks when compared with British banks. Sentiment for Northern Rock and Anglo Irish Bank remain consistently highly positive over the five lags while sentiment for Standard Chartered and Bank of Ireland is strong on the first lag but drops on remaining lags. We can see that on average the sentiment for the Irish banks is higher for every lag except lag 4.

	L1	**L2**	**L3**	**L4**	**L5**	**Sample Size**
Standard Chartered	20%	11%	3%	4%	6%	2012
Northern Rock	38%	26%	25%	24%	26%	1188
HBOS	20%	12%	7%	8%	11%	2266
Barclays	16%	11%	15%	11%	14%	3665
RBS	18%	6%	9%	11%	10%	1233
Average	**23%**	**13%**	**12%**	**12%**	**13%**	
Anglo Irish Bank	48%	28%	24%	20%	21%	503
Bank of Ireland	20%	6%	8%	8%	8%	727
Allied Irish Bank	22%	12%	7%	2%	9%	778
Average	**30%**	**15%**	**13%**	**10%**	**13%**	

Table 4.6: Lagged Correlation Table for Sentiment

4.4 Return Autoregression

Significant coefficients for the British banks test positive for the first lag and revert to negative for the second. The third lag is mostly negative however the fourth lag is mostly not significant. The fifth lag is almost always negative. For the Irish banks there is a corresponding positive correlation on the first lag and negative correlation on the second and fifth. The fourth lag in contrast is negative. (Table 4.8)

From the table we can see that the British banks Standard Chartered, Barclays, HSBC and Lloyds have several significant results however a low adjusted R^2 is also observed and therefore these models are not of high precision. In addition Anglo Irish Bank, Bank of Ireland and the ISEQ Index also have several significant coefficients however also display a low R^2 which is indicative that these are not good models. In contrast to this, Northern Rock, HBOS, RBS, the FTSE100, AIB and Irish Life and Permanent all possess coefficient results of high significance as well as a higher adjusted R^2. From this it can be concluded that these are good models.

In the following table, entries in ***bold and italics*** have a p-value of 0.01 or less. Entries in **bold** have a p-value greater than 0.01 but less than 0.05 and entries in *italics* have a p-value greater than 0.05 but less than 0.1. Also shown is the adjusted R^2 of each model which assists in determining validity of the model.

A low p-value (high significance) and high adjusted R^2 indicate that changes in predictors are related to changes in the response variable and the model produced explains the response variability. The presence of low adjusted R^2 indicates a relationship however the predictions are less precise when compared to a model with high adjusted R^2.

	1	2	3	4	5	R^2	Count	Start	Finish
Standard Chartered		***-0.06***	***-0.055***			0.006	3111	01/01/1999	03/12/2010
Northern Rock	***0.137***		***0.057***		**-0.051**	0.023	2381	01/01/1999	15/02/2008
HBOS	***0.129***	***-0.205***	***-0.062***		**0.039**	0.068	2621	01/01/1999	16/01/2009
Barclays	***0.047***				***-0.056***	0.005	3111	01/01/1999	03/12/2010
RBS	***0.122***	***-0.067***	***0.066***		***-0.058***	0.023	3111	01/01/1999	03/12/2010
Lloyds	***0.075***			*0.029*	*-0.099*	0.014	3111	01/01/1999	03/12/2010
HSBC		***-0.061***	**-0.040**		*-0.031*	0.005	3111	01/01/1999	03/12/2010
FTSE100	***-0.052***	***-0.061***	***-0.088***	***0.060***	***-0.059***	0.019	3111	01/01/1999	03/12/2010
Anglo Irish Bank	*0.034*	**-0.038**	***-0.051***		*-0.035*	0.005	2622	01/01/1999	19/01/2009
Bank of Ireland	***0.065***		**0.045**	*-0.030*	***-0.049***	0.008	3111	01/01/1999	03/12/2010
Allied Irish Bank	***0.217***	***-0.056***		**-0.036**	***-0.102***	0.059	3111	01/01/1999	03/12/2010
Irish Life & Permanent	***0.121***		**-0.46**	***-0.54***	***-0.88***	0.029	3110	04/01/1999	03/12/2010
ISEQ	**0.046**					0.001	3021	04/01/1999	03/12/2010

Table 4.7: Return Autoregression

4.5 Return and Sentiment: A Regression Analysis

The following table looks at the ability of sentiment to impact upon return. It shows the results for sentiment regressed against return for a subset of the banks. We can see there are very few statistically significant values in this table which would appear to indicate the small role sentiment plays in influencing returns.

Of the values computed, most are small with just over half are significant at the 10% level and the remainder significant at the 5% level. It can be observed that Standard Chartered, Barclays, RBS, Anglo Irish Bank and the ISEQ Index show no significance. Of the statistically significant results, estimations show HBOS with the highest Adjusted R^2, more than six times the value of the second highest result.

The following equation derived from the equation in Chapter 3 constitutes this VAR; here α represents the coefficients of 5 lags of return, S represents the coefficients of sentiment and ϵ the error term.

$$v_t = \alpha_0^1 + \alpha^1 \; L_t \; S_t + \epsilon_t$$

	Neg 1	**Neg 2**	**Neg 3**	**Neg 4**	**Neg 5**	**Adj R^2**	**Sample Size**
Standard Chartered						0.002	2012
Northern Rock				*0.0006*		0.011	1188
HBOS	**-0.008**					0.189	2266
Barclays						0.005	3665
RBS						0.030	1233
FTSE100	*0.0002*					0.018	3014
Anglo Irish Bank						0.014	503
Bank of Ireland	**-0.002**					0.011	727
Allied Irish Bank			*0.002*			0.025	778
ISEQ						-0.008	726

Table 4.8: Return against Sentiment

4.6 Return Autoregresion with Exogenous Sentiment

The below tables are a return autoregression with sentiment as an exogenous variable for the ISEQ and FTSE100 indices. FTSE is autoregressed twice, first with the same sample size as available to ISEQ to allow for comparison and later with a larger sample size in order to demonstrate the influence of pre-crisis market values in the data set.

The following tables show autoregression results in which sentiment is set as exogenous to the model, this means it is taken as an influence that comes from outside the model and is therefore unexplained by the model. This analysis is run on the ISEQ index and the FTSE 100 as these contain information on the relevant banks. Results for each of the lags of the coefficients are in basis points. A basis point is one percent of one percent and is a common unit of measure in finance to denote the change in a financial instrument.

For ISEQ we can see that the result for negative sentiment is highly significant. However none of the results for return are of statistical significance meaning they are by chance. The contribution by sentiment therefore, is very small. The result for Adjusted R^2 suggest the model is not of a high quality while the result for Durbin-Watson suggests little to no correlation. One can speculate that this is as a result of banks making up a greater proportion of the ISEQ Index.

For the FTSE100 our estimates show that all lags of the return are of high significance while negative sentiment is also significant. The result for Adjusted R^2 suggests the model is of good quality however the Durbin-Watson result suggests little to no correlation. There is only one instance of mean reversion and apart from this coefficients remain negative.

ISEQ Return (Sentiment Exogenous)

Start Date: 2008-02-01. End Date: 2010-12-03. Sample Size: 720

	Coefficient (BP)	Std. Error	t-ratio	p-value	Significance
const	−3.11 bp	0.0003	−0.80	0.41	.
ISEQReturn$_{t-1}$	147.29 bp	0.03	0.39	0.69	.
ISEQReturn$_{t-2}$	−427.51 bp	0.03	−1.14	0.25	.
ISEQReturn$_{t-3}$	−170.62 bp	0.03	−0.45	0.64	.
ISEQReturn$_{t-4}$	100.99 bp	0.03	0.27	0.78	.
ISEQReturn$_{t-5}$	−278.95 bp	0.03	−0.75	0.45	.
Negative	−17.89 bp	0.0005	−3.35	0.0008	***.

Mean dependent var	−0.0005	S.D. dependent var	0.01
Sum squared resid	0.07	S.E. of regression	0.01
R^2	0.019	Adjusted R^2	0.01
$F(6, 713)$	2.3	P-value(F)	0.03
$\hat{\rho}$	0.0005	Durbin–Watson	1.99

FTSE Return (Sentiment Exogenous)

Start Date: 2008-02-08. End Date: 2010-12-03. Sample Size: 713

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	18.63 bp	0.0003	4.83	0.000	***.
FTSEReturn$_{t-1}$	−797.38 bp	0.03	−2.17	0.03	**.
FTSEReturn$_{t-2}$	−1125.17 bp	0.03	−3.11	0.001	***.
FTSEReturn$_{t-3}$	−987.38 bp	0.03	−2.73	0.006	***.
FTSEReturn$_{t-4}$	1078.73 bp	0.03	2.98	0.002	***.
FTSEReturn$_{t-5}$	−1025.45 bp	0.03	−2.82	0.004	***.
Negativ	−32.25 bp	0.0004	−6.76	0.000	***.

Mean dependent var	2.25e–06	S.D. dependent var	0.007
Sum squared resid	0.036410	S.E. of regression	0.007
R^2	0.1	Adjusted R^2	0.09
$F(6, 706)$	13.49	P-value(F)	1.60e–14
$\hat{\rho}$	0.01	Durbin–Watson	1.97

Here FTSE is auto-regressed with sentiment set as exogenous for a larger sample size in order to observe the difference in results when regression is performed over a more protracted period of time. Estimates show that all lags of the return are of high significance while negative sentiment is also statistically significant. The result for Adjusted R^2 suggests the model is of good quality however the Durbin-Watson result suggests no correlation. There is only one instance of mean reversion and apart from this coefficients remain negative. This is broadly in line with the previous FTSE autoregression however the Adjusted R^2 and statistical significance of the negative sentiment is lower.

FTSE Return (Sentiment Exogenous)

Start Date: 1999-01-11. End Date: 2010-12-03. Sample Size: 3001

	Coefficient (BP)	Std. Error	t-ratio	p-value	Significance
const	−0.24 bp	0.0001	−0.24	0.8	
FTSEReturn_{t-1}	−575.07 bp	0.01	−3.15	0.001	***.
FTSEReturn_{t-2}	−627.26 bp	0.01	−3.44	0.0006	***.
FTSEReturn_{t-3}	−848.68 bp	0.01	−4.66	0.0000	***.
FTSEReturn_{t-4}	567.54 bp	0.01	3.11	0.001	***.
FTSEReturn_{t-5}	−591.57 bp	0.01	−3.24	0.001	***.
Negative	−2.18 bp	0.0001	−2.12	0.03	**.

Mean dependent var	−9.76e–06	S.D. dependent var	0.005
Sum squared resid	0.09	S.E. of regression	0.005
R^2	0.02	Adjusted R^2	0.01
$F(6, 3001)$	11.12	P-value(F)	2.65e–12
$\hat{\rho}$	−0.001	Durbin–Watson	2.002

4.7 Chart Observations

Included below are a number of charts and graphs upon which we can make observations.

From Northern Rock we can see the sporadic nature of sentiment and the difficulty encountered when gathering sentiment data. As Northern Rock is a small bank, previously a building society, and with an operating income of just 627 million in 2006, they are not mentioned as often in the Financial Times in comparison to the larger banks. From the graph we can see that Northern Rock are refered to infrequently between 1999 and 2002. Their largest number of mentions comes at the beginning of 2003 when they are performing successfully and have been promoted to the FTSE100.

After this they are mentioned a little more frequently demonstrating their higher status however it is not until 2007 when they are demoted from the FTSE100 and run into financial difficulties that they begin to appear very frequently in the news. We can compare this to Standard Chartered which had an operating income of $8620 million in 2006 and operates internationally. It is consistently mentioned more frequently. Their performance during and post the crisis means that visually they appear quite consistent. Barclays bank had an operating income of 21,656 million in 2006, 34.5 times greater than Standard Chartered, and being such a large bank with an international presence it is mentioned even more frequently.

For the Irish banks the sentiment discovered was less than the British banks. Sentiment on Anglo Irish Bank in the Financial Times was significantly reduced following its collapse while the sentiment numbers for Allied Irish Bank and Bank of Ireland actually increase in number. As can be observed sentiment detected tends towards negative however the severity of the negative sentiment actually decreases during and following 2009.

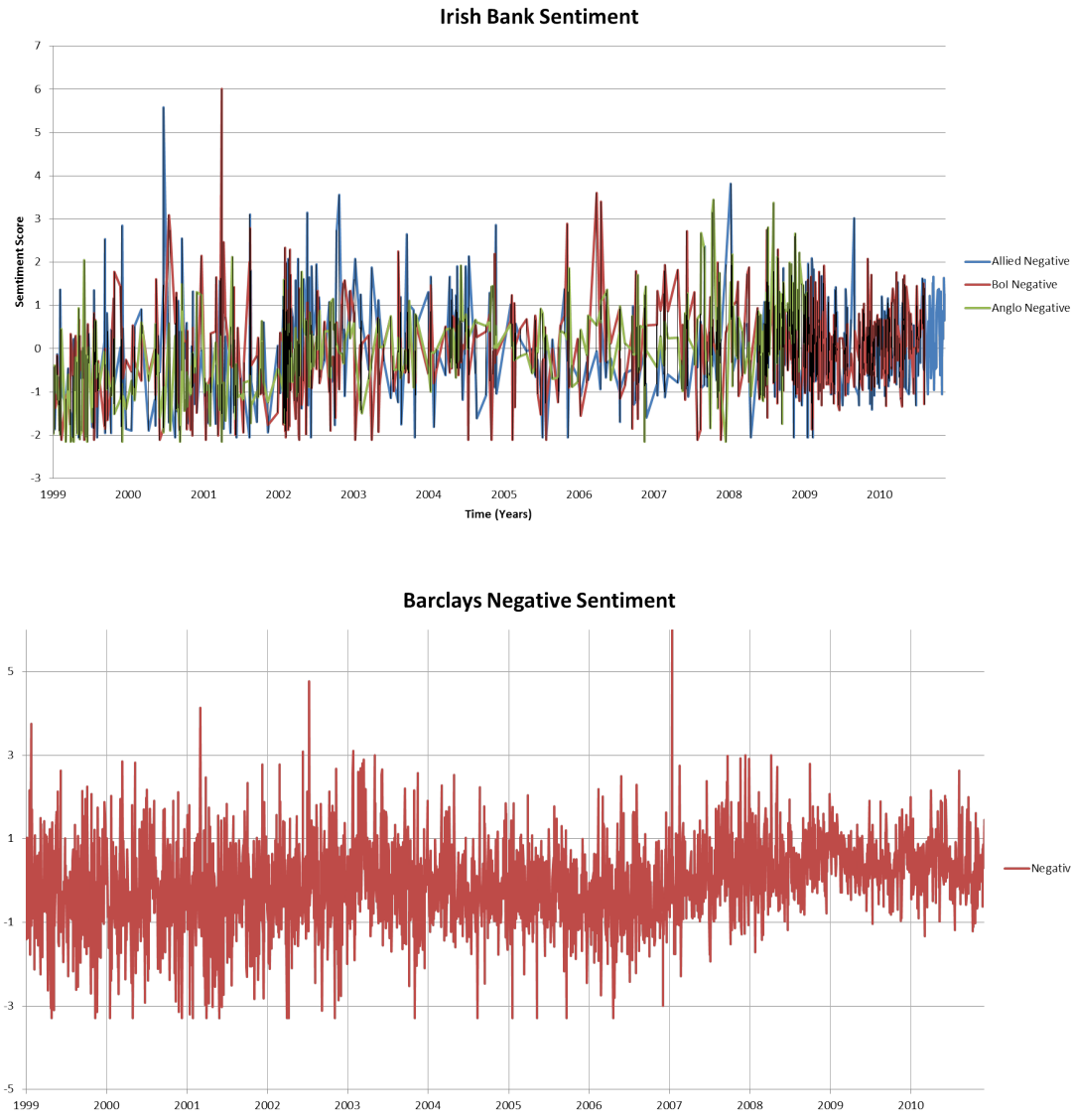


Figure 4.2: Sentiment for Barclays and Major Irish Banks

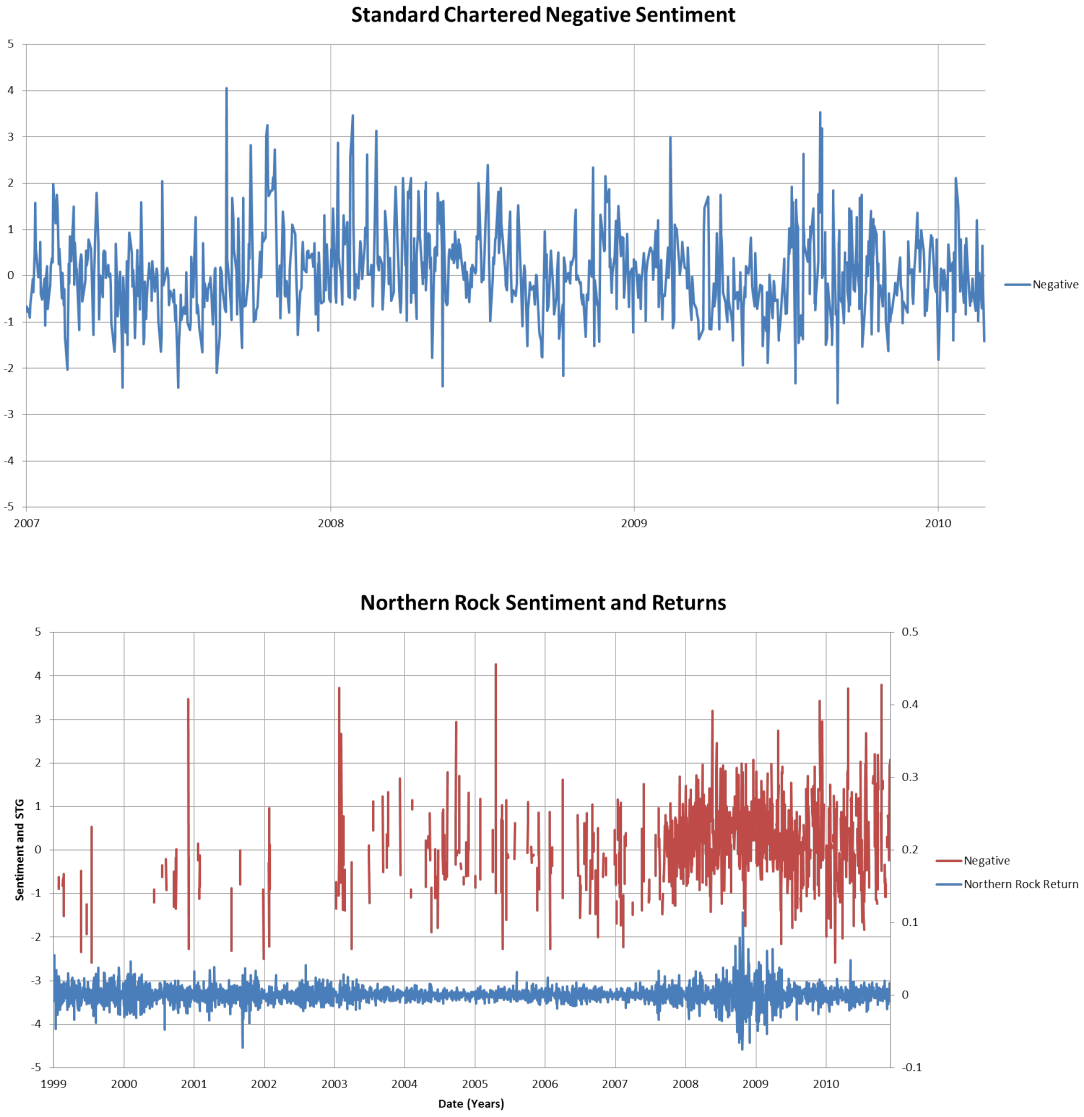


Figure 4.3: Sentiment for Standard Chartered and Northern Rock

4.8 Summary

This chapter presented the research findings of the work. Return and Sentiment time series data were first collated. Statistical properties called stylised facts were presented as well as correlations for each of the banks and the two indices for both return and sentiment.

Histograms for clustering and return were presented and regressions were performed for return, return against sentiment and return with sentiment as an exogenous variable. Finally a series of observations were made based on visual charts created from sentiment data obtained.

Chapter 5

Discussion of Results

5.1 Introduction

This chapter discusses the research findings within the context of the literature and define conclusions arising from the results. The discussion explores how the findings fit into the existing body of knowledge looking for consistency with current theories as well as new insights. The implications of the research findings for policy and practice may be explored and suggestions for further research given.

5.2 Findings and Implications

From our observations of the time series data streams we can say that financial Crisis present an anomaly in price return and sentiment series. In general volatility is low, we can see a period of stability before the crisis, so in times of normal market operation banks appear to be relatively safe and stable. However for the period of time analysed estimates show that every bank with the exception of one and indeed the ISEQ and FTSE100 markets themselves made a loss. The negative result from the markets however is smaller than the banks and so we can conclude that in general banks at this period of time were not worth investing in and any investments made over an extended period of time such as a decade can be wiped out in a much shorter period.

In the beginning of this work a null and alternative hypothesis was formed to be studied. The null hypotheses H_0 was that it is not possible to detect sentiment before a breakdown in a specific sector of the economy while the alternative hypotheses H_1 was that it is possible to detect sentiment before a breakdown in a specific sector of the economy. Evidence has emerged to support a rejection of the null hypothesis, sentiment was detected with significance and would explain a small fraction of the variance of the data however the models presenting this evidence are tenuous and so there is insufficient evidence to reject the null hypothesis.

Sentiment graphs show spikes in sentiment for the period 2000/2001 and 2007/2008 in line with the Global Tech Crash and recession and the more recent financial crisis. It is useful to note the increase in sentiment surrounding Irish Banks in the 2008-2010 era indicating a new amplified focus on Irish banks not previously seen. Interpretation of the data has to be made with this in mind. Media focus is tied to influence, the more influential a bank is in size and scope the more sentiment data there is to be found. This is intuitive however it implies that when looking for evidence of the influence of sentiment it is more desirable to collate and aggregate sentiment data and work with indices.

5.2.1 Stylised Facts and Correlation

The performance of Irish banks is noticeably worse than British banks, this can be attributed to the relative size and scope of British banks as well as their international reach. The exceptions are those banks in Britain which collapsed, these show much worse performance in line with their fortune. In general however the British banking sector appears more mature than the Irish banking sector. The application of linear models to these series is complex and requires careful consideration as estimates show that all banks and indices show a non-normal distribution.

From the graphs we can observe the stark differences between the correlation of prices and returns. Prices can be observed to be not independent and not sta-

tionary. The correlation of prices, returns and sentiment is in line with previous research. The lack of mean reversion and skew towards negative is indicative of the markets moving away from the historical averages of price and return and towards a loss period and is evidence against the assumption of mean reversion. The Irish market is more correlated but this appears influenced by a select few.

The histograms of return and clustering again show that the markets are not functioning correctly. Movements happen which are far outside what would be expected from a normal distribution and what can be modelled correctly with a linear model. This provides further evidence against the Efficient Market Hypothesis. Estimates show that the ISEQ index has clustered at a high degree than the FTSE100. Banks that performed worse during the crisis can be observed to have a greater propensity towards clustering.

In contrast to returns, stylised facts of sentiment estimates show a random walk with an approximate normal distribution. A smaller range was observed for failed banks when compared to surviving banks, this is indicative of the concentrated sentiment inspired by these controversial banks. Looking at sentiment correlation we can see that it is heaviest for those banks which performed worst however the results here for the Irish banks are likely influenced by the foreign nature of the data source. Any news that makes it into the British Financial Times regarding Irish institutions would have to do more to earn it's way there when compared with news regarding British banks and this correlation provides evidence for that.

5.2.2 Regression and Autoregression

Return: The Adjusted R^2 for a number of the models is low indicating that these models are not of high precision. A low value for Adjusted R^2 makes it difficult to draw any direct conclusions or derive any implications however at least some of the variance can be explained. The highest results are 0.068 (6.8%) and 0.059 (5.9%) for HBOS and Allied Irish Bank respectively. Based upon the

weakness of these models any conclusions drawn are tenuous however estimates show elements of mean reversion in both Irish and British banks and the markets appear to move together in unison on several lags. This may be indicative of the relationship between the markets or the mutual effect of some external source.

Return and Sentiment: Sentiment results with ISEQ Index shows no significance, this is likely to be because the subset of the banks examined only make up a small proportion of the index and the sample size obtained is also small. Sentiment results for the FTSE100 show slight significance for the first lag however the Adjusted R^2 of the model is low so the variance this explains is low. It is interesting to note the relative strength of the Adjusted R^2 for HBOS: 0.189 (indicating this model explains 18.9% of the variance) as well as the significance of the first lag. The model for each bank however must be regarded as tenuous because we cannot be certain that the occurrence of the name of the bank is just about that bank. Market indices inspire more confidence since it is likely that all pieces in the Financial Times have a degree of impact so even articles that are about multiple topics aside from the bank in question will feed into the index sentiment.

Return and Exogenous Sentiment: For FTSE, an f-test result and p-value of the magnitude observed suggest it unlikely that values predicted by these models are by chance. For ISEQ the f-test result is lower and p-value higher but still statistically significant. Looking at negative sentiment results for ISEQ, estimates show the coefficient results at less than 20 basis points which is in line with previous research by Tetlock (2007) as discussed in Chapter 2. The value for Adjusted R^2 is low, this model only explains 1% of the variation, however current negative sentiment is statistically significant at a 99.9% level. This indicates that sentiment plays a role but it's part is very small. This is in line with previous research suggesting variance explained by sentiment is at most 20 basis points. Looking at FTSE the basis points are high, this is likely as a result of the time period this sample is taken from as when examined over a longer period the basis points figure drops as can be seen in the second FTSE table. The contribution of sentiment is very small, it could be speculated that this is as a result of banks making up a greater proportion of the ISEQ Index.

5.3 Future Research

The impact of sentiment on returns has been estimated however the reverse, the impact of return on sentiment, has not been examined. Linear models appear to be insufficient for the purposes of market analyses and other modelling attempts could be made using OLS (Ordinary least Squares) or Weighted Least Squares.

While research has been limited to a select few banks, analysing additional financial institutions in both the UK and Ireland could help achieve better results. From here another possible angle is to expand outside of the banking sector to look at related areas in the markets. Obtaining supplementary reliable sources of sentiment data could help make the data obtained more robust while picking an equivalent source of Irish sentiment data to the Financial Times and either pulling Irish data from the Irish source, British from the British or both from both could help create a more uniform accumulation.

Previous research by Tetlock (2007) concluded that there is a link between volume and sentiment, as mentioned in Chapter 2 a change in sentiment precedes a change in volume. Expanding this work to factor in volume could be beneficial. Benefit could be obtained from building up a comprehensive collection of data and performing analysis over a number of time periods, including and excluding the crisis, both before and after.

Another possible avenue would be to include a number of foreign markets for example the German stock market which is another important market in Europe and which has a degree of influence over UK and Irish markets. This would be in the same vein as Zouaoui et al (2011) and Garcia (2013).

Finally the number of articles being created couldn't be analysed for sentiment but could be an indicator that an event is taking place in the market. Possible future research might include this variable as a factor.

5.4 Summary

An analysis was performed which looked at the banking sector of the UK and Irish markets. Elements of technical and sentiment analysis were performed and results obtained.

Where sentiment is present and is significant it's apparent influence on returns is small and in line with previous research. When problems arise in the markets linear models are insufficient for the purposes of analyses yet these are still widely used.

Estimates obtained indicate the presence and influence of sentiment in the movement of returns however due to the nebulous nature of the data underpinning the model there is not sufficient evidence available to reject the null hypothesis.

Chapter 6

Conclusion

The interpretation of exogenous and endogenous variables in GRETL is involved. Results appear to be in line with previous research indicating sentiments role in the market however certain models must be approached with scepticism due to the difficulty in obtaining accurate data. The market indices allow for greater interpretation because while banks make up only a small percentage, they have an influence. There influence can be seen in Ireland but is not as obvious in the UK. While estimates obtained indicated the presence and influence of sentiment in the movement of returns. Due to the tenuous nature of the data underpinning the model there is not sufficient evidence available to reject the null hypothesis.

Analysis indicates evidence against the Efficient Market Hypothesis and Mean Reversion in times of market crisis, in line with previous research. The implications are that in times of crisis traditional models fail to function correctly and attempts should be made to discover a correct model, this will be non linear.

The approach to proving or disproving the null hypotheses in this case is in the analytical method category, where a formal theory is developed and the results derived can be compared to empirical observations. Additional material is included in the appendices including the full results of return autoregression from which the Table 4.8 was constructed and the remaining return and clustering histograms from chapter 4

Appendix A - AutoRegression

Return vs Return

Standard Chartered Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	0.000168887	0.000203809	0.8287	0.4074		Mean dependent var	0.000142	S.D. dependent var	0.011388
Return_{t-1}	−0.0240674	0.0179541	−1.3405	0.1802		Sum squared resid	0.399537	S.E. of regression	0.011353
Return_{t-2}	−0.0641287	0.0178968	−3.5833	0.0003	***.	R^2	0.007798	Adjusted R^2	0.006197
Return_{t-3}	−0.0554394	0.0179116	−3.0952	0.0020	***.	$F(5, 3100)$	4.872470	P-value(F)	0.000192
Return_{t-4}	−0.00401987	0.0179016	−0.2246	0.8223		$\hat{\rho}$	−0.002100	Durbin–Watson	2.002057
Return_{t-5}	−0.0270907	0.0178967	−1.5137	0.1302					

Northern Rock: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	−0.000281144	0.000257395	−1.0923	0.2748		Mean dependent var	−0.000336	S.D. dependent var	0.012683
Return_{t-1}	0.137449	0.0205412	6.6914	0.0000	***.	Sum squared resid	0.372174	S.E. of regression	0.012531
Return_{t-2}	0.00185829	0.0207332	0.0896	0.9286		R^2	0.025753	Adjusted R^2	0.023697
Return_{t-3}	0.0573263	0.0207558	2.7619	0.0058	***.	$F(5, 2370)$	12.52952	P-value(F)	4.98e–12
Return_{t-4}	0.0259295	0.0208414	1.2441	0.2136		$\hat{\rho}$	0.000471	Durbin–Watson	1.996284
Return_{t-5}	−0.0515981	0.0206365	−2.5003	0.0125	**.				

HBOS: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	−0.000430180	0.000268889	−1.5998	0.1098		Mean dependent var	−0.000403	S.D. dependent var	0.014220
Return_{t-1}	0.129916	0.0195546	6.6438	0.0000	***.	Sum squared resid	0.491476	S.E. of regression	0.013722
Return_{t-2}	−0.205614	0.0197075	−10.4333	0.0000	***.	R^2	0.070590	Adjusted R^2	0.068810
Return_{t-3}	−0.0625904	0.0201436	−3.1072	0.0019	***.	$F(5, 2610)$	39.64678	P-value(F)	2.19e–39
Return_{t-4}	0.0288043	0.0197742	1.4567	0.1453		$\hat{\rho}$	0.002461	Durbin–Watson	1.994538
Return_{t-5}	0.0396954	0.0196227	2.0229	0.0432	**.				

Barclays: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	−3.55669e–005	0.000242619	−0.1466	0.8835		Mean dependent var	−0.000036	S.D. dependent var	0.013557
Return_{t-1}	0.0478515	0.0179338	2.6682	0.0077	***.	Sum squared resid	0.566766	S.E. of regression	0.013521
Return_{t-2}	0.0207650	0.0179414	1.1574	0.2472		R^2	0.006843	Adjusted R^2	0.005241
Return_{t-3}	−0.0249331	0.0179413	−1.3897	0.1647		$F(5, 3100)$	4.271747	P-value(F)	0.000712
Return_{t-4}	0.0210995	0.0179454	1.1758	0.2398		$\hat{\rho}$	−0.003893	Durbin–Watson	2.007469
Return_{t-5}	−0.0567837	0.0179285	−3.1672	0.0016	***.				

RBS: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	−0.000243745	0.000289654	−0.8415	0.4001		Mean dependent var	−0.000267	S.D. dependent var	0.016329
Return_{t-1}	0.122982	0.0179293	6.8593	0.0000	***.	Sum squared resid	0.806949	S.E. of regression	0.016134
Return_{t-2}	−0.0675571	0.0180498	−3.7428	0.0002	***.	R^2	0.025347	Adjusted R^2	0.023775
Return_{t-3}	0.0666037	0.0180574	3.6885	0.0002	***.	$F(5, 3100)$	16.12369	P-value(F)	1.00e–15
Return_{t-4}	0.0234759	0.0180582	1.3000	0.1937		$\hat{\rho}$	−0.003109	Durbin–Watson	2.006209
Return_{t-5}	−0.0588290	0.0179266	−3.2817	0.0010	***.				

Lloyds: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	−0.000271333	0.000250557	−1.0829	0.2789		Mean dependent var	−0.000267	S.D. dependent var	0.014054
Return_{t-1}	0.0754788	0.0178710	4.2235	0.0000	***.	Sum squared resid	0.603474	S.E. of regression	0.013952
Return_{t-2}	9.39502e–005	0.0179135	0.0052	0.9958		R^2	0.016014	Adjusted R^2	0.014427
Return_{t-3}	−0.0193931	0.0179190	−1.0823	0.2792		$F(5, 3100)$	10.09047	P-value(F)	1.34e–09
Return_{t-4}	0.0296388	0.0179217	1.6538	0.0983	*.	$\hat{\rho}$	−0.005750	Durbin–Watson	2.011410
Return_{t-5}	−0.0997517	0.0178801	−5.5789	0.0000	*.				

HSBC: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance				
const	4.52798e–005	0.000146523	0.3090	0.7573		Mean dependent var	0.000039	S.D. dependent var	0.008188
Return_{t-1}	−0.0182367	0.0179513	−1.0159	0.3098		Sum squared resid	0.206680	S.E. of regression	0.008165
Return_{t-2}	−0.0617480	0.0179063	−3.4484	0.0006	***.	R^2	0.007044	Adjusted R^2	0.005442
Return_{t-3}	−0.0403634	0.0179267	−2.2516	0.0244	**.	$F(5, 3100)$	4.398165	P-value(F)	0.000541
Return_{t-4}	0.0220475	0.0179005	1.2317	0.2182		$\hat{\rho}$	−0.002189	Durbin–Watson	2.004106
Return_{t-5}	−0.0311383	0.0179021	−1.7394	0.0821	*.				

FTSE100: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	−9.91307e–006	9.97728e–005	−0.0994	0.9209	
Return_{t-1}	−0.0522362	0.0179280	−2.9137	0.0036	***.
Return_{t-2}	−0.0610115	0.0179094	−3.4067	0.0007	***.
Return_{t-3}	−0.0882854	0.0178774	−4.9384	0.0000	***.
Return_{t-4}	0.0606147	0.0179151	3.3834	0.0007	***.
Return_{t-5}	−0.0597462	0.0179288	−3.3324	0.0009	***.

Mean dependent var	−8.40e–06	S.D. dependent var	0.005616
Sum squared resid	0.095848	S.E. of regression	0.005560
R^2	0.021312	Adjusted R^2	0.019733
$F(5, 3100)$	13.50105	P-value(F)	4.71e–13
$\hat{\rho}$	−0.002852	Durbin–Watson	2.005495

Anglo Irish Bank: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	−0.000312556	0.000341538	−0.9151	0.3602	
Return_{t-1}	0.0347032	0.0195565	1.7745	0.0761	*.
Return_{t-2}	−0.0386056	0.0195605	−1.9736	0.0485	**.
Return_{t-3}	−0.0518646	0.0195468	−2.6534	0.0080	***.
Return_{t-4}	0.0284229	0.0198604	1.4311	0.1525	
Return_{t-5}	−0.0357221	0.0198601	−1.7987	0.0722	*.

Mean dependent var	−0.000295	S.D. dependent var	0.017507
Sum squared resid	0.796087	S.E. of regression	0.017461
R^2	0.007070	Adjusted R^2	0.005169
$F(5, 2611)$	3.718483	P-value(F)	0.002343
$\hat{\rho}$	−0.001025	Durbin–Watson	2.001953

Bank Of Ireland: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	−0.000395747	0.000337782	−1.1716	0.2414	
Return_{t-1}	0.0652116	0.0179335	3.6363	0.0003	***.
Return_{t-2}	0.0227855	0.0179606	1.2686	0.2047	
Return_{t-3}	0.0451028	0.0179429	2.5137	0.0120	**.
Return_{t-4}	−0.0308592	0.0179570	−1.7185	0.0858	*.
Return_{t-5}	−0.0490741	0.0179628	−2.7320	0.0063	***.

Mean dependent var	−0.000417	S.D. dependent var	0.018888
Sum squared resid	1.096219	S.E. of regression	0.018805
R^2	0.010420	Adjusted R^2	0.008823
$F(5, 3100)$	6.528151	P-value(F)	4.76e–06
$\hat{\rho}$	−0.003051	Durbin–Watson	2.006056

Allied Irish Bank: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	−0.000519623	0.000313148	−1.6593	0.0971	*.
Return_{t-1}	0.217911	0.0178640	12.1983	0.0000	***.
Return_{t-2}	−0.0569905	0.0182706	−3.1193	0.0018	***.
Return_{t-3}	0.0219539	0.0182948	1.2000	0.2302	
Return_{t-4}	−0.0361887	0.0182700	−1.9808	0.0477	**.
Return_{t-5}	−0.102610	0.0178600	−5.7453	0.0000	***.

Mean dependent var	−0.000547	S.D. dependent var	0.017963
Sum squared resid	0.940923	S.E. of regression	0.017422
R^2	0.060832	Adjusted R^2	0.059317
$F(5, 3100)$	40.15869	P-value(F)	3.96e–40
$\hat{\rho}$	−0.009169	Durbin–Watson	2.018181

Irish Life and Permanent: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	−0.000417250	0.000313245	−1.3320	0.1830	
Return_{t-1}	0.121351	0.0178939	6.7817	0.0000	***.
Return_{t-2}	−0.0189132	0.0180065	−1.0504	0.2936	
Return_{t-3}	−0.0461377	0.0179884	−2.5649	0.0104	**.
Return_{t-4}	−0.0540937	0.0180358	−2.9992	0.0027	***.
Return_{t-5}	−0.0889255	0.0183246	−4.8528	0.0000	***.

Mean dependent var	−0.000377	S.D. dependent var	0.017697
Sum squared resid	0.942074	S.E. of regression	0.017435
R^2	0.030951	Adjusted R^2	0.029387
$F(5, 3099)$	19.79613	P-value(F)	1.79e–19
$\hat{\rho}$	−0.004732	Durbin–Watson	2.009164

ISEQ Index: Return

	Coefficient	Std. Error	t-ratio	p-value	Significance
const	−9.36957e–005	0.000117445	−0.7978	0.4251	
Return_{t-1}	0.0468825	0.0182210	2.5730	0.0101	**.
Return_{t-2}	−0.0101229	0.0182388	−0.5550	0.5789	
Return_{t-3}	−0.0137837	0.0182201	−0.7565	0.4494	
Return_{t-4}	0.0132099	0.0182163	0.7252	0.4684	
Return_{t-5}	−0.0262983	0.0181984	−1.4451	0.1485	

Mean dependent var	−0.000095	S.D. dependent var	0.006452
Sum squared resid	0.125090	S.E. of regression	0.006447
R^2	0.003235	Adjusted R^2	0.001579
$F(5, 3010)$	1.953594	P-value(F)	0.082420
$\hat{\rho}$	−0.001387	Durbin–Watson	2.002743

Appendix B - Clustering and Return Histograms

Table 1: Standard Chartered Return and Clustering Histograms

Return			Clustering		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.074568533	0	-6.54169	0	0	0.0%
-0.064568533	4	-5.66629	1	1496	48.1%
-0.054568533	2	-4.79089	2	787	25.3%
-0.044568533	7	-3.91548	3	412	13.2%
-0.034568533	12	-3.04008	4	214	6.9%
-0.024568533	39	-2.16468	5	113	3.6%
-0.014568533	121	-1.28927	6	51	1.6%
-0.004568533	675	-0.41387	7	26	0.8%
0.005431467	1495	0.46153	8	7	0.2%
0.015431467	551	1.33693	9	3	0.1%
0.025431467	137	2.21234	10	1	0.0%
0.035431467	44	3.08774	11	1	0.0%
0.045431467	13	3.96314	12	0	0.0%
0.055431467	4	4.83854	13	0	0.0%
0.065431467	4	5.71395	14	0	0.0%
0.075431467	1	6.58935	15	0	0.0%
0.085431467	1	7.46475	16	0	0.0%
0.095431467	0	8.34015	17	0	0.0%
0.105431467	0	9.21556	18	0	0.0%
0.115431467	1	10.09096	19	0	0.0%
0.125431467	0	10.96636	20	0	0.0%

Table 2: Northern Rock Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	Z Score	*Bin*	*Frequency*	Percentage
-0.190071497	1	-14.97580631	0	0	0.0%
-0.170071497	0	-13.39722385	1	1158	48.6%
-0.150071497	1	-11.81864139	2	608	25.5%
-0.130071497	2	-10.24005892	3	301	12.6%
-0.110071497	0	-8.66147646	4	165	6.9%
-0.090071497	2	-7.082893996	5	76	3.2%
-0.070071497	2	-5.504311533	6	35	1.5%
-0.050071497	4	-3.92572907	7	17	0.7%
-0.030071497	14	-2.347146607	8	10	0.4%
-0.010071497	189	-0.768564143	9	3	0.1%
0.009928503	1963	0.81001832	10	2	0.1%
0.029928503	182	2.388600783	11	1	0.0%
0.049928503	16	3.967183247	12	1	0.0%
0.069928503	1	5.54576571	13	1	0.0%
0.089928503	1	7.124348173	14	1	0.0%
0.109928503	1	8.702930637	15	1	0.0%
0.129928503	1	10.2815131	16	1	0.0%
0.149928503	0	11.86009556	17	0	0.0%
0.169928503	1	13.43867803	18	0	0.0%

Table 3: HBOS Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	Z Score	*Bin*	*Frequency*	Percentage
-0.233158191	1	-16.36959754	0	0	0.0%
-0.208158191	0	-14.61130836	1	1481	48.8%
-0.183158191	0	-12.85301918	2	774	25.5%
-0.158158191	0	-11.09473	3	395	13.0%
-0.133158191	1	-9.336440819	4	197	6.5%
-0.108158191	1	-7.578151638	5	100	3.3%
-0.083158191	7	-5.819862457	6	43	1.4%
-0.058158191	6	-4.061573276	7	23	0.8%
-0.033158191	18	-2.303284095	8	10	0.3%
-0.008158191	379	-0.544994914	9	5	0.2%
0.016841809	2077	1.213294267	10	3	0.1%
0.041841809	112	2.971583448	11	2	0.1%
0.066841809	9	4.729872629	12	1	0.0%
0.091841809	5	6.48816181	13	1	0.0%
0.116841809	4	8.246450991	14	0	0.0%
0.141841809	1	10.00474017	15	0	0.0%

Table 4: Lloyds Return and Clustering Histogram

Return Histogram			Clutering Histogram		
Bin	*Frequency*	Z Score	*Bin*	*Frequency*	*Percentage*
-0.180113775	1	-12.80380359	0	0	0.0%
-0.160113775	2	-11.38001056	1	1876	48.1%
-0.140113775	0	-9.956217526	2	979	25.1%
-0.120113775	0	-8.532424492	3	516	13.2%
-0.100113775	1	-7.108631459	4	265	6.8%
-0.080113775	1	-5.684838425	5	136	3.5%
-0.060113775	8	-4.261045391	6	66	1.7%
-0.040113775	14	-2.837252358	7	32	0.8%
-0.020113775	98	-1.413459324	8	14	0.4%
-0.000113775	1385	0.010333709	9	6	0.2%
0.019886225	1475	1.434126743	10	3	0.1%
0.039886225	99	2.857919776	11	2	0.1%
0.059886225	20	4.28171281	12	2	0.1%
0.079886225	3	5.705505844	13	0	0.0%
0.099886225	1	7.129298877	14	0	0.0%
0.119886225	1	8.553091911	15	0	0.0%
0.139886225	1	9.976884944	16	0	0.0%
0.159886225	0	11.40067798	17	0	0.0%
0.179886225	1	12.82447101	18	0	0.0%

Table 5: HSBC Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	Z Score	*Bin*	*Frequency*	*Percentage*
-0.090330392	1	-11.01335792	0	0	0.0%
-0.080330392	0	-9.794769275	1	1544	49.6%
-0.070330392	0	-8.576180627	2	760	24.4%
-0.060330392	2	-7.357591979	3	392	12.6%
-0.050330392	0	-6.139003332	4	189	6.1%
-0.040330392	1	-4.920414684	5	106	3.4%
-0.030330392	10	-3.701826036	6	58	1.9%
-0.020330392	38	-2.483237389	7	30	1.0%
-0.010330392	141	-1.264648741	8	16	0.5%
-0.000330392	1228	-0.046060093	9	9	0.3%
0.009669608	1444	1.172528554	10	4	0.1%
0.019669608	201	2.391117202	11	1	0.0%
0.029669608	28	3.60970585	12	1	0.0%
0.039669608	11	4.828294497	13	1	0.0%
0.049669608	3	6.046883145	14	0	0.0%
0.059669608	2	7.265471793	15	0	0.0%
0.069669608	1	8.48406044	16	0	0.0%
0.079669608	0	9.702649088	17	0	0.0%

Table 6: RBS Return and Clustering Histogram

Return Histogram			Clustering Histogram		
Bin	*Frequency*	Z Score	*Bin*	*Frequency*	*Percentage*
-0.475871486	1	-29.12534145	0	0	0.0%
-0.450871486	0	-27.59443291	1	1504	48.3%
-0.425871486	0	-26.06352438	2	794	25.5%
-0.400871486	0	-24.53261585	3	410	13.2%
-0.375871486	0	-23.00170731	4	202	6.5%
-0.350871486	0	-21.47079878	5	105	3.4%
-0.325871486	0	-19.93989025	6	55	1.8%
-0.300871486	0	-18.40898171	7	27	0.9%
-0.275871486	0	-16.87807318	8	8	0.3%
-0.250871486	0	-15.34716465	9	4	0.1%
-0.225871486	0	-13.81625611	10	1	0.0%
-0.200871486	1	-12.28534758	11	1	0.0%
-0.175871486	0	-10.75443905	12	0	0.0%
-0.150871486	0	-9.223530513	13	0	0.0%
-0.125871486	1	-7.692621979	14	0	0.0%
-0.100871486	0	-6.161713446	15	0	0.0%
-0.075871486	3	-4.630804913	16	0	0.0%
-0.050871486	12	-3.099896379	17	0	0.0%
-0.025871486	59	-1.568987846	18	0	0.0%
-0.000871486	1297	-0.038079312	19	0	0.0%
0.024128514	1652	1.492829221	20	0	0.0%
0.049128514	69	3.023737755	21	0	0.0%
0.074128514	9	4.554646288	22	0	0.0%
0.099128514	5	6.085554821	23	0	0.0%
0.124128514	1	7.616463355	24	0	0.0%
0.149128514	1	9.147371888	25	0	0.0%

Table 7: Barclays Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	Z Score	*Bin*	*Frequency*	Percentage
-0.124050338	1	-9.147525694	0	0	0.0%
-0.099050338	1	-7.303676055	1	1512	48.6%
-0.074050338	2	-5.459826416	2	800	25.7%
-0.049050338	11	-3.615976777	3	407	13.1%
-0.024050338	70	-1.772127138	4	185	5.9%
0.000949662	1682	0.071722501	5	104	3.3%
0.025949662	1263	1.91557214	6	53	1.7%
0.050949662	68	3.759421779	7	27	0.9%
0.075949662	9	5.603271418	8	14	0.5%
0.100949662	2	7.447121057	9	6	0.2%
0.125949662	1	9.290970696	10	3	0.1%
0.150949662	0	11.13482034	11	0	0.0%
0.175949662	0	12.97866997	12	0	0.0%
0.200949662	0	14.82251961	13	0	0.0%
0.225949662	0	16.66636925	14	0	0.0%
0.250949662	1	18.51021889	15	0	0.0%

Table 8: Anglo Irish Bank Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.269715933	1	-15.395394	0	0	0.0%
-0.244715933	0	-13.96688332	1	1213	46.3%
-0.219715933	1	-12.53837264	2	628	24.0%
-0.194715933	0	-11.10986196	3	350	13.3%
-0.169715933	1	-9.681351276	4	186	7.1%
-0.144715933	3	-8.252840594	5	105	4.0%
-0.119715933	2	-6.824329912	6	58	2.2%
-0.094715933	6	-5.39581923	7	36	1.4%
-0.069715933	6	-3.967308548	8	20	0.8%
-0.044715933	10	-2.538797866	9	13	0.5%
-0.019715933	64	-1.110287184	10	7	0.3%
0.005284067	1961	0.318223498	11	6	0.2%
0.030284067	522	1.74673418	12	0	0.0%
0.055284067	29	3.175244861	13	0	0.0%
0.080284067	9	4.603755543	14	0	0.0%
0.105284067	3	6.032266225	15	0	0.0%
0.130284067	2	7.460776907	16	0	0.0%
0.155284067	0	8.889287589	17	0	0.0%
0.180284067	1	10.31779827	18	0	0.0%
0.205284067	0	11.74630895	19	0	0.0%
0.230284067	1	13.17481963	20	0	0.0%

Table 9: Bank of Ireland Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.344447367	1	-18.21671554	0	0	0.0%
-0.319447367	0	-16.89296842	1	1438	46.2%
-0.294447367	0	-15.5692213	2	792	25.5%
-0.269447367	0	-14.24547418	3	439	14.1%
-0.244447367	0	-12.92172706	4	220	7.1%
-0.219447367	0	-11.59797994	5	114	3.7%
-0.194447367	0	-10.27423282	6	53	1.7%
-0.169447367	0	-8.950485698	7	28	0.9%
-0.144447367	1	-7.626738578	8	12	0.4%
-0.119447367	2	-6.302991459	9	7	0.2%
-0.094447367	4	-4.979244339	10	4	0.1%
-0.069447367	12	-3.655497219	11	2	0.1%
-0.044447367	29	-2.3317501	12	2	0.1%
-0.019447367	152	-1.00800298	13	0	0.0%
0.005552633	2200	0.31574414	14	0	0.0%
0.030552633	614	1.639491259	15	0	0.0%
0.055552633	63	2.963238379	16	0	0.0%
0.080552633	17	4.286985499	17	0	0.0%
0.105552633	9	5.610732618	18	0	0.0%
0.130552633	2	6.934479738	19	0	0.0%
0.155552633	4	8.258226858	20	0	0.0%
0.180552633	1	9.581973977	21	0	0.0%

Table 10: AIB Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.383216752	1	-21.30811984	0	0	0.0%
-0.358216752	0	-19.916108	1	1445	46.4%
-0.333216752	0	-18.52409615	2	777	25.0%
-0.308216752	0	-17.13208431	3	423	13.6%
-0.283216752	0	-15.74007247	4	227	7.3%
-0.258216752	0	-14.34806063	5	117	3.8%
-0.233216752	0	-12.95604878	6	67	2.2%
-0.208216752	0	-11.56403694	7	28	0.9%
-0.183216752	0	-10.1720251	8	15	0.5%
-0.158216752	0	-8.780013256	9	5	0.2%
-0.133216752	1	-7.388001414	10	2	0.1%
-0.108216752	2	-5.995989571	11	2	0.1%
-0.083216752	10	-4.603977729	12	2	0.1%
-0.058216752	16	-3.211965886	13	1	0.0%
-0.033216752	51	-1.819954044	14	0	0.0%
-0.008216752	455	-0.427942202	15	0	0.0%
0.016783248	2392	0.964069641	16	0	0.0%
0.041783248	131	2.356081483	17	0	0.0%
0.066783248	31	3.748093326	18	0	0.0%
0.091783248	14	5.140105168	19	0	0.0%
0.116783248	4	6.53211701	20	0	0.0%
0.141783248	2	7.924128853	21	0	0.0%
0.166783248	1	9.316140695	22	0	0.0%

Table 11: ILP Return and Clustering Histograms

Return Histogram			Clustering Histogram		
Bin	*Frequency*	*Z Score*	*Bin*	*Frequency*	*Percentage*
-0.301029996	1	-16.9968525	0	0	0.0%
-0.276029996	0	-15.58359778	1	1412	45.4%
-0.251029996	0	-14.17034306	2	737	23.7%
-0.226029996	0	-12.75708835	3	413	13.3%
-0.201029996	0	-11.34383363	4	245	7.9%
-0.176029996	0	-9.930578915	5	138	4.4%
-0.151029996	2	-8.517324199	6	68	2.2%
-0.126029996	3	-7.104069482	7	42	1.4%
-0.101029996	3	-5.690814766	8	19	0.6%
-0.076029996	6	-4.27756005	9	10	0.3%
-0.051029996	17	-2.864305334	10	7	0.2%
-0.026029996	80	-1.451050617	11	5	0.2%
-0.001029996	1206	-0.037795901	12	3	0.1%
0.023970004	1672	1.375458815	13	3	0.1%
0.048970004	92	2.788713531	14	2	0.1%
0.073970004	15	4.201968248	15	2	0.1%
0.098970004	8	5.615222964	16	2	0.1%
0.123970004	1	7.02847768	17	2	0.1%
0.148970004	1	8.441732396	18	0	0.0%
0.173970004	2	9.854987112	19	0	0.0%
0.198970004	0	11.26824183	20	0	0.0%
0.223970004	1	12.68149654	21	0	0.0%
0.248970004	0	14.09475126	22	0	0.0%

References

Geert Bekaert Andrew Ang. Stock return predictability: Is it there? *The Review of Financial Studies /*, 20(3):651707, 2007.

M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock return. *Journal of Finance*, 61(4):1645–1680, 2006.

M. Baker and J. Wurgler. Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21:129–151, 2007.

Yuan Baker, Wurgler. Global, local and contagious investor sentiment. *Journal of Financial Economics*, 104:pp272–287, may 2012.

Barclays Bank. Barclays bank annual report, 2006a.

Standard Chartered Bank. Standard chartered annual report, 2006b.

N. Shleifer Barberis, N. and R. Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49:307–343, 1998.

Bredin D Bathia D. An examination of investor sentiment effect on g7 stock market returns. *European Journal of Finance*, 19(9):909–937, 2013.

Mohamad. Zouaoui Mohamed Beer, Francisca. Watfa. Do investors care about noise trader risk? *The Capco Institute Journal of Financial Transformation*, 35, sep 2012.

Fischer Black. Noise. *Journal of Finance*, 41(3):529–543, 1986.

Ronald Braeutigam. *Microeconomics*. Wiley, fourth edition, 2010.

Roger. Casella, George. Berger. *Statistical Inference*. Duxbury Press, second edition, jun 2001.

Julien Chevallier. Global imbalances, cross-market linkages, and the financial crisis: A multivariate markov-switching analysis. *Economic Modelling*, 29(3): 943–973, 2012.

Fratzscher Chudik. Identifying the global transmission of the 20072009 financial crisis in a gvar model. *European Economic Review*, 55:325339, 2011.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223236, 2001.

A. Shleifer L. H. Summers De Long, J. B. and R. J. Waldmann. Noise trader risk in financial markets. *Journal of Political Economy*, 98(4):703–738, 1990.

J. Bradford DeLong and Andrei Shleifer. The stock market bubble of 1929: Evidence from closed-end funds. *Journal of Economic History*, 52(3):675–700, Sept 1991.

Trinity College Dublin. Tcd good research practice. Online, www.tcd.ie/about/policies/assets/pdf/TCDGoodResearchPractice.pdf, 2009.

Anna Zalewska Elena Ferrera, Julie Salaberb. Consumer confidence indices and stock markets meltdowns. *European Journal of Finance*, pages 1–26, oct 2014.

Zouaoui et al. How does investor sentiment affect stock market crises? evidence from panel data. *The Financial Review*, 46:723747, 2011.

J. Pennings Hoffmann A. O., T. Post. Individual investor perceptions and behavior during the financial crisis. *Journal of Banking and Finance*, 37:60–74, 2013.

McInnis Hribar. Investor sentiment and analysts earnings forecast errors. *Management Science*, 58(2):293307, feb 2012.

C. X. Jiang Lee, W. Y. and D. C. Indro. Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking and Finance*, 26:2277–2299, 2002.

Brooks R Mun M. The roles of news and volatility in stock market correlations during the global financial crisis. *Emerging Markets Review*, 13:1–7, 2012.

L. T. Orlowski. Financial crisis and extreme market risks/ evidence from europe. *Review of Financial Economics*, 21:120–130, 2012.

S.J. Taylor S. Poon. Stock returns and volatility: an empirical study of the u.k. stock market. *Journal of Banking and Finance*, 16:37–59, 1992.

M Schmeling. Investor sentiment and stock returns: Some international evidence. *Journal of Empirical Finance*, 16(3):394408, 2009.

Robert Schwartz and Reto Francioni. *Equity Markets in Action: The Fundamentals of Liquidity, Market Structure and Trading*. Wiley, 2004.

Robert J. Shiller. *Irrational Exuberance*. Princeton University Press, 2000.

Jerold B. Warner Stephen J. Brown. Using daily stock returns: The case of event studies. *Journal of Financial Economics*, 14(1):3–31, 1985.

S.J. Taylor. *Asset Price Dynamics, Volatility and Prediction*. Princeton University Press, 2005.

S.J. Taylor. *Modelling Financial Time Series. 2nd Edition*. World Scientific Publishing, 2008.

P.C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62:11391168, 2007.

Baguley Thom. *Serious Stats, Advanced Statistics.* Palgrave Macmillan, jun 2012.

John D. Turner. *Banking in Crisis: The Rise and Fall of British Banking Stability, 1800 to the Present*. Cambridge University Press, 2014.

Helmut Wagner. The causes of the recent financial crisis and the role of central banks in avoiding the next one. *International Economics and Economic Policy*, 7(1):63–82, may 2010.

Marvin V. Zelkowitz and Dolores R. Wallace. Experimental models for validating technology. *Computer*, 31(5):23–31, 1998.

Jiahui Zivot, Eric. Wang. *Modeling Financial Time Series with S-PLUS*. Springer New York, 2006. pp 385-429, Vector Autoregressive Models for Multivariate Time Series. ISBN 978-0-387-27965-7.