University of Strathclyde

Department of Economics

Three Essays on Information, Volatility, and Crises in Equity Markets

Shane K. Clark

This thesis is presented in fulfilment of the requirements for the degree of Doctor of Philosophy

2015

Acknowledgements

The completed work of this thesis would not have been possible without the help and support of many individuals. First and foremost, I would like to thank Julia Darby for her continued support and patience throughout the past 18 months of study at Strathclyde. I also thank Professor Peter McGregor has provided continued support and opportunities for me to conduct and present this research effectively.

A number of individuals have provided helpful comments and feedback on this research. I thank, Ian Marsh and Alex Dickson, for their helpful comments.

I would also like to thank my colleagues and friends at Strathclyde and other Scottish universities: Chris, Harry, Fraser, Nikos, Patrizio, Colin, and many others. Their support was paramount.

Finally, I would like to thank my family, and in particular, Marie Tamba, my wife and colleague. Without her I would never have attempted or finished this adventure that the PhD has been. My brother Brendan, I am deeply grateful for his guidance and sustained interest in my research. Finally, I would like to express my extreme gratitude to my parents for their support and patience over the years.

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Shane K. Clark Date: 26/06/2015

Table of Contents

Table of Figures

Table of Tables

Thesis Overview

This thesis consists of three essays, which examine the behaviour of stock market indices in light of the recent crises, in relation to volatility, information and sentiment.

Essay 1 focuses on the link between the flow of public information and stock market volatility in the FTSE, DJIA and the S&P 500 indices. This essay builds on existing literature in several ways. First, a new proxy for the daily public information flow is created to encompass a wider range of publication than previous contributions. It is also disaggregated by media type. This proxy is tested as an explanatory variable in index return volatility. Through the use of augmented GARCH models, the essay explores (a) whether information flow is a significant explanatory variable in volatility persistence, (b) whether the type of media the information flow is carried in impacts the volatility-information relation, (c) whether information backlog is incorporated into volatility when market re-open, and (d) whether there is a lag in the way information is incorporated into volatility (including by media type). Results show conclusive evidence of a strong relation between information and volatility. Further, the media types which deliver the most current information, i.e. *Wire-Feeds*, tend to have the largest impact on returns volatility. The evidence on the role of backlog and information lags in volatility is mixed, with significance only in the S&P 500 for Backlog, and only in S&P 500 and DJIA for lags.

Essay 2 further examines the stock market volatility and public information relation in light of daily market anomalies. Building on the literature on the Day of Week Effect, this essay investigates whether a Day of the Week effect is present in US and UK indices both before and during the latest financial crisis. Augmented GARCH models are constructed to test for a Day of the Week Effect in either returns and/or volatility, and are further explored in light of the information arrival proxy created in Essay 1. Results show weak evidence of Day of the Week effect in the data, except for a Friday Effect in the Russell 2000, both before and during the crisis; which confirm that small size stocks are more likely to exhibit Day of the Week effects, as suggested by previous uses of equal-weighted indices. Through the interaction variables, results show that information flow can satisfactorily explain Day of the Week effects.

Essay 3 investigates the relation between proxies for investor sentiment and stock market crises and recoveries on international indices. Using an Early-Warning-System (EWS) model, the essay examines whether investor sentiment is a useful predictor for the occurrence of stock market crises and early signs of recovery. Three alternative proxies are used to measure investor sentiment, including previously cited measures of stock market riskiness, investors' risk aversion and investors' optimism about stock markets. The results show that investor sentiment is overall a significant predictor of the occurrence of crises within a one year period, and that the addition of sentiment into early warning signal models of stock market crises can improve the predictive performance of the model (increases in investor sentiment increase the probability of occurrence of a crisis, which is in line with previous contributions finding a negative lead-lag relation between sentiment and stock returns). The extension of the model to early signs of recoveries also shows that sentiment is a reliable predictor. The measure of stock market riskiness (Baker and Wurgler, 2006) is found to be a better predictor than the Volatility Index (VIX) and the Put-to-Call Ratio (PCR). The cross-country comparison results confirms the literature findings that the link between sentiment and stock market returns varies across indices and cultures, as the predictive power of the variable appears strongest in the French and U.S. indices.

1 Essay 1: Public Information Arrival and Stock Market Volatility

1.1 Introduction

The flow of information has long been established as a key driving force behind security price changes. In recent times, financial research has focused primarily on the relation between stock market volatility and the flow of information arrival. Prompted by early empirical observations that stock return variance tend to exhibit differences between trading and non-trading hours (French and Roll, 1986), this research has attempted to establish this relation using a variety of information flow proxies, and increasingly sophisticated models. This essay builds on this literature by explicitly modelling information arrival and its impact on stock market volatility, in an augmented GARCH model of index returns.

The first innovation is in the dataset. The arrival of new publicly available information is captured via the ProQuest database search engine. To my knowledge, this is the first time that data from the ProQuest search engine has been used to provide a count of the aggregate number of publications per day, although as will be explained, the methodology used to relate the publication count to the volatility of stock returns draws on the existing literature. The ProQuest-based daily publication count is used to proxy the inflow of new information in the stock market. In addition, the daily publication count is disaggregated according to the medium by which information is delivered i.e. Magazines, Wire-Feeds, Newspapers, etc. Having generated several proxy indicators for information flow and examined the data, the ability of the constructed variables to capture movements in the volatility of stock returns is systematically investigated in augmented GARCH models. Three stock market indices are used in the essay: the Financial Times Stock Exchange (FTSE) for the UK, and the Standard and Poor's 500

(S&P 500) and Dow Jones Industrial Average (DJIA) for the US. These series were chosen since they are among the most widely used indices in the literature.

The chosen methodology allows for several relevant hypotheses to be tested. First, by explicitly capturing the total flow of public information, the new information flow proxy can be used to further test the well-established relation between information flow and volatility, i.e. whether increases in the flow of public information help to explain spikes in stock market volatility.

A second question is motivated by previous research that has stressed a distinction between the impact of arrival of public information during trading hours and the likely impact of information that accumulated when the markets are closed (e.g. over weekends and bank holidays). The premise is that a backlog of information builds up when markets are closed and this can only be incorporated into stock returns when markets reopen. Therefore, we might expect the volatility of stock returns to spike when the market reopens. In order to test the validity of this expectation, a "backlog" proxy is constructed to take into account the accumulated arrival of information published while markets are closed, and to test whether indeed this is incorporated in stock returns when markets reopen. Third, through varying the lag structure applied to the daily information flow variable, it is possible to test whether the relation is contemporaneous – i.e. newly published information impacts on the volatility of stock returns on the same day, or whether there might be significant delay in this impact.

Finally, the essay also builds on previous arguments in the literature that different types of information content can have varying impacts on volatility, by examining whether the media type in which information is captured (Wire-Feeds, Newspapers, Magazines etc.) can contribute in explaining these differences. Five proxies for information flow disaggregated by media type are used in order to shed light on whether different types of publication have varying impacts on the volatility of stock returns. The differences in media types are further examined with respect to the timing of their impact on volatility (lags and backlogs investigation). By answering these four major questions, this essay contributes to improving the understanding of how information flow impacts volatility of stock returns.

The remainder of the essay is organized as follows. Section 1.2 briefly introduces the notion of volatility in the stock market, both its definition and common properties identified in financial research of stock market returns, i.e. volatility clustering. This is followed by a discussion of some key aspects of the development of generalised autoregressive conditional heteroskedasticity (GARCH) models and presents the results of the basic GARCH(1,1) model in the context of the three indices of this essay. Section 1.3 provides a review of the literature that has related information flow to stock market volatility, and identifies some outstanding questions which this essay helps to address. Section 1.4 presents the information flow data used in the essay, provides details of the construction of the various proxies for the daily inflow of public information, and discusses the properties of constructed proxy variables. Section 1.5 outlines the estimation methods and explains how the augmented GARCH model will be used to systematically assess the validity of three major hypotheses. The results are presented and discussed in Section 1.6 and finally, Section 1.7 summarises the key conclusions and offers some suggestions for future research in this area.

1.2 Volatility in the Stock Market

1.2.1 Definition and Properties

Volatility is a phenomenon of particular interest in financial markets and academic research. According to the New Palgrave Dictionary of Economics, volatility is defined as: “a measure of asset price variability over some period of time, which typically describes the standard deviation of an investment's return in a particular context.”[1] Volatility is thus often simply defined as the variance of stock returns, while it is important to note that the size of the measurement period is an intrinsic part of the definition.

The concept is primarily associated with financial risk, i.e. the higher the variance in returns, the greater levels of risk; it is of major importance for finance practitioners as a central concept in asset pricing models and portfolio investment and risk management. It has also been a primary focus in financial academic research.

Empirical studies of financial time series have revealed a number of properties which are common in stock return volatility, e.g. heavy tails with positive excessive kurtosis, absence of linear autocorrelation of returns, volatility clustering (see for example, Cont, 2001). Of all of these, volatility clustering has received the most attention in the literature. In simple terms, it is defined by Mandelbrot (1963) as “large changes tend to be followed by large changes of either sign and small changes tend to be followed by small changes” (p.418). This phenomenon is generally present in high frequency security price data (e.g. daily returns). Volatility clustering is a relatively straightforward concept to grasp graphically when looking at returns over time, which

[1] Partial definition of volatility from the New Palgrave Dictionary of Economics.

show sustained periods of high or low volatility. This volatility clustering is illustrated in Section 2.3 for all three indices used in this essay.

1.2.2 ARCH and GARCH Models

Building on empirical observations, the above properties of stock market volatility, and in particular volatility clustering, have led to the development of improved models of stock returns, which aim to explicitly account for time-varying nature of volatility.

By far, the most prevalent models are built upon the AutoRegressive Conditional Heteroskedasticity (ARCH) model proposed by Engle (1982), and later generalized by Bollerslev (1986) in the GARCH model. The popularity of the ARCH and GARCH models is explained by their ability to model time-varying volatility and volatility clustering (Jondeau et al., 2007). A large number of extensions to the GARCH model have emerged, as well as a huge literature on their empirical applications (for reviews, see Bollerslev et al., 1992; Bollerslev et al., 1994; Li et al., 2002), but the original GARCH model is still widely used and an adequate model used for analysing volatility in financial times series (Kalev et al., 2004).

The GARCH (1,1) model is as follows:

$$r_t = \omega + \mu r_{t-1} + \varepsilon_t \tag{1}$$

$$\varepsilon_t | \, \Omega_{t-1} \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 \tag{2}$$

where r is the return, ω is a constant, ε_t is the serially uncorrelated error terms with a mean of zero and a conditional variance of σ_t^2. In Eq. (2), explicitly modelling the conditional variance of returns, the coefficients β and γ represent the persistence (or clustering) of returns volatility, i.e. $\beta + \gamma$ embodies the extent to which volatility is determined by its past level.

1.2.3 Volatility in UK and US Stock Returns

The stock returns data used in the analysis of this paper cover two countries and three indices: the Financial Times Stock Exchange (FTSE, UK), the Standard & Poor's 500 (S&P 500, US) and the Dow Jones Industrial Average (DJIA, US). The essay uses daily data over a five year sample period from January 1, 2005 through December 31, 2010[2]. The period selected corresponds to a heightened period of volatility occurring during the subprime debt crises. Although a longer time frame would present a more detailed analysis of whether the relation between information flow and volatility holds over a longer time frame, the construction of the information proxy is a highly manual and time consuming task. Therefore, a five year sample period is selected which covers recent periods of heightened volatility.

1.2.3.1 Graphical Examination and Descriptive Statistics

Figure 1 to Figure 3 below represent the FTSE, DJIA, and S&P 500 returns respectively, constructed as the daily percentage change in price for each index (for example a figure of 0.01 represents a 1% increase). All three indices show similar patterns, with sustained periods of high volatility (particularly during the financial crisis) and periods of persistent lower volatility (e.g. 2006-2007). While evidence of volatility clustering is by no means new or unexpected, these graphs present the first documented evidence of the presence of volatility clustering in the data used for this

[2] These data series were obtained from the Thompson Reuters Datastream service.

essay. All three indices exhibit similar volatility clusters over the sample period, the most distinct of which is see between 2008 and 2009, documenting a large spike in volatility at the time of the latest financial crisis.

Figure 1 FTSE Return Chart

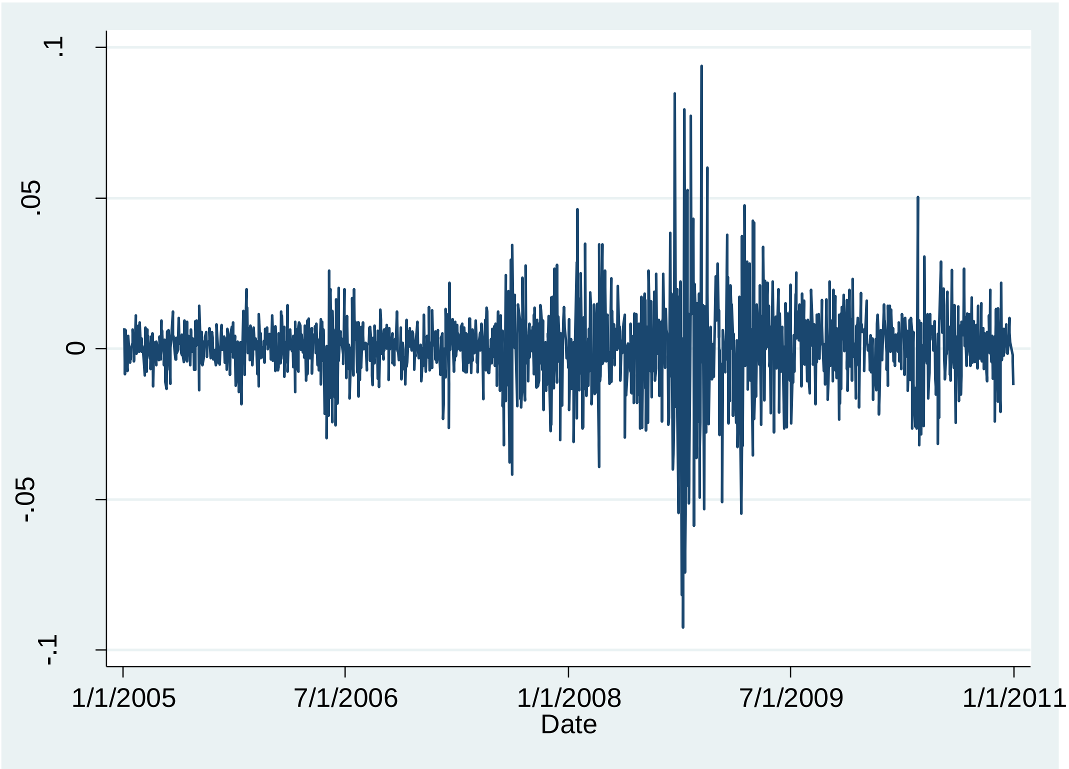


Figure 2 DJIA Return Chart

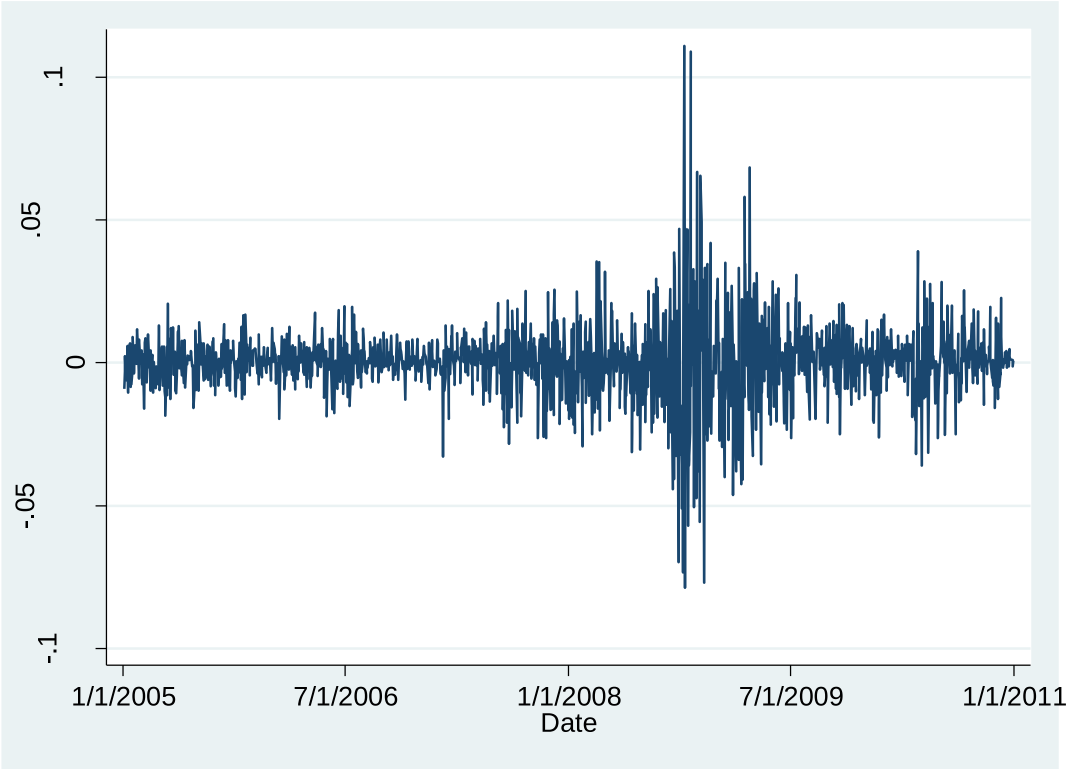


Figure 3 S&P 500 Return Chart

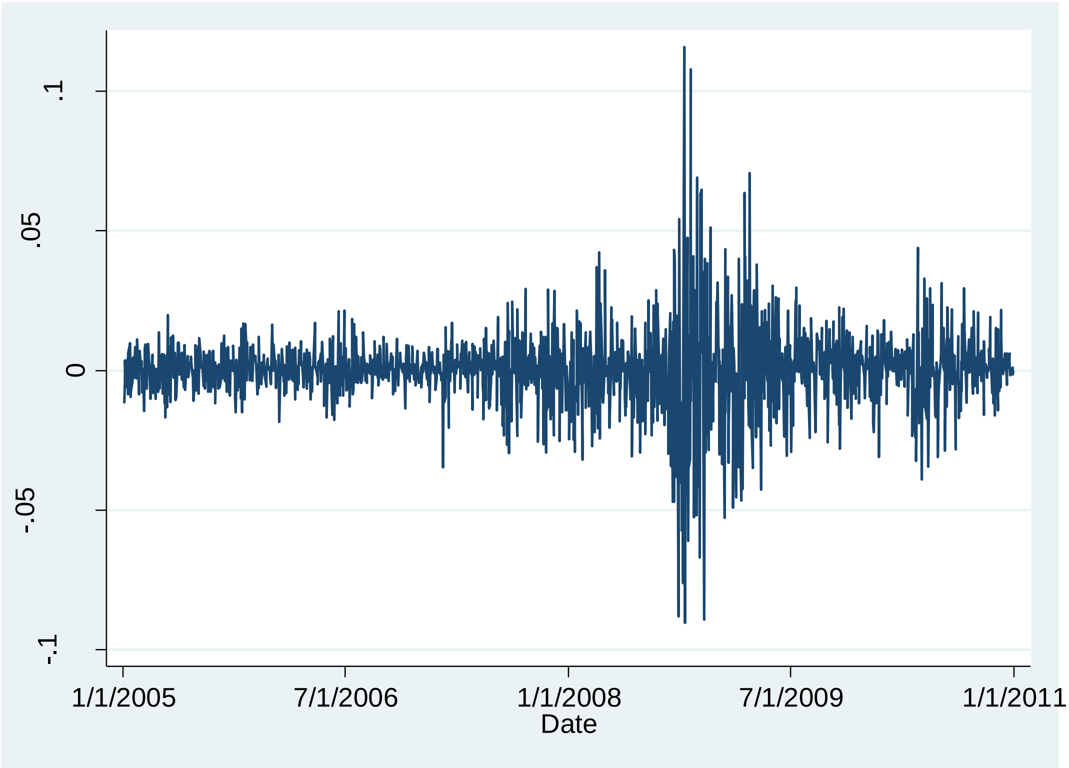


Table 1 below presents some descriptive statistics of returns for the three indices over the sample period. All three indices show similar mean, standard deviations, minimum and maximum values, suggesting that all three time series are likely to exhibit the same properties, and that the model chosen to represent volatility will be the same. Further, **Table 2** presents the correlation of indices, as expected the US indices exhibit strong positive correlation. These datasets will be further explored econometrically.

Table 1 Descriptive Statistics of Index Returns

Variable	Observations	Mean	Std. Dev.	Min	Max
DJIA	1486	0.0001095	0.0134378	-0.0787328	0.1108033
S&P 500	1486	0.0000921	0.0146623	-0.0903498	0.1158004
FTSE	1486	0.0001262	0.0135823	-0.0926557	0.0938434

Table 2 Correlation of Indices

Correlation	DJIA	S&P 500	FTSE Returns
DJIA	1.00		
S&P 500	0.98	1.00	
FTSE Returns	0.56	0.56	1.00

The presence of volatility clustering in the data is identified econometrically in the next section, through explicit tests for autoregressive conditional heteroskedasticity in returns.

1.2.3.2 Presence of Clustered Volatility

In order to obtain accurate results in testing the hypotheses of this paper, it is necessary to appropriately model conditional volatility. As a starting point, a simple Auto Regression AR(1) model of the stock returns is estimated, and Engle's ARCH test is used to show that the null hypothesis of no Auto-Regressive Conditional Heteroskedasticity (ARCH effects) is easily rejected, for each of the stock returns indices.

The results of the tests for Arch effects are presented in **Table 3** for all three indices. First, a simple auto regression of returns is performed using Ordinary Least Squares estimation[3] (OLS). The residuals of the AR models are saved and squared for further estimation.

Table 3 presents the results of the Cameron and Trivedi's information matrix, which tests for heteroskedasticity, skewness and kurtosis, and the results show that we can reject the Null hypothesis of homoscedasticity: the results of the OLS regression are not efficient.

Table 3 Test Results for Autoregressive Conditional Heteroskedasticity

		FTSE	DJIA	S&P
AR Specification		AR(4)	AR(3)	AR(3)
Cameron and Trivedi's Information Matrix	Heteroskedasticity	466.01	380.28	369.87
	p-value	(0.000)	(0.000)	(0.000)
	Skewness	29.51	49.26	47.19
	p-value	(0.000)	(0.000)	(0.000)
	Kurtosis	12.33	8.31	11.68
	p-value	(0.000)	(0.003)	(0.000)
Engle's Test	F-Test specification	F (5,1474)	F (7,1468)	F (7,1468)
	F-Test result	90.73	90.87	96.23
	Prob > F	(0.000)	(0.000)	(0.000)

It is possible to formally test for the presence of conditional heteroskedasticity (time-varying variance) in the index returns using Engle's ARCH test. The lower section of **Table 3** shows the results of an F-test, on a regression of the squared standardized

[3] For FTSE an AR (4) is chosen as the 5th lag is not found to be significant. For the DJIA and S&P 500 an AR (3) is chosen.

residuals from the AR models above on a constant and five or seven lags[4] of the squared residuals. The presence of ARCH effects should be reflected in autocorrelation in the squared residuals. The results of the F-test show that the F-test statistics are above the critical value, so that the p-value is lower than 0.05. These results allow for rejection of the Null hypothesis of no autocorrelation in the squared residuals, and suggest the presence of significant ARCH effects.

These findings suggest that there is systematic, but so far un-modelled volatility clustering in FTSE, DJIA, and S&P 500 returns. This motivates the use of explicit models of Autoregressive Conditional Heteroskedasticity, in order to properly capture this volatility clustering. In line with most recent contributions on volatility modelling in the financial economics literature, this essay uses a Generalized Autoregressive Conditional Heteroskedasticity model to capture volatility clustering in the data.

1.2.3.3 GARCH (1, 1) Results

A GARCH (1, 1) is conducted on the three return indices, and the results are presented in **Table 4**. In the cases of all three stock return indices, the key parameters, β and γ, in the GARCH specification are statistically significant at the 1% level. In line with the literature, the use of the GARCH (1, 1) model is now used as the baseline model, which will be built upon to investigate whether the flow of information plays a significant role in capturing systematic movements in the volatility of stock returns. The extension of the model used in the essay is detailed in Section 5.

[4] The 6^{th} lag was found to be not significant in the FTSE, similarly the 8^{th} lags were found to be not significant in the DJIA and S&P 500 data.

Table 4 GARCH (1, 1) Results

GARCH (1,1)

$r_t = \omega + \mu r_{t-1} + \varepsilon_t$

$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

T=1486	Estimates of parameters						Diagnostics
Returns	ω	μ	α	β	γ	β+γ	LR test
FTSE	0.001	-0.058	0.000	0.121	0.876	0.997	4690.979
p-value	(0.006)	(0.039)	(0.003)	(0.000)	(0.000)		(0.0388)
DJIA	0.001	-0.082	0.000	0.090	0.899	0.989	4769.350
p-value	(0.005)	(0.006)	(0.000)	(0.000)	(0.000)		(0.0059)
S&P	0.001	-0.095	0.000	0.089	0.901	0.990	0.002
p-value	(0.011)	(0.002)	(0.000)	(0.000)	(0.000)		(0.0021)

**A p-value inferior to 0.01 corresponds to a coefficient statistically significant at the 1% level*

**A p-value inferior to 0.05 corresponds to a coefficient statistically significant at the 5% level*

**A p-value inferior to 0.10 corresponds to a coefficient statistically significant at the 10% level*

1.3 Literature Review: The impact of information arrival on the variance of stock returns

Time varying volatility properties of stock returns have been extensively studied in the literature. While GARCH models seek to capture systematic movements in volatility purely by including lagged squared residuals and lagged variance terms in the conditional volatility equation, a great deal of research has focused on identifying other key drivers. One strand of this literature has focused on information flow as a potential driver of the volatility of stock returns. A number of studies have covered a variety of assets and markets, and have made use of different methodologies and a variety of indicators of market activity. This essay attempts to explore the relationship between information flow and stock market volatility in detail, and seeks to build upon past contributions to the literature on volatility and information flow.

1.3.1 Volatility and Information Arrival

French and Roll (1986) is considered a seminal empirical paper in the field. The paper developed the understanding of how information is processed in financial markets and reflected in the variance of stock returns (as a measure of volatility). A key contribution was identifying the empirical regularity that the variance of returns during trading hours, i.e. from open to close of trading on an average day, is considerably larger than the variance of returns over non-trading hours, e.g., close-to-open returns over a weekend. Three potential explanations for this phenomenon are proposed (see French and Roll, 1986, p.6):

i. The volatility is caused by the arrival of public information i.e. "news" that tends to arrive predominantly during normal working hours;
ii. Most volatility is caused by informed investors trading in response to private information received predominantly when the exchanges are open[5];
iii. The process of trading introduces noise into stock returns as investors potentially over-react to each other's trades.

They investigate the relevance of these explanations through three hypotheses tests (summarised in **Table 5**). The methodology consisted of comparing volatility occurring during trading periods against that occurring during non-trading periods in the sample (these include holidays, election days, and market closure days). Specifically, the authors calculate the volatility from the change in price from the close on the day before to the open of the day following the non-trading period. They then go on to investigate the cause(s) behind the differences in trading and non-trading

[5] As defined by Berry and Howe (1994), the distinction between public and private information lies in its availability either to the general public or to a "narrow segment" of the market (p.1331).

hours by testing hypotheses that focus on the arrival of public, private information, and trading errors, while these are admittedly not mutually exclusive hypotheses.

Table 5 French and Roll (1986) Hypotheses [6]

Proposition	Interpretation
"High trading-time volatility is caused by public information which is more likely to be observed during normal business hours."	This hypothesis suggests that volatility should not hinge on whether the exchange is open or closed. Public information is more likely to be produced during normal business hours, when exchanges are open. So a higher volatility during trading hours is simply the reflection of more public information available to trade on.
"High trading-time volatility is caused by private information which is more likely to affect prices when the exchanges are open."	This hypothesis suggests that volatility should be reduced during non-trading hours, as private information only affects prices when informed agents trade. Further, it is possible that more private information is produced when exchanges are open, and security analysts are working. However, even if private information is constant, it will only be reflected into volatility when markets re-open and agents can trade.
"High trading-time volatility is caused by pricing errors that occur during trading."	This hypothesis posits that high trading day volatility is partly caused by a noise component due to the process of trading. The volatility should then fall when the exchanges are closed and the volatility which is lost would not be recovered upon opening.

[6] Here, a rejection of the Null Hypothesis in each of these tests can be seen as providing evidence in support of these three explanations proposed by French and Roll (1986).

French and Roll (1986) results show that despite the fact that a significant share of the difference in returns variance between trading and non-trading hours can be attributed to mispricing and trading errors ("approximately 4 to 12% of daily variance is caused by mispricing"), the effect appears largely driven by the arrival of information. The comparison between holiday volatility and trading day volatility shows more support in favour of Hypothesis 2, which explains that French and Roll (1986) initiated a debate on the relative importance of public and private information in the setting and volatility of security prices.

Although a number of papers have gone on to investigate how private information impacts stock market volatility (including e.g., Admati & Pfleiderer, 1988; Vega, 2006), the majority of the literature has focused on the impact of the arrival of public information. This is also the focus of this essay. However, a key issue that has to be addressed, as Baklaci et al. (2011) suggest, is the fact that the flow (or rate) of information arrival in the stock market is extremely difficult to measure, and there is no consensus in the literature on the most appropriate. This difficulty is reflected in the wide variety of methods so far proposed, and is further discussed below.

1.3.2 Measuring Public Information

1.3.2.1 Early Proxies for News Arrivals

A number of notable contributions to the empirical literature testing the relation between volatility and information arrival have relied upon quantifying the flow of public information using a variety of different proxies for news arrival. Among these contributions, Berry and Howe (1994) use a count of all news releases sent by The Reuters News Service using the North American Securities Wire. Mitchell and Mulherin (1994) study the daily number of headlines released by The Dow Jones

Company news database. They find the volume of news stories, (defined as a count of the daily total number of headlines reported by Dow Jones and Company) and market activity are strongly related; and that they share similar patterns for the day of the week. The relation between market activity and news is significant in regressions which control for the day of the week. Mitchell and Mulherin (1994) did not find a strong link between the chosen proxy for public information and index volatility, as measured through absolute value of daily returns. They conclude that their results suggest a complex relation between public information and stock market activity and recognise that continued development of models of information volatility relation could further develop the understanding of the relation.

Accordingly, the principal criticism relating to these early studies is that they link information flow proxies to relatively simple measures of market volatility. In Mitchell and Mulherin (1994), stock market activity, or volatility, is captured as either the absolute value of daily stock market returns, or the sum of the absolute values of daily market returns for each firm in the index. Kalev et al. (2004) attribute Mitchell and Mulherin's (1994) lack of conclusive results to a primary weakness in these measures of volatility, in that they fail to incorporate the well-documented phenomenon of *volatility clustering* in financial time series. However, these studies pioneered research on the flow of information arrival.

In summary, the previous literature has documented several suggestions of proxy for public information. However, these proxies for public information are limited in scope likely due to the availability of information at the time. Further, these studies have used information proxies in relatively simple models of stock market volatility, which do explicitly address issues of volatility clustering. Since then, contributions in modelling techniques have become more developed, such as the GARCH model, which can capture previously un-modelled conditional volatility.

1.3.2.2 Volume of Trade

In parallel to innovations in proxies derived from the release of news, many empirical studies looking at the relation between volatility and public information have proposed that trading volume can serve as a proxy for information arrival in the market (e.g., Andersen, 1996; Bollerslev & Domoqitz, 1993; Bollerslev & Jubinski, 1999; Dieobold, 1986; Fleming et al., 2006; Fleming & Kirby, 2011; Lamoureux & Lastrapes, 1990; Liesfenfeld, 2001; Yuksel, 2002). These studies justify their belief in trading volume as proxy for the arrival of information based on a (older) large body of empirical evidence showing a positive relationship between trading volume[7] and price volatility, and that volume moves in response to the arrival of information, justifying it (for a review, see Karpoff, 1987). In an influential contribution, Lamoureux and Lastrapes (1990) found that daily trading volume is a significant explanatory variable in stock returns volatility using a GARCH model on the 1986 CRSP database.

Sharma et al. (1996) also tested whether trading volume played a significant role in the conditional variance equation and find that although it led to significant reductions in GARCH effects, the results were less clear than those in Lamoureux and Lastrapes (1990). Specifically, they find with the inclusion of volume in the variance equation, GARCH effects are reduced but do not disappear completely. They argue this difference is likely due to the fact that Lamoureux and Lastrapes (1990) use individual securities, while they use an index; and that although trading volume might be an

[7] Volume as a proxy for information would reflect both public and private information sources.

acceptable proxy for information arrival which affects individual stocks, it might not appropriately reflect the systematic information factors which affect volatility in indices. One major issue with using trading volume as a proxy for information arrival is the fact that it is an indirect indicator, which cannot differentiate between private or public information.

Kalev et al. (2004) also summarises a number of other flaws with the use of trading volume to proxy for information arrival. First and foremost, the use of trading volume is linked to obvious endogeneity issues. Volatility and trading volume are simultaneously influenced by the underlying information arrival process and therefore volume cannot be assumed to be exogenous to volatility (Tauchen & Pitts, 1983; Foster & Viswanathan, 1993, 1995; Harris, 1987; etc.). Second, trading volume may not be accurate in proxying for the rate of information arrival, since a large portion of trading might in fact occur without the arrival of any new information (Anderson, 1996). Anderson's (1996) findings suggest 34-75 percent of daily trading volume is unrelated to information arrival. An example would be liquidity motivated trading. Further, from a micro approach to the market, trading volume might be too "noisy" a proxy for information arrival; as opinions and interpretations of information can vary greatly. This assertion generates a further question of whether the volume-volatility correlation also depends on the complexity of the structure underlying the information. In other words, trading volume might not only incorporate information arrival but also its interpretation by market participants, which makes it too "crude" of a proxy for information flow (Kalev et al., 2004). Finally, trading volume could be a consequence of the process of trading by agents with heterogeneous or asymmetric information; such as situations where informed traders may make strategic trades (e.g. increase in smaller size trades) based on their information lead, which further weakens the relationship between trading volume and information arrival.

Although these concerns help provide motivation for the development of the new information arrival proxy used in this essay, trading volume is still likely to capture some private information, in addition to public information. This means that the contribution of trading volume in explaining volatility cannot be fully dismissed, or must be at least tested for in this essay. However, due to the endogeneity issue highlighted previously, the model used in this essay will test for the contribution of lagged trading volume, in the absence for a better way to control for endogeneity

1.3.2.3 Reconciling Public Information and Trading Volume

In an augmented GARCH model of returns in the Australian stock market, Kalev et al. (2004) chose to relate specific company announcements to return volatility by introducing them in the conditional variance equation. They found evidence that the flow of public information, in the form of the number of company announcements positively impacts the conditional variance of stock returns. In addition, they showed that a lagged volume variable did not affect the significance of the variable capturing the flow of public information. The inclusion of the *lagged trading volume* term was chosen in view of the likely endogeneity of contemporaneous trading volume (mentioned above). In an additional OLS regression, it was found that trading volume is indeed correlated with news arrival.

Cousin and Launois (2005, 2008) extended the investigation to the French and U.S. stock markets respectively. Using the daily count of Reuters announcements as their proxy for the flow of the arrival of new information, their studies confirm the significance of public information in contributing to explaining systematic movements in stock market volatility.

1.3.3 Different Types of Public Information

1.3.3.1 Categorising Information Content

In addition to providing evidence in favour of a volatility-information flow relationship, both Kalev et al. (2004) and Cousin and Launois (2005) also introduce a classification of news announcements in terms of their information content about a specific company. Kalev et al. (2004) examines this relation at the index and individual security level[8]. Cousin and Launois, (2005) use the CAC 40 index as well as the component securities for analysis. They build on Andersen's (1996) argument that because "different types of news have different stochastic arrival processes and hence their effect on return volatility may not be the same" (Kalev et al., 2004). They find that the effects of different types of news announcements on volatility varies depending on the frequency of publishing for each type of information content, specifically referring to content of information related to business. For example, Cousin and Launois (2005) find information relating to corporate governance, and regulation has the highest impact on volatility. Kalev et al. (2004) posit the category of information impacts returns differently, such as quarterly cash flow reports compared with dividend announcements.

Kalev et al. (2004) separate news into the 12 following categories: mergers and acquisitions, change in shareholdings, periodic earnings reports, quarterly activities reports, quarterly cash flow reports, issued capital changes, asset acquisition and

[8] Kalev et al., 2004 uses the 5 most traded securities on the Australian index as a sample of individual securities.

disposal, notice of meeting, stock exchange announcements, dividend announcements, progress reports, and company administration. Based on the analysis of these categories, they find that conditional volatility is significantly influenced by all categories with the exception of the quarterly cash flow report. They also find that volatility persistence is reduced in different ways depending on the type of news included in the variance equation. One interesting finding is that news types associated with a higher frequency (e.g. periodic earnings reports, change in issued capital, asset acquisition and disposal, notice of meeting, and company administration) tend to have a greater impact on the persistence of volatility than news published in lower frequency publications. This findings supports the Mixture of Distributions Hypothesis, which posits that the rate of information arrival should impact volatility and volume. These findings suggest that the categorical content of information and frequency of publishing can have a categorical effect on volatility.

Similarly, Cousin and Launois (2008) also disaggregated their information proxy dataset into nine categories of news by content: merger and acquisition, earnings, earnings projections, analyst comments and recommendations, funding/capital, regulation and government policy, contracts and orders, ownership changes, and earnings surprises. Their results showed that earnings surprises and regulation have the most significant impact on volatility. Additionally, separating the data between news announced during trading or non-trading hours shows that the information flow has a larger impact on volatility when released during market operating hours.

All these findings support Andersen (1996) in that different types of news might have different impacts on stock return volatility, as they have with different arrival processes. These findings largely influence and motivate the creation of the disaggregated information proxy by media type used in this essay (see Section 4). This disaggregation enables this essay to build on the literature to examine the different impacts of information types disaggregated not by content but by medium.

1.3.3.2 Behavioural Finance

In addition to the findings from the empirical literature on volatility modelling and information, there are a number of contributions from the behavioural finance literature, based on psychological notions, which can arguably be useful in examining the relation between various types of information and volatility in the stock market. Using these psychological notions can help to further investigate the proposition that different types of news could have varying impacts on stock market adjustments, and thus volatility, by not only looking at the timing and frequency of news announcements, but also other characteristics.

These psychological notions or principles, relevant to this analysis include *framing* (Shefrin, 1999; Marsh & Tversky, 2004), *availability heuristics* (Andreassen, 1990), *selective perception* (Kahneman & Tversky, 1984) and the *adaptation of mass opinion* (Shiller, 2000).

The principle of *framing,* based on prospect theory, posits that information is perceived differently depending on how it is presented (Shefrin, 1999; Marsh & Tversky, 2004). In the context of securities markets, framing suggests that the way an announcement (either firm specific or macroeconomic) is reported through the media to investors and traders, might influence their perceptions of the announcement. As an example, let us assume that a company's earnings announcement reports a loss for the quarter. While in theory this earnings announcement should be a clear negative signal for an investor, if this information were presented in a "could have been worse" context, investors could interpret this information in a positive manner. The concept of framing clearly has implications which can be extended to the channel through which information is received. The type of medium relating the information can either be purely objective and factual reporting (i.e. a newswire), or could incorporate some degree of opinion or subjectivity (i.e. analyst reports). Thus, framing suggests that the type of media used to report news could potentially influence the digestion of news by market participants.

Another relevant psychological concept is *availability heuristics.* It suggests that investors place a higher weight on information that is easy to access (Andreassen, 1990). For instance, an agent may place a higher weight on the opinion of his/her stockbroker, than on having to form a personal opinion from conducting a thorough analysis of analyst reports, which would have to be sought out and read. This principle reflects the concept of opportunity costs, where individuals favour making trading decisions based on more accessible information sources. For instance, this theory would suggest that more accessible and easily digested media (e.g. Wire-Feeds) could have more impact on trading decisions, and thus prices and volatility, than less accessible media (i.e. analyst reports).

Selective perception is based on cognitive dissonance, asserting that investors are more likely to place a higher weight on information which supports their current beliefs (Kahneman & Tversky, 1984). In a stock market example, an agent who plans to purchase a stock might conduct due diligence by reviewing several analyst reports on this company. If these analyst reports present conflicting evidence, the agent might place a higher weight on the data supporting his/her original plan to purchase the stock when making the decision.

An interpretation of combining availability heuristics and selective perception can go further still in explaining the impact of information on stock market volatility. For example, a type of media published with higher frequency could potentially represent a wider range of opinions (or information which needs digesting) and thus lead to higher disagreement amongst investors. This could be due to two phenomena: First, investors could spend more time reading through information to find opinions which conform to their own. Second, this could generate more disagreement amongst investors as a higher volume of information is likely to produce a higher number of conflicting opinions which investors could find as support for their own beliefs, even if they are in the minority. See **Table 6** below for a simplified example (flows from left to right).

Table 6 Selective Perception and Availability Heuristics Illustration

Media Source	Volume of daily publication	Conflicting news signal (negative versus positive Reports)	(1st point) Digestion period	(2nd point) Probability of finding information which conforms to, and perpetuates your beliefs if holding a minority view point
A	1	No	short	Low
B	100	Yes	Long	High

Finally, the idea of *adaption of mass opinion* argues that investors generally place their trust in information that is recognised as true by notable and/or popular sources (Shiller, 2000). An agent might be more trusting of the analysis of a well-known television security analyst, such as seen on CNN Money or similar programs, since the program is successful. This further relates to the concept of *herding* (Shiller, 2001), which suggests that investors can have financial incentives to behave in accordance with a group. Herding behaviour suggests that investors are afraid of missing opportunities (by not following a group) or failing alone (if one chooses one's own path, which proves to be less prosperous). Given the difficulties in detecting "group trends" in real time, this theory supports the notion of prolonged patterns in prices, which could be perpetuated by information and have a lagged impact on prices depending on the source of information. For example, a popular magazine highlighting a growing industry or sector (e.g. technologies), could both confirm a price trend to the reader, and deliver the message through a trusted and notable venue. This would encourage herding behaviour, and confirm the effect of information by media type. This example also illustrates the possibility of a variation in the lagged impact of various types of media to relay information on investors' behaviour, as discussed next.

1.3.4 Contemporaneous or Sequential

Up to now, this literature review has focused on *empirical results* of the information-volatility literature. These do resonate with early theoretical contributions in financial economics and their suggestions as to the timing of information arrival, and the speed at which information is incorporated into stock returns and thus reflected in volatility.

The relationship between information arrival and stock market volatility has been theorized as early as the 1970s, starting with the Mixture of Distribution Hypothesis (MDH) (Clark, 1973). In an example of cotton futures markets, Clark (1973) proposes that price changes (returns) can vary from one day to the next, in relation to the flow of information received by traders, which also varies every day. Accordingly, the MDH posits that the variance of returns (or volatility) is driven by the rate of information arrival (Clark, 1973; Epps & Epps, 1976; Harris, 1986). Further, this theory formally suggests that price volatility and the volume of trades are jointly distributed, i.e., contemporaneously dependent on the same underlying variable: the flow of information (Harris, 1986). This proposition has influenced the empirical studies of the information-volatility relation discussed previously.

Another theory relevant to this literature is the Sequential Information Arrival Hypothesis or SIA (Copeland, 1976; Morse, 1980; Jennings et. al., 1981; Jennings & Barry, 1983). This model differs from the MDH in that it proposes that the volume-volatility relation is sequential, and not contemporaneous. Copeland (1976) proposes a model in which information is disseminated gradually to the market and traders sequentially adjust their demand curves. In effect, this corresponds to a situation where there may be a time lag between the initial release of information and action by investors in responding to the information (Jennings & Barry, 1983).

Karpoff (1987) argues that the MDH and SIA are not mutually exclusive, but subsequent studies have investigated these hypotheses and explored their empirical validity (Davidson, 2014; Epps & Epps, 1976; Najand & Yung, 1991; Harris, 1986; Alsubaie & Najand, 2009; Boubaker & Makram, 2011). Harris (1986) and Epps and Epps (1976) findings support the MDH, but argue the model is mis-specified to capture non-normal distributions, and suggest an augmentation to the model which would help explain the leptokurtic residuals apparent in Clark's (1973) MDH model specification.

Mougoue and Aggarwal (2011) find evidence in favour of the SIA hypothesis but against the MDH, showing a lag relationship between volume and volatility in currency markets: specifically, they find a negative contemporaneous correlation between the volume of daily trades and the volatility of returns across three futures contracts, but they do find a lead-lag relation between trading volume and return volatility, using both linear and non-linear Granger causality tests. These findings support the SIA. Herbert (1995) also finds that past trading volume influences current volatility in the natural gas market. In the stock market, Chen et al. (2001) find that neither contemporaneous nor past volume can explain the persistence of volatility satisfactorily. Davidson (2014) finds evidence in support of the SIA, finding a positive correlation between trading volume and volatility, across 15 internationally-traded indices, through the existence of a lead-lag relation between volume and volatility (this relation presents evidence for sequential leading effects in volume and lagged effects in volatility). It is apparent that the majority of these studies have used trading volume to proxy for information arrival, informed by the previously referenced literature. However, the issue of contemporaneous or sequential impact of information on securities markets has not been explicitly explored in a study using proxy for information arrival based on actual media publication. This essay builds on the existing literature by examining the impact of lagged information on the volatility of stock returns to further document the possibility of a sequential relation.

1.3.5 Gaps in the Literature

In light of this review, a number of issues regarding the relation between information arrival and stock market volatility remain unanswered or lack consensus in the literature. This essay contributes to the literature by addressing some of these questions as summarized in the table below.

Table 7 Literature Review Summary and Motivations Table

Questions from the Literature	Contributions of This Paper
What is an appropriate way to define and quantify public information?	While previous papers use trading volume, or relatively constrained proxies constructed from news feeds on specific stocks, there has not been a proxy constructed to encompass public information in a more general sense, which could be appropriately related to volatility in an index. The essay proposes the use of a new proxy for the flow of public information, attempting to capture information in a broad sense, which is not restricted to firm-level news. The database used to construct the information flow proxy is arguably more wide ranging than previous attempts as it captures the largest collection of business related publications.[9]
To what degree can the flow of public information arrival contribute to explain movements in stock market volatility?	This new proxy for public information is used in a GARCH (1, 1) model[10] of stock returns to contribute to the literature on the relationship between information flow and volatility. The model is augmented to examine the influence of the proxy for flow of information on the volatility of returns.
Do different types of information affect volatility differently (due to the way it is presented to market participants)?	Recent contributions to the literature described above suggest that different types of information content have varying impacts on the volatility of stock returns. This essay further examines this finding: The new proxy presented in this analysis is designed to capture the total flow of public information, but it is also disaggregated by media type. While the content of information arrival has been previously studied in the literature, there has not yet been empirical testing of how the impact on volatility of different types of media to convey the information may vary. Previous literature has used different types of media to proxy for information flow (e.g. wires or newspaper headline articles) however, this essay makes the direct comparison of the impact of various media

[9] The method of construction for the dataset used in this essay will be detailed in Section 4.
[10] See Sections 1.2 and 1.5.

	types possible. Using the disaggregated data by media type, this question can thus be answered in a systematic estimation exercise. The issue of timing is also further investigated with regards to the disaggregated dataset. An analysis of the impact of backlog and lags for each media type is performed. The relationship is estimated for the total flow of information, but is also investigated by media type to determine whether different reporting media have a contemporaneous or sequential relationship with volatility in the stock market
Is the relation between the flow of public information and volatility contemporaneous or sequential?	Through the use of lags and backlogs in the flow of public information, this essay also addresses the debate of contemporaneous or sequential influence of information on stock market volatility. This essay examines whether, as suggested in the literature, information accumulated during non-trading days is reflected in changes in volatility when the markets re-open. The construction of a backlog variable enables the direct testing of this question. Additionally, the use of lags in the information flow variable can help further answer the question of whether all public information is digested by markets contemporaneously.

1.4 Datasets

This essay uses two sources of data for analysis: Datastream for FTSE, S&P 500, and DJIA index price data (as described in Section 2.3) and the ProQuest database for the information proxy. The daily data covers a sample period from January 1, 2005 through December 31, 2010.

The following section describes this data in detail, including the collection and creation of the information dataset (which proxies for the flow of information). Several graphical analyses and descriptive statistics are presented.

1.4.1 A Proxy for the Flow of Information Arrival

One of the aims of this essay is to create a dataset to proxy for the daily flow of public information. As noted in section 3.2.2 above, the use of trade volume has been criticized in the literature, hence there is a need for other proxies in order to improve our understanding of the role of public information in the stock market. Rather than restricting attention to a single category of publications, a question asked here is, "Can one capture all public information which affects trades?" The answer is most certainly no. However, and importantly, we can try to extend and hopefully improve on the information proxies already identified in the literature. Specifically, this essay attempts to contribute to this research gap by creating a new proxy for the flow of new public information, which can also be disaggregated and used to test a number of hypotheses on the relevance of the role of the flow of information in capturing systematic movements in stock market volatility.

A new dataset is constructed to quantify the flow of public information using ABI/Inform Complete, which is the business database sub-part of ProQuest search engine. Once collected, this dataset is used to construct a number of variables to proxy for the daily flow of new information.

1.4.1.1 Using ProQuest Database

ProQuest Corporation provides a searchable database of databases recording publications[11]. A ProQuest query is in essence searching a set of databases of published media articles in a variety of formats. These formats can include: Books, Conference Papers and Proceedings, Dissertations and Theses, Encyclopaedias and Reference Works, Government and Official Publications, Historical Newspapers, Historical Periodicals, Magazines, Newspapers, Other Sources, Pamphlets and Ephemeral Works, Reports, Scholarly Journals, Trade Journals, and Wire-Feeds.

ProQuest has been previously used in the financial literature relating media to market activity. However, the approach used in this essay differs significantly from previous financial research.

1.4.1.2 Previous Use of Information Databases in the Literature

In past contributions, ProQuest has been used as a search engine to retrieve information about a specific event or entity, or information relating to specific date or time period. Thus, it has been used to perform general queries of the literature, in a similar manner to other search engines, but enabling a restriction to certain types and topics of publications.

For instance, Shiller (2000) describes the use of ProQuest in an event study examining the securities' market crash of 1987. In this case Shiller (2000) searched the ProQuest

[11] The number of searchable databases with ProQuest will vary depending on the subscription agreements for each institution.

database to review the content of the news articles released just prior to the crash in an attempt to identify substantial information which would instigate the crash. Engelberg and Parsons (2011) also use ProQuest to filter results by region and extract local media coverage from specific newspapers in order to assess the impact of local news coverage on local securities. Dougal et al. (2012) use ProQuest to specifically retrieve Wall Street Journal articles from the newspaper's archives. They examine the linguistic pattern used by the contributing columnists, identifying bullishness (positive associative words) or bearishness (negative associative words) and relate this measure of prevailing sentiment to DJIA index prices from 1970-2007.

These contributions have used information collection from databases to retrieve either publications restricted to a specific region or a timeframe, and/or to examine the content of information (for example bullish or bearish news). In contrast, this essay aims to provide a new proxy for the flow of public information which might impact stock market returns and their volatility, in a wider sense. Thus, in contrast to previous uses, here ProQuest is used as a counting tool, which presents the number of publications in a given day. In essence, ProQuest is used to create a measure of the daily volume of public information. This method of quantifying the volume of media by counting the number of publications is consistent with Fang and Peress (2009) who have used the LexisNexis database to count the number of published newspaper articles about a specific security. However, it is the first time that such a method has been employed to construct a proxy for the total flow of public information which is capable of being disaggregated by media type.

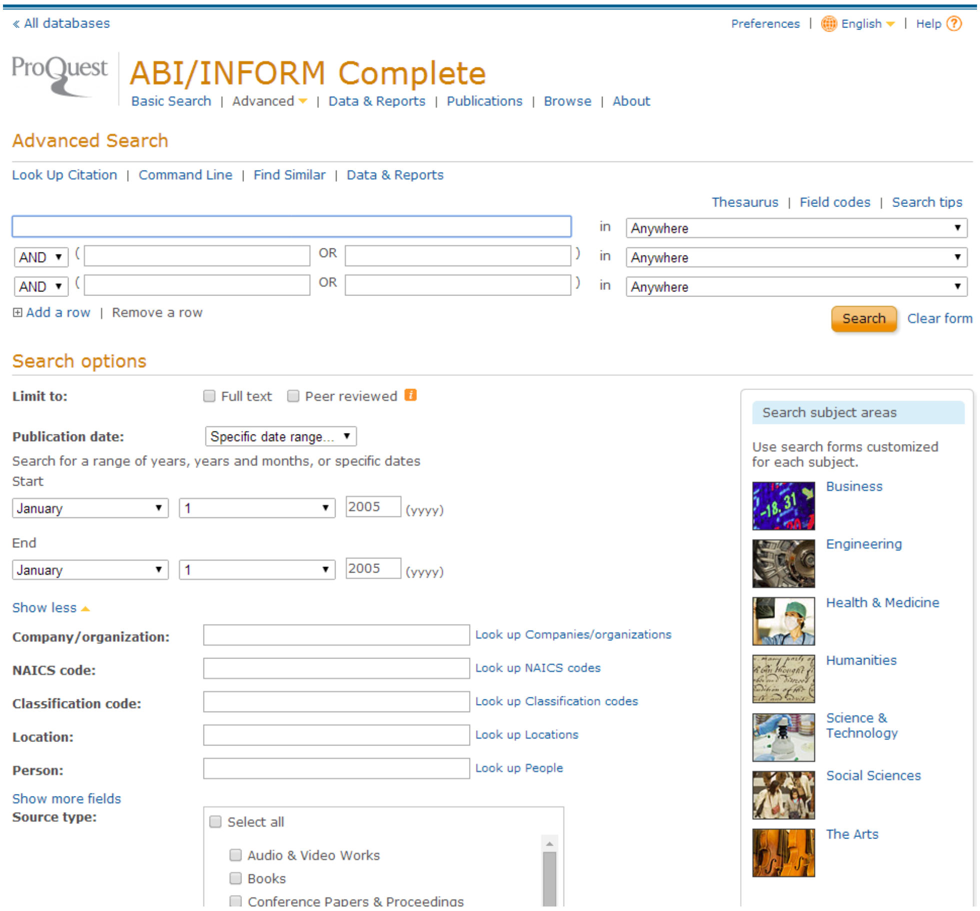
1.4.1.3 Data Collection

For the purpose of this essay, only one of the multiple databases contained in ProQuest was used in the proxy construction, namely the ABI/INFORM Complete database (hence forward ABI). Because the essay is concerned with the impact of information flow on volatility in stock markets, the searchable databases were restricted to ABI, the sub-database in ProQuest that focuses solely on business-related publications[12]. This is an aim to eliminate any news which would not contribute to share price change. ABI is considered one of the principal searchable databases for business, which, as ABI claims, is due to having special relationships with the major publishers for complete coverage of media types such as: The Wall Street Journal, The Financial Times, The Economist, as well as leading trade journals.

Although ABI offers many search options, this essay is interested in the daily flow of information in business-related media. Figure 4 shows a screenshot of the ABI search engine. The query options include: a range of keyword searches including an "and, or" feature, a date range, and media type, among other features.

[12] The author would like to thank Professor Ian Marsh for highlighting the importance of the choice of a more targeted database to reflect investors' public information.

Figure 4 ABI Search Engine Caption



In this case, the search of the ABI database must be as encompassing as possible to incorporate all daily published information which might impact stock market participants. Therefore, the search is conducted without keywords. By not restricting the search field with a handful of terms (e.g. NASDAQ, technology), the dataset can proxy for the total daily flow of public information in contrast with information relating to a specific security or event. An additional advantage of quantifying information in this way is the possibility that results include several publications relating to the same event (e.g. several Reports related to one earnings report). This enables the quantification of stale information and noise which have been shown to have an impact on security prices (Tetlock, 2011).

The only restrictions applied to the ABI queries are the date range (one single day at a time), and the "source type," which qualifies the type of media the information is published in.

As stated previously, a large number of media formats are recorded in ProQuest and ABI. However, it is unlikely that all of them influence securities trading in the same way (see Section 2). It is also likely that certain media formats have little or no impact on security prices, as they would simply not be considered or seen by investors. Based on these observations, several restrictions are imposed to the type of media counted and recorded in this essay, in order to improve the quality of information dataset[13]. Some of the media formats available on ABI are excluded from the database after careful consideration and analysis, based on two criteria:

1. Some forms of media are published at relatively low frequency or only once and are difficult to assign to specific days. In particular, some publications are released during a given year, quarter, or month. Although they are still released on a specific date, this is often not captured and collected in databases on that date. In fact, they are often misleadingly allocated to the first day of the year or the first of the month, corresponding to the year/month of the publication. If these publications were included in the data, the flow of information on these specific days would not actually correspond to the information available to investors on that given day. Two forms of such publications are identified in the search engine: Books and Scholarly Journals.

[13] The author would like to thank Professor Ian Marsh for his constructive comments on how to improve the information proxy used in this essay.

2. Some forms of media are also less likely to influence investors' decisions, in that their informational content is not perceived as relevant by market participants. This judgment could be linked to whether the content is more theoretical/less applied, or considered less timely. Coincidentally, these publications also correspond to the media forms which are published at lower frequency, as investors potentially dismiss the importance of publications relating past events, which are presumably already reflected in prices.

The types of publications that are most likely to have the characteristics which could give rise to the problems outlined above,: Books, Conference Papers & Proceedings, Dissertations and Theses, Encyclopaedias and Reference Works, Government and Official Publications, Historical Newspapers, Historical Periodicals, Pamphlets and Ephemeral Works, and Scholarly Journals. To avoid distortion, these categories of publication are excluded from the analysis. Through the exclusion of these media types and the restricted choice of ABI/Inform Complete (business restricted database) the dataset should be more reflective of the information digested by investors.

In order to illustrate the impact of excluding these types of publications from the proxy measure, two figures of the aggregate count of publication for a sample year are shown below. The first series (Figure 5) includes all publication types (i.e. raw search, without un-selecting certain types of information, when running the search and counting the results), the second series shows the same time series, but excludes the publications listed above. A large variation can be seen between Figure 5 and Figure 6, both in the volume of information and the monthly distribution. Predominantly, in the unrestricted publication count, a large spike occurs in January, as some media types published during the year tend to be grouped into January when they are not associated to a specific publication date, e.g. books and scholarly journals. This seems to confirm that a dataset including publications of these types of media would greatly distort the information proxy.

Figure 5 Annual Unrestricted ProQuest Article Yield

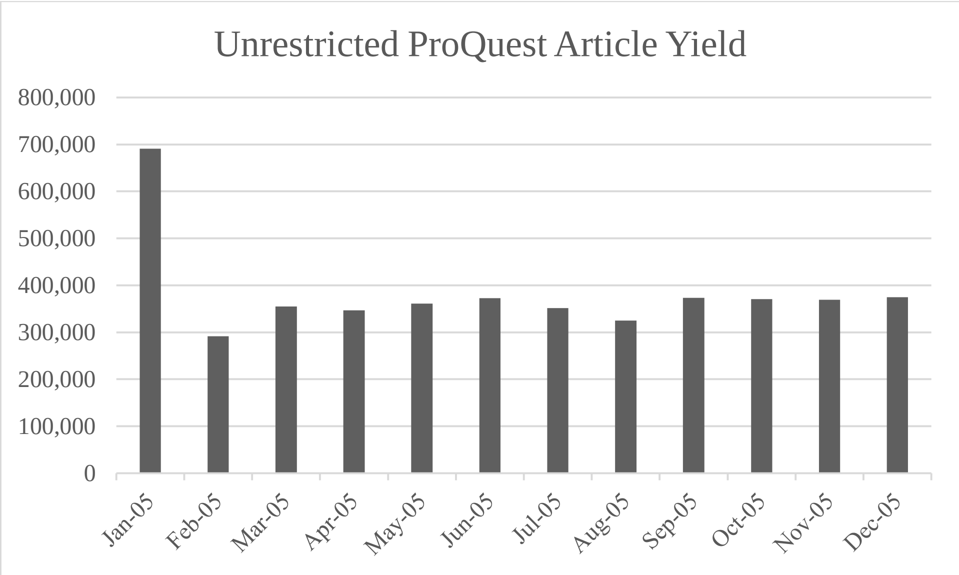
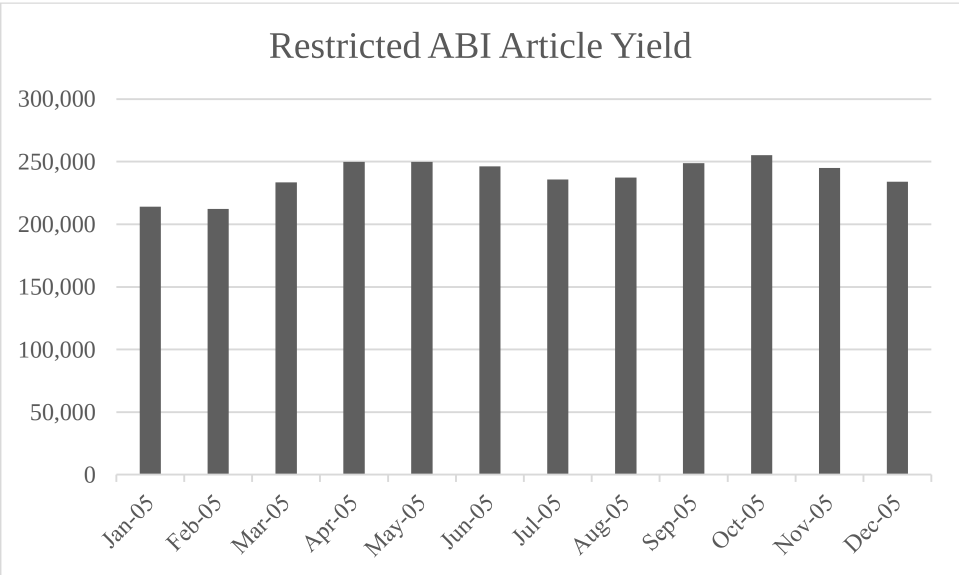


Figure 6 Annual Restricted ABI Article Yield



After this filtering process, five categories of media are considered in the analyses and recorded in the daily count for the selected time period. These five categories include: *Wire-feeds*, *Newspapers*, *Trade Journals*, *Reports*, and *Magazines*.

This approach differs significantly from previous efforts to quantify media coverage, as it attempts to proxy for the majority of information released that is publicly available to investors and can therefore feed into their trading decisions. However, while this dataset represents a new contribution to the information flow literature, it is limited in a number of ways.

First, this method limits the dataset of public information to publications recorded in the ABI database. This dataset is likely to only be able to capture a fraction of the total available public information. Second, this method will almost certainly include some information that might not have been considered by investors, and could still omit other information sources[14]. Improving this proxy, and more generally proxies for the flow of public information arrival represents a vast area for future research contributions. The method used in this thesis should therefore be considered simply a new attempt to proxy for the flow of public information, while suggestions to extend this proxy will be discussed in the conclusion section.

1.4.1.4 Information Variables

The newly created dataset is composed of recorded number of publications, collected daily from January 1, 2005 to December 31, 2010. This 6-year timeframe represents a total of 2,192 daily observations. These observations correspond to 1,513 trading days.

14 Although, the search *is* limited to business publications through the choice of the ABI/Inform database.

This essay attempts to create a good proxy for the flow of public information arrivals. For that purpose, a number of variables are constructed using the raw dataset, which will then be used in the regressions in Section 5.

Table 8 summarizes all these variables created and used in the analyses. The first variable aims to capture the daily flow of new information available to investors only on trading days. This variable, called "new information" is essentially the result of the count of publications on a given trading day. For example, the value of the "New" variable on a Monday corresponds to the flow of information released on that Monday only. The "*New-All*" variable captures all publications regardless of their types (given the constraints imposed and previously discussed). Additionally, a "new" variable is also created for each category of media, e.g. "*New Wires*," "*New Magazines*," etc.

Additionally, given the questions identified in the literature review, this essay will also address the issue of whether past information, which has accumulated during non-trading days, can have an impact on stock market volatility[15]. Thus, the issue of information backlog (which accumulates during non-trading days) is addressed explicitly through the creation of backlog variables. First, a "*backlog dummy*" is created to indicate whether today follows one or more non-trading days, in other words whether an information backlog has previously accumulated. In essence, the *backlog dummy* is equal to 1 for all trading days which follow non-trading days (i.e. every Monday and every day that follows a bank holiday). The dummy is equal to zero on all other days.

[15] Because of the daily nature of the dataset, it is out of the scope of this essay to study the impact of information accumulated during non-trading "hours". Instead, this question is examined using a daily measure, i.e. the backlog during days when the market is closed (weekends, holidays).

Then using the *backlog dummy*, a "*Backlog-All*" variable is created to represent the flow of information which occurs during non-trading days, and thus cannot be captured in price that day. For example, the publications occurring during a weekend or a bank holiday are recorded in this *backlog* variable. A separate *backlog* variable is created for the total publications, as well as for each media type.

Finally, the backlog is then incorporated in the flow of information on the day following the non-trading days in a new "Total" variable. For example, on a Monday, the "Total" variable would be a sum of the daily flow of information on that day plus the backlog of publications over the weekend. A "Total" variable is created both at an aggregate level of all media types, and for each media type separately.

Table 8 Flow of Information Variables

Name	**Definition**	**Notes**
Flow of New Information		
New-All	Number of new publications released on the current trading day. *Does not include backlog data.*	Includes all media categories. *Monday data = only Monday data and so on.*
New-Wire	Number of Wire-Feeds released on a given trading day	Only includes current day Wire-Feed data, does not include backlog data
New-Trade Journals	Number of Trade Journal publications on a given day	Only includes current Trade Journal data, does not include backlog data
New-Newspapers	Number of Newspapers published on a given day	Only includes current Newspapers data, does not include backlog data
New-Magazines	Number of new Magazine issues published on a given day	Only includes current day Magazine data, does not include backlog data
New- Reports	Number of Reports published on a given day	Only includes current Reports data, does not include backlog data
Backlog Information		
Backlog-All	Number of publications released while the markets were closed.	Includes all media categories

	Resulting in the Backlog of information to be incorporated into trading when markets reopen, counted across all categories	
Backlog for each media type[16]	Number of publications during non-trading days for each media category	One backlog variable is constructed for each of the 5 media categories
Backlog	Equals 1 on a trading day with an information Backlog and equals 0 on days without.	Dummy variable
Total Information		
Total-All	Total number of publications on a given trading day (covering all media types) including the Backlog of publications that was issued over the weekend/holiday	Total-All = New-All+Backlog-All. This represents the sum of the total categories for (Wire-Feeds, Newspapers, Magazines, Reports, and, Trade Journals) equalling the total count of all information
Total for each media type	Total number of publications for each media category including the backlog data for each category.	Total-"Category" = New Data+ Backlog data for each category (Wire-Feeds, Magazines etc…)

1.4.1.5 Descriptive Statistics and Graphical Analysis

[16] A backlog variable is created for each media type, e.g. *Backlog-wire, backlog-magazine*…

Table 9 Descriptive Statistics for Information Proxy Variables[17]

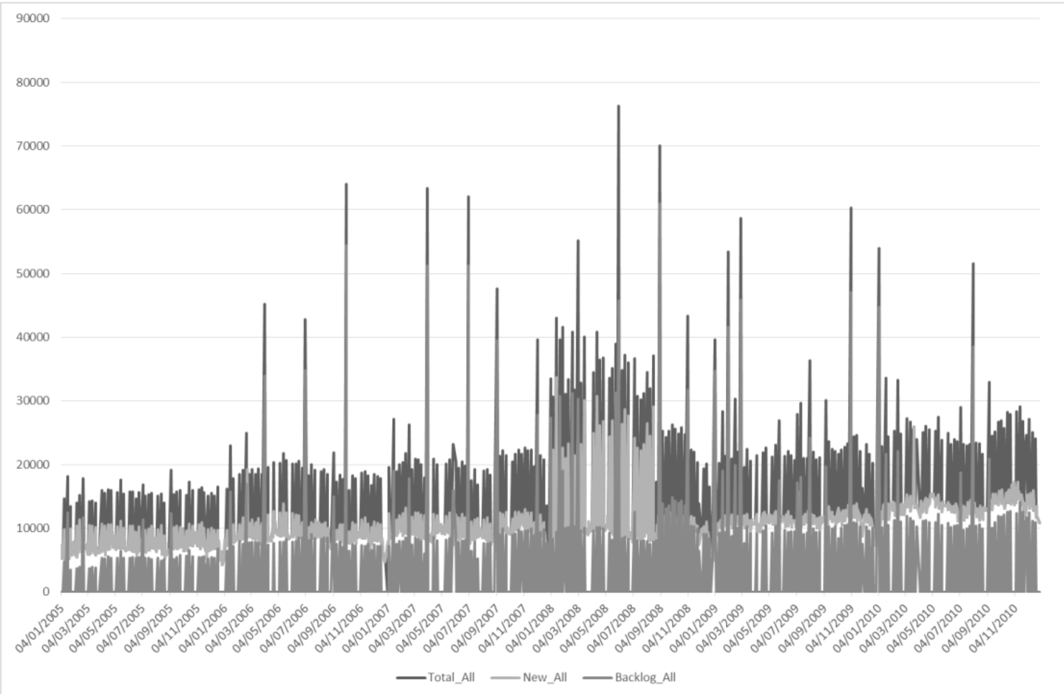
Descriptive Statistics

Variables	Mean	Max	Min	Std Dev	Skewness	Kurtosis	Count
Total All	14250.07	76305.00	4032.00	9341.41	2.50	8.08	1486.00
Wire-feeds	5651.07	21321.00	283.00	1703.80	0.70	5.55	1486.00
Trade Journals	4276.44	38978.00	565.00	4775.82	3.08	9.35	1486.00
Reports	257.40	14243.00	0.00	521.16	18.48	436.40	1486.00
Magazines	504.07	9860.00	0.00	961.54	3.74	15.42	1486.00
Newspapers	1200.88	2872.00	214.00	438.88	0.93	-0.32	1486.00

[17] It should be noted the minimum values of zero for both *Magazines* and *Reports* occurred a total of 4 times, once for *Magazines*, 3 times for *Reports*.

Table 9 above presents the descriptive statistics for the information variables used in the analyses for this essay. It can be seen that there are 1,486 observations across all variables, with the remaining 6 columns presenting statistics on the mean, maximum, minimum, standard deviation, skewness, and kurtosis. The following Figure 7 presents the total information variable graphically (created using ProQuest). This includes all of the five sub-categories of information. This includes the backlog of information which is published on non-trading days and included on Monday. It can be seen this data shows periods of significant clustered information.

Figure 7 Total Information



The *Total-All* variable is constructed using the *New-All* variable and the *Backlog-All* variables representing all the information on a given trading day including previous information generated during the previous non-trading days (such as weekends or holidays). There appears to be a higher average in the middle of the sample corresponding to the start of the financial crises.

The statistics in Table 9 show leptokurtic distributions for *Total-All* and all individual information types, as well as excess skewness. *Wire-Feeds* exhibit the largest mean daily number of publication, which is to be expected given their relatively concise nature. Of the five categories there are the fewest *Reports* with a mean close to 260. Figure 8 presents histograms for the distribution of the *Total* information variables by media types. The histograms confirm that Magazines, Reports and Trade journals show high skewness and kurtosis in comparison to Wirefeeds and Newspapers, reflecting the fact that the latter two tend to be published less in clusters. Reports show extreme leptokurtic distribution suggesting that many of the observations are concentrated around the mean, as well as extreme (right) skewness.

Figure 8 Histograms of Individual Information Types

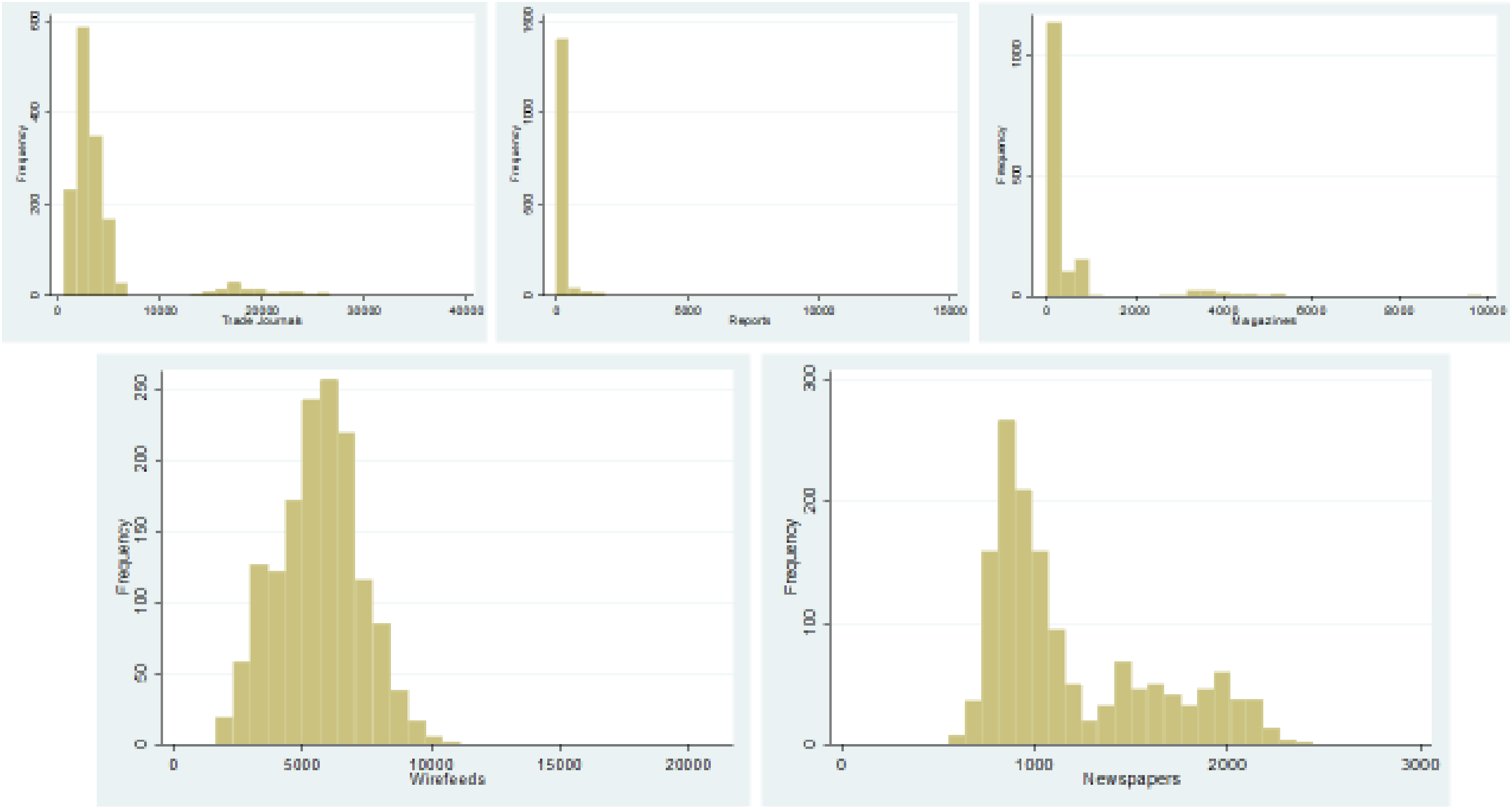


Figure 9, Figure 10, Figure 11, Figure 12, Figure 13, and Figure 14 represent each of the variables in Table **9** graphically. Figure 12 confirms the descriptive statistics observation on *Reports*. The time-series data exhibits rare and large departure from the mean, suggesting *Reports* tend to be published in large clusters, with several extreme outliers. This reflects the quarterly nature of report publishing frequency (for example, corporate earnings reports). *Newspapers* and *Wire-Feeds* together have the lowest kurtosis and skewness of all the variables. Graphical observation also suggests both experience higher frequency of variability and some degree of drift overtime. *Magazines* and *Trade Journals* experience a similar level of kurtosis and skewness. However, graphical observation shows that *Trade Journals* experienced higher variability during mid-year 2008, coinciding with the financial crisis. While *Magazines* have continuously the same pattern over the entire sample. This reflects the regular and monthly nature of magazine publication frequency. *Total-All* shown in Figure 9 represents the aggregate of all variables previously discussed.

Figure 9 ***Total-All***

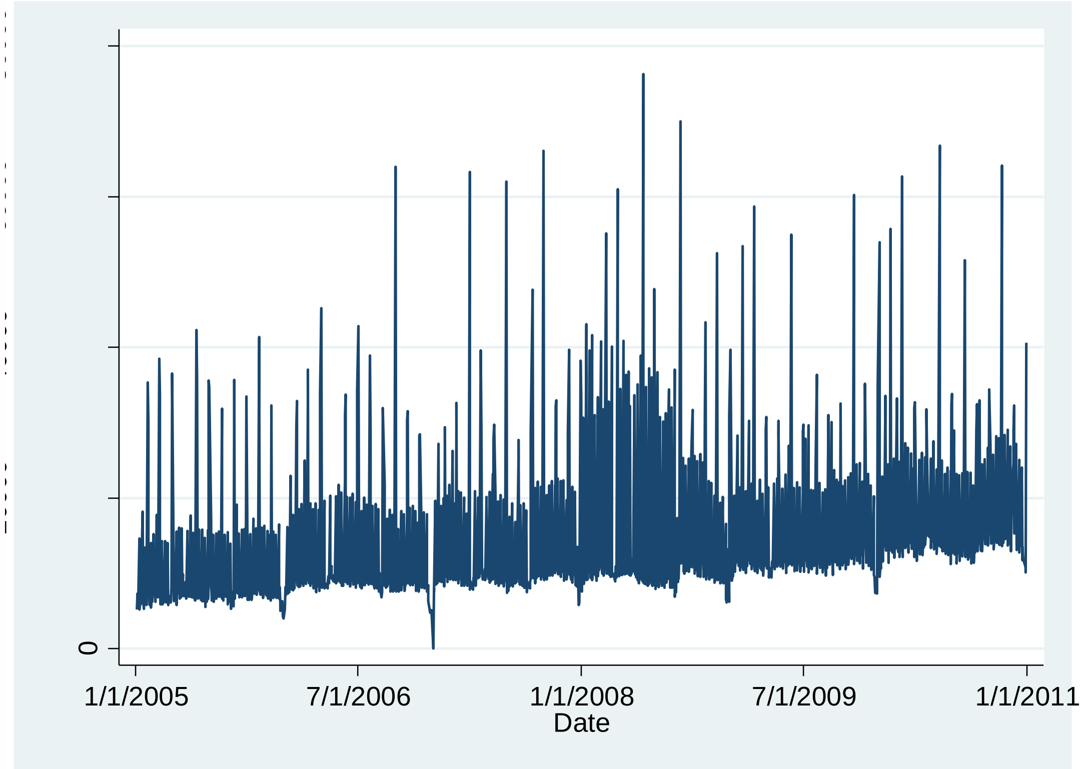


Figure 10 ***Total Wire-Feeds***

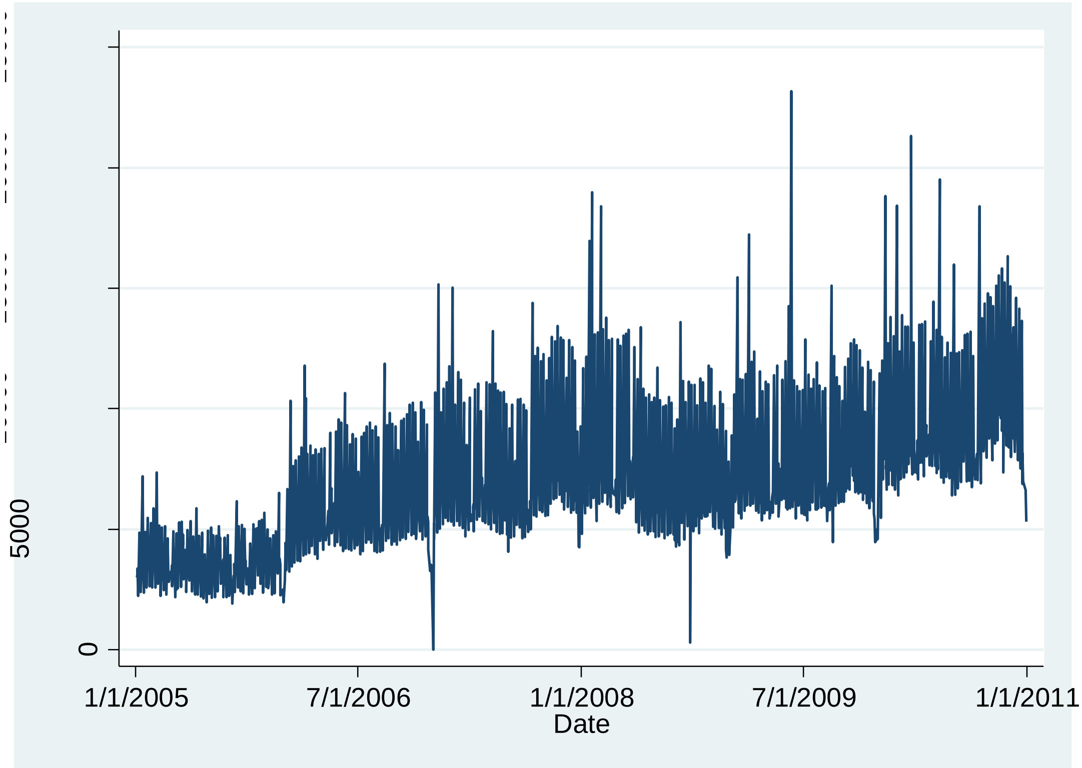


Figure 11 Total *Trade Journals*

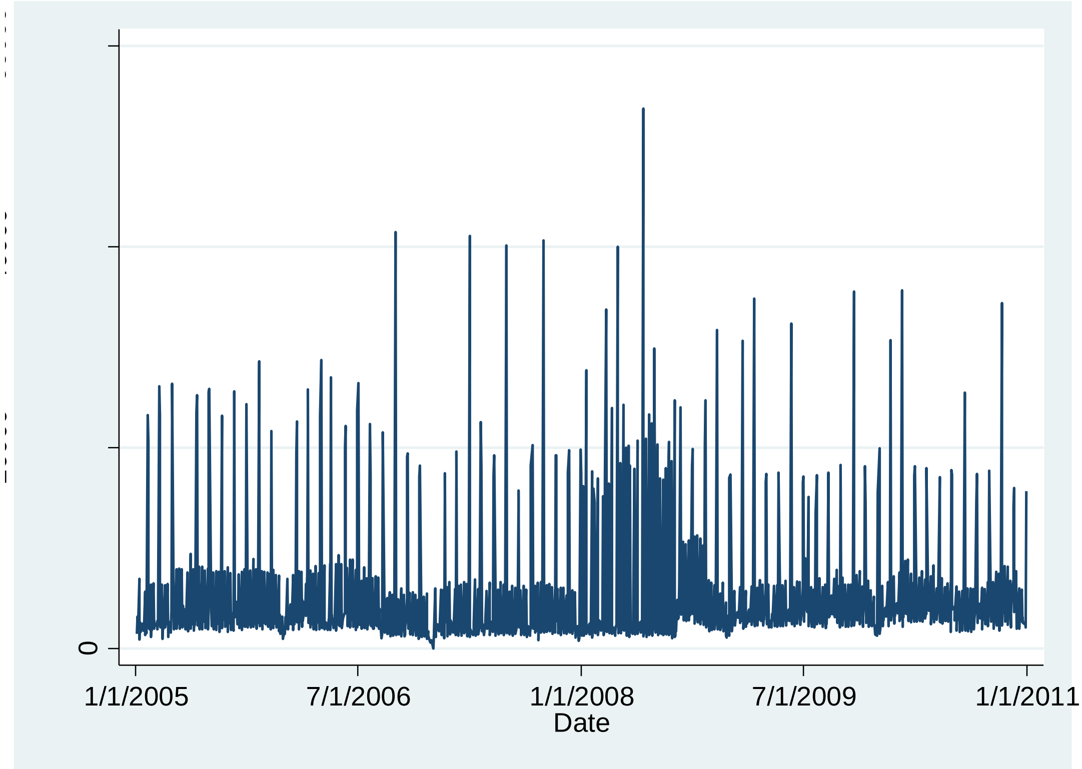


Figure 12 *Total Reports*

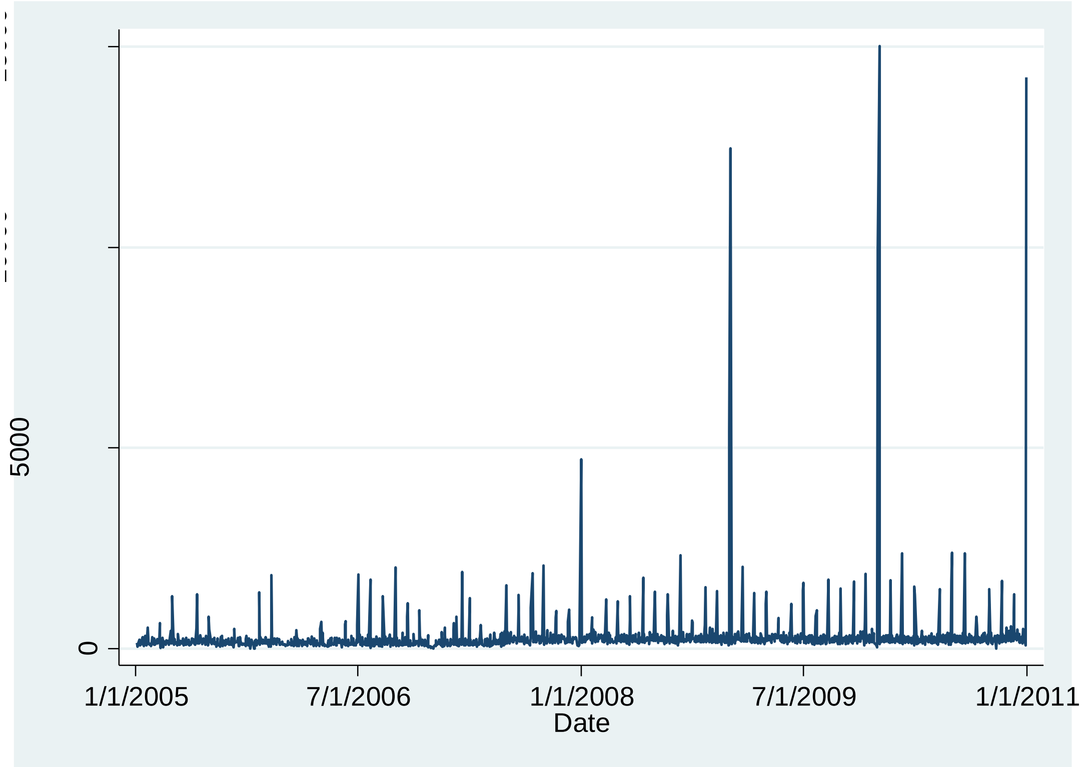


Figure 13 ***Total Magazines***

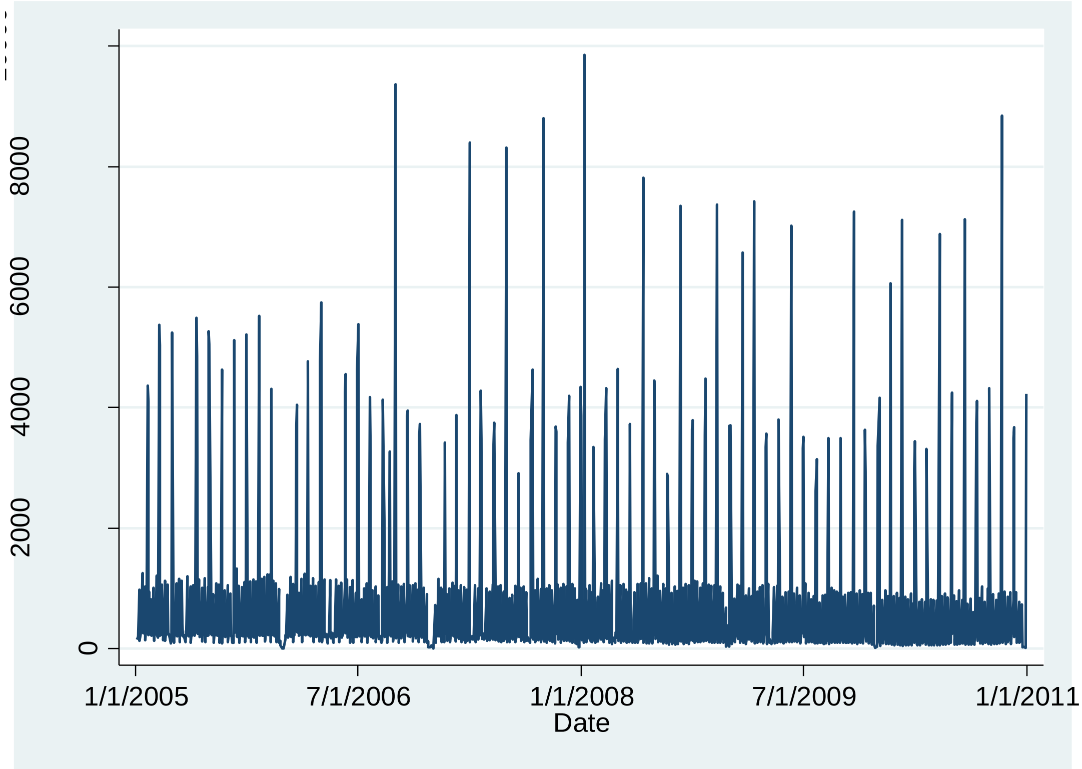
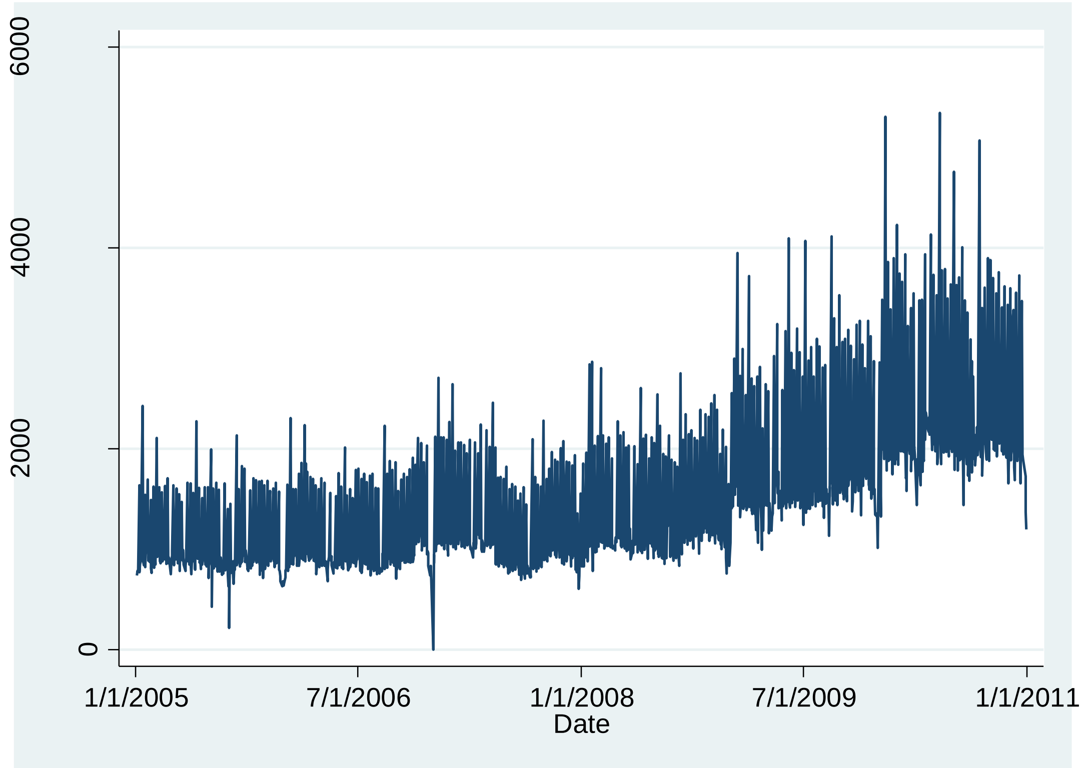


Figure 14 ***Total Newspapers***



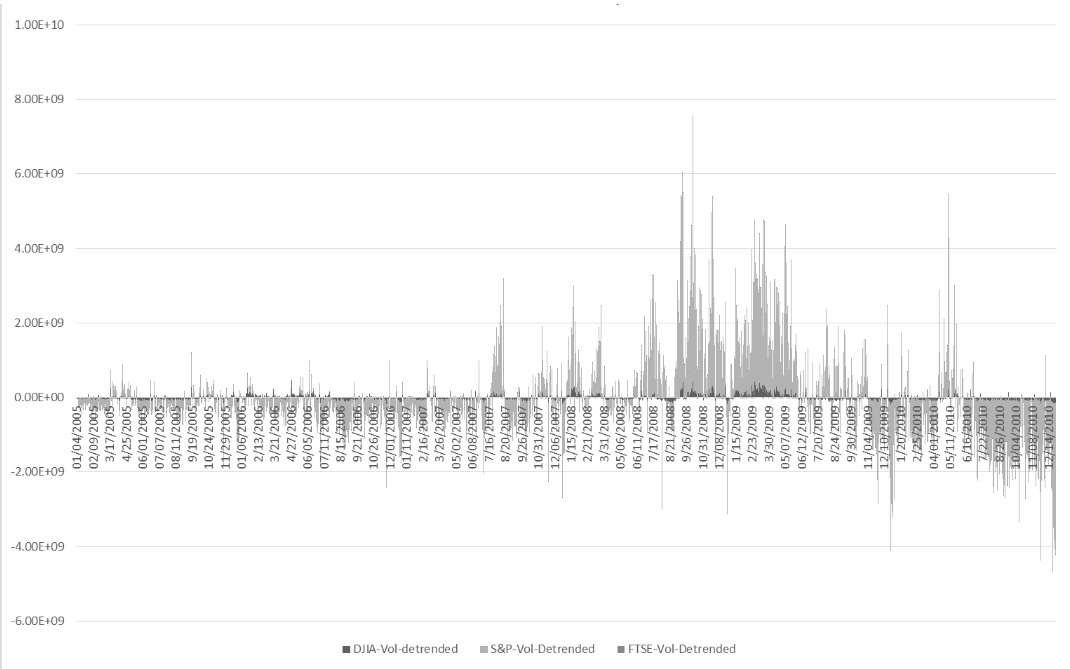
1.4.2 Volume of Trades

In addition to the information variables used in the variance equation, this essay follows Kalev et al. (2004) and others in considering the additional information contained in the volume of trades. Some of the models described in the next section include volume as an explanatory variable. As Kalev et al. (2004) point out, in the absence of a way to account for endogeneity of contemporaneous trading volume and volatility of stock returns, it is preferable to use a lagged volume variable to proxy for the flow of information. Additionally, graphical observation of the trading volume data shows that across all indices, there is a gradual increase over time, as also identified in previous contributions (Tauchen & Pitts, 1983; Andersen, 1996; Kalev et al., 2004). The volume of shares traded increases over time as the market participants and market capitalization grow. This can be removed as the growth trend in volume of shares traded is not of interest to the hypotheses of this essay. In order to remove this deterministic time trend, volume is de-trended using Ordinary Least Squares to regress trading volume on time (t). The coefficient associated with t is found to be significant at the 5% level for all three indices[18]; the residuals of the regression are considered the de-trended volume and are used in the rest of the analyses.

Figure 15 shows the de-trended trading volume for the FTSE, DJIA, and S&P 500 respectively. It can be seen there is a great deal more trading volume in the S&P 500 than the other two indices.

[18] Results are shown in Appendix 1.

Figure 15 De-trended Volume Data by Index



The de-trending has removed the gradual increase over time. The FTSE de-trended volume appears to be more clustered around the mean but also shows more frequent deviations from the mean compared to the US indices. The DJIA volume shows more clustering of volume than the FTSE as well as a significant increase in the range of values corresponding to the financial crises. The S&P 500 chart displays relatively low variations in volume until late 2007/early 2008 where the variations in trading volume appear larger and highly clustered, and seem to correspond to the timing of the latest financial crisis.

1.5 Empirical Methods

The focus of this essay is to explore the role of the flow of information as a determinant of stock market volatility. Thus, the method chosen to model the volatility of asset returns is of primary importance. As described in Section 2, the most prevalent model of return volatility is the GARCH model (Bollerslev, 1986). The popularity of the ARCH and GARCH models is explained by their ability to model time-varying volatility and volatility clustering. A large number of extensions to the GARCH model have emerged, as well as a huge literature on their empirical applications (for quality reviews, see Bollerslev et al., 1992; Bollerslev et al., 1994; Li et al., 2002), but the original GARCH model is still widely adequate and used to model volatility in financial times series (Kalev et al., 2004).

1.5.1 An Augmented GARCH Model

As suggested by theoretical and empirical contributions reviewed, the time-varying volatility is linked to the underlying process of information arrival. In other words, the conditional variance of returns could be determined by the flow of information arrival. This paper builds on this literature, by proposing a new proxy for information arrival that various types of information impact volatility differently.

To test this empirically, an augmented-GARCH specification is used, which introduces a new variable (or set of variables, depending on the hypothesis tested) to embody the flow of information in the conditional variance equation of a GARCH (1, 1) model shown in equation 2 (see Section 2).

In theory, according to efficient market theories (Fama, 1965), the arrival of information results in a change in asset prices, thus the persistence of volatility (or volatility clustering) should be explained by a serial correlation in news arrival itself. In other words, this can be tested by introducing a variable accounting for the information flow in the conditional variance equation: if this hypothesis is correct, then the coefficient associated with the additional term (s) in the conditional volatility equation should be statistically significant. Additionally, the ARCH and GARCH parameters $(\beta + \gamma)$, embodying volatility persistence after inclusion of the additional variable should decrease.

As stated previously, the choice of proxy to embody the flow of information is a major topic for debate in the literature (see Section 3). Instead of focusing on trading volume, this essay proposes a new proxy for the flow of information, collected specifically to broaden previous contributions (see Section 4).

The new proxy for the flow of information (I_t), embodied in the new information variables listed in Section 4, is introduced in the conditional variance in a GARCH (1,1) model, following a method proposed by Bomfim (2003) and Kalev et al. (2004), among others. The new conditional variance equation is as follows:

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \lambda \, I_t \qquad (3)$$

Where I_t represents a vector of variables capturing the flow of information. It will be replaced by the specific information variable (s) considered under each hypothesis described below (e.g. *Total-All*, *New-All*, *Total-Wires*, etc…)

Following Kalev et al. (2004), two major observations will be possible to be derived from this augmented GARCH model:

i) the significance of the coefficient associated with I_t reports whether the flow of information arrival is a determinant of volatility and,

ii) if the volatility clustering is explained by the information flow, then the GARCH(1,1) original coefficients $(\beta + \gamma)$should be reduced when using Eq. (3) compared to Eq. (2)[19].

Additionally, this essay further follows Kalev et al. (2004) and tests for an additional model where the trading volume is considered in the conditional variance equation. This allows accounting for unobserved elements of the flow of information in the proxy, such as private information. The trading volume is used as explanatory variable instead of the flow of information as in Eq. (4). It is lagged, as suggested in the literature to escape the issue of endogeneity of volatility and volume (as previously mentioned).

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \theta V_{t-1} \tag{4}$$

Finally, a third model is used which includes both the trading volume and the information flow as explanatory variables, so that the role of additional information captured in trading volume (e.g. private information) can be explored, as in Eq. (5).

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \lambda\, I_t + \theta V_{t-1} \tag{5}$$

The augmented-GARCH models described here are used as a basis to test a number of hypotheses detailed below. The “information flow” variable(s) added to the conditional variance equation will vary depending on the hypothesis under consideration.

[19] Although $\alpha + \beta$ is not expected to be reduced to zero, as this would imply that the proxy used in these analyses fully encompasses all information arrival influencing index returns, which is unrealistic.

1.5.2 Hypotheses

Three major hypotheses are tested in this essay, corresponding to the gaps identified in the literature in Section 1.3, as described below:

Hypothesis 1: *An Increase in flow of public information is associated with higher volatility in stock returns.*

This hypothesis proposes to test the theoretical relationship between the flow of information arrival and the volatility of returns (as stated in the Mixture of Distribution Hypothesis, Clark, 1973). It is tested using the augmented GARCH (1, 1) model, where the daily flow of total published information is included as an exogenous variable in the conditional variance equation.

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t \qquad (6)$$

where *Total-All* (TA) is the total daily flow of information collected through ProQuest[20]. The significance of the associated coefficient λ determines the extent to which the theoretical relationship holds given the proxy used in this essay.

Using the model in Eq. (6) suggests that volatility on a given day is determined by the information published on that day, but the proxy also considers the information published during any preceding non-trading days e.g. Saturday and Sunday information are captured in *Total-All* (Monday). This information "backlog" is likely to be incorporated in prices on the next trading day, thus the backlog of information is embodied in the *Total*-All variable.

[20] As previously stated, this model is also tested in presence of the lagged volume, but the equation is not repeated here.

If the volatility in index returns follows a similar pattern as the flow of public information as measured by this proxy; then one could conjecture that this pattern could be caused by the timing of information arrival. However, the inclusion of the backlog of information in the *Total-All* variable could in itself be the driving force behind the identified volatility-information relation. Therefore, it is important to investigate whether this relation is sustained when controlling for the existence of a backlog in the *Total-All* variable.

It is possible to formally test for the impact of this backlog of information on volatility. A dummy variable, representing whether or not there is a backlog on a given day (*Backlog*) is created and incorporated in the augmented GARCH alongside the flow of information, as in Eq. (7):

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \delta Backlog_t \qquad (7)$$

The significance of δ, the coefficient associated with the dummy variable, determines whether the presence of backlog information modifies the daily relationship between information flow and volatility. In other words, if δ is positive and significant, the volatility of returns is larger on days with an information backlog than others. If the sum of the original GARCH coefficients $(\beta + \gamma)$ is reduced after inclusion of the *Backlog Dummy*, volatility clustering can be interpreted as a consequence of accumulation of information during non-trading days, which is then incorporated into returns when the market opens, generating more volatility. Additionally, if the coefficient associated with the information flow variable remains significant after the inclusion of the *Backlog Dummy*, it will be acceptable to conclude that the relation is driven not only by this backlog effect.

Hypothesis 2: *The flow of public information impacts volatility differently depending on the information medium it is carried in.*

Hypothesis 2 addresses the question of whether the type of media embodying the information flow is a determinant of the information-volatility relationship. According to the literature, different categories of information have different impacts on volatility (Andersen, 1996; Kalev et al., 2004; Cousin & Launois, 2008). However, the influence of the type of media on the way information is incorporated into prices has not yet been explicitly, empirically tested.

As described in Section 4, the dataset collected in this essay enables the distinction between five media types, namely: *Wire-Feeds*, *Newspapers*, *Magazines*, *Trade Journals*, and *Reports*. Under this hypothesis, the type of media affects the way information is incorporated by investors into stock prices, and thus will show a different impact on stock market volatility

To test for this, the GARCH model is augmented with the total flow of information, (both with and without the *Backlog Dummy* variable), for each type of media recorded in the data (5 types). Hypothesis 2 is thus an extension of Hypothesis 1, but disaggregated by media type, as shown in Eq. (8):

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t} \tag{8}$$

Where $Total_{i,t}$ is the daily flow of information associated with a type of media is, where i=[*Wire-Feeds, Newspapers, Trade Journals, Magazines, Reports*].

Because stock market volatility can be impacted by different media types simultaneously, the above equation, which examines each type separately, could suffer from omitted variable bias. In order to address this concern, another regression is run where conditional volatility is explained by all media types simultaneously. This is shown in Eq. (9):

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5} \lambda_i \, Total_{i,t} \tag{9}$$

Where $Total_{i,t}$ is the daily flow of information associated with a type of media is, where i=[*Wire-Feeds*, *Newspapers*, *Trade Journals*, *Magazines*, *Reports*].

The coefficients associated with each media type, and their significance will determine whether various types of media are more or less associated with stock returns volatility. Preliminary results from Kalev et al. (2004) suggest that information published at higher frequency should have a larger impact on stock market volatility. Using this model, I can test for this explicitly, as well as testing for the joint explanatory powers of different media types.

Hypothesis 3: *Past increases in the flow of public information are associated with higher volatility today.*

This third hypothesis is intended to test whether there might be a lag in the way information is processed by investors and traders. In contrast to the previous hypothesis testing the MDH, this hypothesis is consistent with the sequential information arrival hypothesis, which proposes that information is processed sequentially by investors on the market. This leads to information having a lagged impact on volatility.

The first test of hypothesis consists of examining whether information accumulated during non-trading days can have an impact on volatility when the market re-opens. To test for this, Eq. (6) is modified so that the information variable (s) consider the information published on that given day (using *New-All* instead of *Total-All*), and the backlog information (using *Backlog-All*) separately. As shown in Eq. (10) below:

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \tau Backlog\_All_t + \delta Backlog_t \qquad (10)$$

To further test Hypothesis 3, the previous model can also be augmented to include lags of the total information variable. To represent a trading week, n lags are included in the model[21], as in Eq. (11)

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \lambda_1\, Total\_All_{t-1} + \cdots + \lambda_n\, Total\_All_{t-n} \qquad (11)$$

where λ_i , i =1...n embody the impact of past flows of information on the volatility of returns.

Finally, and more importantly, Hypothesis 3 can also be explored with respect to the joint explanatory powers of various media types, and their timing. Tests are conducted to explore whether the type of media the information is captured in can have an impact on how quickly the information is incorporated into stock prices and thus volatility.

For example, it is possible that media types published at higher frequency, or based on less regular publication schedule (e.g. wires) would have more contemporaneous

[21] The number of lags is tested through our estimations depending on the index used, to ensure the robustness of the findings.

impact on volatility than others. They could arguably be more likely to contain unexpected news, in comparison with more regular publications such as *Newspapers*.

Further, it is possible that some media types could contain complementary explanatory powers at various time horizons. For example, a *Magazine* publication could take a few days to digest by investors, while *Newspapers* would arguably be read daily and be incorporated in volatility on the same day. If this is the case, it could be possible to observe that volatility is explained contemporaneously by some media types, but with a lag for others.

This hypothesis is tested by extending the two previous tests (with *Backlog-All* and lags) to different media types in the conditional variance equation. Essentially, the models used are modifications of Eq. (10) and Eq. (11), to respectively include: *Backlog-All and Backlog Dummy,* and *n lags* for each of the five media types.

1.6 Results

This section presents the econometric results for the hypotheses tested in this essay using the new information proxy variables.

1.6.1 Volatility Persistence and Information Arrival

Tables 10, 11, and 12 summarize the results for testing Hypothesis 1 on the three major indices considered in this essay: the FTSE, S&P 500 and DJIA, respectively. The results confirm that an increase in the flow of public information (as measured by the information proxy created with ProQuest) has a positive impact on the volatility of index returns. For all indices considered, the coefficient associated with the *Total-All* variable is always positive and highly statistically significant in all cases (at the 1% level). This is shown in column B in each of the Tables (10,11, and 12**)** below.

The level of volatility persistence in the GARCH model is measured by the sum of the ARCH and GARCH coefficients ($\beta+\gamma$). In order to interpret the significance of the flow of public information as measured by this proxy, it is useful to compare the sum of ($\beta+\gamma$) in the simple GARCH (1, 1) in model A and the augmented conditional variance model B.

The results show that $\beta+\gamma$ is reduced when the *Total-All* variable is included in the variance equation, signifying that the level of volatility persistence is reduced when considering this new variable, thus the flow of information arrival is a determinant of volatility. However, it is important to note that the volatility persistence is not reduced to zero when including the information variable (the β and γ coefficients remain statistically significant). This could be due to a number of factors. First, the *Total-All* variable is a proxy and thus inherently imperfect. It will not be able to capture <u>all</u> public information. Additionally, trading can occur as a consequence of factors which are not attempted to be captured in this proxy. For example, it does not include private information (which has been linked to volatility) (French & Roll 1986; Vega, 2006), nor does it capture trading associated with non-information processes such as liquidity trading and portfolio balancing (Jegadeesh & Titman, 1993; Lee et al., 1994). Overall, these findings are consistent with findings in the literature (see Kalev et al., 2004) and can be considered supporting evidence in favour of the MDH, which posits a contemporaneous relationship between information flow and stock volatility.

As the literature suggests the role of trading volume[22] as a proxy for information, this is also tested in this essay. In Model C, the first lag of de-trended volume replaces

[22] Among others, Bollerslev & Domoqitz, 1993 find trading volume an adequate proxy for the flow of public information.

Total-All as an explanatory variable in the conditional variance equation. In Model D, the joint significance of the lagged volume and *Total-All* is tested. The results show that the lag of de-trended volume has little explanatory power in the conditional variance equation. The significance of *Total-All* remains strong, when the lagged trading volume is introduced in the model. This confirms the explanatory power of the *Total-All* proxy.

One important feature of the *Total-All* variable is that it captures the information published on a given day, but it also aggregates it from any preceding non-trading days. This is the case for any day following a non-trading day, including weekends, bank holidays, etc. It is possible that the very definition of this variable could create a pattern in the flow of information that would have an impact on the volatility information relationship.

To test for this, I include a dummy indicating the presence of a backlog in the *Total-All* variable as shown in (Model E). For the DJIA and the S&P 500, the relation between the flow of public information (as measured by *Total-All*) and volatility in index returns remains statistically significant at the 5% level, indicating that the backlog impact is not the sole driver of the volatility information relation. Using the FTSE results, the information volatility relationship remains significant at the 10% level. The level of volatility persistence ($\beta+\gamma$) is reduced in all three indices when introducing the *Backlog Dummy*. This suggests the information accrued during non-trading days is a driver of volatility persistence[23].

[23] This issue of information backlog and its impact on Stock Market price and volatility is explored in great detail in the third essay of this PhD thesis.

In terms of the fit of the model, B and E seem to be preferable specifications and include respectively *Total*-All, or *Total-All* and *Backlog*. Overall, these results are satisfying, and largely in line with the literature. The new proxy for the flow of public information is found to be a significant explanatory variable in modelling conditional volatility across the three indices. This proxy expands on the previous literature by increasing the breadth of the information captured. Additionally, by testing for the impact of backlog (through a dummy), the essay confirms that the information backlog accrued during non-trading days is a driver of volatility persistence.

Table 10 FTSE Hypothesis 1 Results

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance Equation

Model (A) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

Model (B) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t$

Model (C) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \theta V_{t-1}$

Model (D) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \theta V_{t-1}$

Model (E) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \delta Backlog_t$

Equation	Variable	A	B	C	D	E
Returns	ω	0.000	0.001	0.001	0.001	0.001
		(0.006)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
	μ	-0.0579112	-0.057	-0.058	-0.057	-0.057
		(0.039)**	(0.04)**	(0.04)**	(0.04)**	(0.05)**
Volatility	β	0.121	0.123	0.120	0.123	0.122
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.875	0.869	0.877	0.868	0.869
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	λ		0.054		0.055	0.051
			(0.01)***		(0.00)***	(0.08)*
	θ			0.000	-0.000	
				(0.93)	(0.64)	
	δ					0.212
						(0.91)
	α	1.24e-06	-14.157	-13.593	-14.217	-14.175
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		4690.97	4,692.93	4,684.17	4,686.10	4,692.93
Prob>chi2		0.038	0.00	0.00	0.00	0.00
Number of obs		1,486	1,486	1,484	1,484	1,486

p<0.1; ** *p*<0.05; *** *p*<0.01

Table 11 S&P 500 Hypothesis 1 Results

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance Equation Model (A) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

Model (B) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t$

Model (C) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \theta V_{t-1}$

Model (D) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \theta V_{t-1}$

Model (E) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \delta Backlog_t$

Equation	Variable	A	B	C	D	E
Returns	ω	0.001	0.001	0.001	0.001	0.001
		(0.01)**	(0.01)**	(0.01)**	(0.01)**	(0.01)**
	μ	-0.095	-0.096	-0.095	-0.096	-0.096
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Volatility	β	0.089	0.088	0.083	0.083	0.089
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.901	0.899	0.907	0.904	0.898
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	λ		0.049		0.040	0.043
			(0.00)***		(0.02)**	(0.01)**
	θ			0.000	0.000	
				(0.15)	(0.23)	
	δ					0.675
						(0.62)
	α	1.44e-06	-14.076	-13.535	-14.002	-14.211
		(0.000)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		4672.936	4,674.12	4,668.38	4,669.07	4,674.22
Prob>chi2		0.002	0.00	0.00	0.00	0.00
Number of obs		1,486	1,486	1,484	1,484	1,486

* p<0.1; ** p<0.05; *** p<0.01

Table 12 DJIA Hypothesis 1 Results

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance Equation

Model (A) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

Model (B) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t$

Model (C) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \theta V_{t-1}$

Model (D) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \theta V_{t-1}$

Model (E) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \delta Backlog_t$

Equation	Variable	A	B	C	D	E
Returns	ω		0.001	0.001	0.001	0.001
			(0.01)***	(0.01)***	(0.01)***	(0.01)***
	μ		-0.083	-0.082	-0.083	-0.083
			(0.01)***	(0.01)***	(0.01)***	(0.01)***
Volatility	β		0.089	0.087	0.086	0.089
			(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ		0.897	0.901	0.899	0.897
			(0.00)***	(0.00)***	(0.00)***	(0.00)***
	λ		0.042		0.037	0.037
			(0.01)***		(0.03)**	(0.04)**
	θ			0.000	0.000	
				(0.23)	(0.31)	
	δ					0.677
						(0.64)
	α		-13.998	-13.488	-13.945	-14.160
			(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood			4,770.02	4,763.24	4,763.75	4,770.12
Prob>chi2			0.00	0.00	0.00	0.00
Number of obs			1,486	1,484	1,484	1,486

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

1.6.2 Information by Media Type

The tables below summarize the estimation results of the second hypothesis, which explores the impact of different types of media on index volatility. As described in Section 4, the daily number of publications is disaggregated into 5 sub-categories, each representing a different type of media. Through this disaggregation, it is possible to investigate whether different media types have varying impacts on stock market volatility. While previous contributions have explored the varying impacts of different information content, this essay enables us to examine whether the type of media information is carried in can affect the way investors digest this information.

Tables 13 through 15 summarise the results of several models examining the impact of each of the five media categories separately, and then jointly. Comparing Models A through E in the FTSE, S&P 500, and DJIA, Total-Wires have a positive impact on index volatility, and are highly statistically significant (at the 1% level for all indices). *Total-Newspapers* also positively impacts stock market activity, although less strongly. The coefficient is statistically significant at the 1% level for FTSE, at the 5% level for S&P 500, and at the 10% level for DJIA. In comparison, *Total-Reports*, *Total-Magazines*, and *Total-Trade Journals* do not show statistically significant results.

Models F through J examine augments from the previous models with the *Backlog Dummy*, in order to determine whether the relation found previously is mainly driven by the backlog effect. For all three indices, the results show that *Total-Wires* and *Total-Newspapers* remain statistically significant explanatory variables after the inclusion of the Dummy, which is found to be insignificant. However, Backlog is found significant in Model G when paired with Total-*Trade Journals*, which also becomes statistically significant, in the S&P 500 and DJIA only. This suggests that the Backlog in *Trade Journals* tends to be absorbed into prices when markets reopen following non-trading days. Further, the *Backlog Dummy* is also found to be highly statistically significant when paired with *Total-Reports* in the S&P 500 only. Overall, it appears that media types which are typically more concise, such as *Wire-Feeds* and *Newspapers* tend to have larger contemporaneous impacts on stock market volatility. In addition, they do not seem to exhibit particular impacts of *Backlog*. However, media types which are typically more lengthy or technical in nature, for example Trade Journals or Analyst Reports do not appear to significantly impact volatility, but seem to exhibit a backlog effect. One potential explanation for this finding would be that the period for digesting the information is dependent on the media type. This will be explored further with the use of lags when testing for Hypothesis 3 below.

By investigating the impact of each media type separately, it is possible that the regressions would suffer from omitted variable bias. Consequently, to address this issue a further regression is run in Model K which includes all media types jointly. When comparing the log-likelihood of Model K with the previous model including *Total-All*, the fit of the model has improved, suggesting that a disaggregation of information can provide additional insights into stock market volatility. The coefficient of Total-Wires remains significant at the 10% level for the S&P 500 and FTSE only, suggesting that previous results might reflect some degree of omitted variable bias. This is definitely the case for *Newspapers*, which is no longer significant. However, *Total-Trade Journals* becomes statistically significant at the 1% level for the FTSE and 10% level for the S&P 500 and DJIA.

The robustness of these results is further examined by adding the lagged trading volume to the augmented model (Model L). The results hold across the three indices, despite the inclusion of the volume variable, which is found to be insignificant. Further, the log-likelihood result worsens with the inclusion of the trading volume variable.

Overall, these results suggest that *Wire-Feeds* show the strongest contemporaneous relation to stock market volatility, while *Trade Journals* could also be a significant explanatory variable. Surprisingly, *Newspapers* lose statistical significance in the model which combines all media types. This could suggest a high correlation between *Newspapers* and one of the other information variables. Correlations between the variables by media types are presented in Table 13 below. The results show a high correlation between the number of *Newspapers* and *Wire-Feeds* in particular, as well as a high correlation between *Magazines* and *Trade Journals*. These results could suggest two things: first, that these two media types could be published at the same time coincidentally or that they could be published to report the same piece of information.

Table 13 Information Variables Correlation Matrix

Variable	*Total-Trade Journals*	*Total-Reports*	*Total-Magazines*	*Total-Wire*	*Total Newspapers*
Total-Trade Journals	1				
Total-Reports	0.3771	1			
Total-Magazines	0.8986	0.4277	1		
Total-Wire	0.3174	0.0972	0.2557	1	
Total Newspapers	0.2891	0.0921	0.2356	0.844	1

Interestingly, these results appear in line with previous contributions to the literature (Kalev et al., 2004; Cousin & Launois, 2008) examining the relation between information and volatility: the publication with the highest frequency of publication (in this case *Wire-Feeds*), is more likely to have a significant impact on stock market volatility. It is likely that this is explained by the timeliness of the information released (i.e. more frequently released information is likely to be more relevant to daily market activity) and the digestibility of the medium (e.g. short *Wire-Feeds* are easier to be incorporated into trading decisions than long magazine articles).

Thus far, the essay has modelled the volatility information relation as contemporaneous. However, as suggested in the literature review, it is possible that information released previously could have a lagged impact on stock market volatility. This lag relation has not yet been explored with respect to different types of publications. This is the focus of Hypothesis 3 descripted in the next section.

Table 14 Hypothesis 2 Test Using FTSE Data

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance Equation

Models (A) to (E) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t}$

Model (F) to (J) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t} + \delta Backlog_t$

Model (K) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5} \lambda_i\, Total_{i,t}$

Model (L) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5} \lambda_i\, Total_{i,t} + \theta V_{t-1}$

Equation	Variable	A	B	C	D	E	F	G	H	I	J	K	L
Returns	ω	0.001 (0.00)***	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.00)***	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.00)***	0.001 (0.00)***	0.001 (0.00)***
	μ	-0.058 (0.04)**	-0.057 (0.04)**	-0.058 (0.04)**	-0.057 (0.04)**	-0.058 (0.04)**	-0.057 (0.04)**	-0.057 (0.04)**	-0.058 (0.04)**	-0.058 (0.04)**	-0.057 (0.04)**	-0.055 (0.05)*	-0.054 (0.06)*
Conditional Variance	β	0.121 (0.00)***	0.123 (0.00)***	0.123 (0.00)***	0.123 (0.00)***	0.123 (0.00)***	0.121 (0.00)***	0.123 (0.00)***	0.123 (0.00)***	0.118 (0.00)***	0.123 (0.00)***	0.117 (0.00)***	0.114 (0.00)***
	γ	0.866 (0.00)***	0.874 (0.00)***	0.873 (0.00)***	0.874 (0.00)***	0.869 (0.00)***	0.865 (0.00)***	0.873 (0.00)***	0.873 (0.00)***	0.877 (0.00)***	0.869 (0.00)***	0.867 (0.00)***	0.869 (0.00)***
	λ (Total-Wires)	0.196 (0.00)***					0.173 (0.01)***					0.166 (0.07)*	0.159 -0.11
	λ (Total-Tradejournal)		0.047					0.033				0.167	0.165

		-0.46					-0.59				(0.00)***	(0.00)***
λ (Total-Reports)			0.227					0.198			-2.662	-2.349
			(0.08)*					-0.16			-0.41	-0.48
λ (Total-Magazines)				0.266					-6.947		-0.375	-0.408
				-0.15					-0.17		-0.31	-0.33
λ (Total-Newspapers)					0.626 (0.01)***					0.529	0.224	0.232
										(0.06)*	-0.63	-0.61
δ						1.376	0.996	0.738	3.825 (0.04)**	1.034		
						-0.45	-0.65	-0.74		-0.61		
θ												0
												-0.73
α	-14.463 (0.00)***	-13.825 (0.00)***	-13.607 (0.00)***	-13.816 (0.00)***	-14.23 (0.00)***	-14.947 (0.00)***	-14.013 (0.00)***	-13.763 (0.00)***	-12.153 (0.00)***	-14.497 (0.00)***	-14.759 (0.00)***	-14.733 (0.00)***
Log-Likelihood	4,695.43	4,691.22	4,691.41	4,691.18	4,692.78	4,695.98	4,691.27	4,691.43	4,691.79	4,692.90	4,699.82	4,692.89
Prob>chi2	0	0	0	0	0	0	0	0	0	0	0	0
Number of obs	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,484

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 15 Hypothesis 2 Test Using S&P 500 Data

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance Equation

Models (A) to (E) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t}$

Model (F) to (J) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t} + \delta Backlog_t$

Model (K) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5}\lambda_i\, Total_{i,t}$

Model (L) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5}\lambda_i\, Total_{i,t} + \theta V_{t-1}$

Equation	Variable	A	B	C	D	E	F	G	H	I	J	K	L
Returns	ω	0.001 (0.01)***	0.001 (0.01)**	0.001 (0.01)**	0.001 (0.01)**	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.00)***	0.001 (0.01)***	0.001 (0.01)**	0.001 (0.01)***	0.001 (0.01)***	0.001 (0.01)***
	μ	-0.095 (0.00)***	-0.094 (0.00)***	-0.095 (0.00)***	-0.095 (0.00)***	-0.096 (0.00)***	-0.095 (0.00)***	-0.093 (0.00)***	-0.095 (0.00)***	-0.095 (0.00)***	-0.096 (0.00)***	-0.094 (0.00)***	-0.096 (0.00)***
Conditional Variance	β	0.088 (0.00)***	0.09 (0.00)***	0.088 (0.00)***	0.086 (0.00)***	0.089 (0.00)***	0.088 (0.00)***	0.093 (0.00)***	0.087 (0.00)***	0.087 (0.00)***	0.089 (0.00)***	0.09 (0.00)***	0.086 (0.00)***
	γ	0.898 (0.00)***	0.901 (0.00)***	0.901 (0.00)***	0.902 (0.00)***	0.899 (0.00)***	0.898 (0.00)***	0.905 (0.00)***	0.905 (0.00)***	0.901 (0.00)***	0.899 (0.00)***	0.9 (0.00)***	0.903 (0.00)***
	λ (Total-Wires)	0.177 (0.00)***					0.169 (0.00)***					0.187 (0.05)*	0.184 (0.07)*
	λ (Total-Tradejournal)		-0.279 -0.3					-0.778 (0.01)**				-0.545 (0.06)*	-0.503 (0.07)*

λ (Total-Reports)			0.197 -0.12					0.145 -0.1			0.396 -0.84	0.358 -0.89
λ (Total-Magazines)				-5.446 -0.11					-5.042 -0.12		0.391 (0.10)*	0.37 -0.11
λ (Total-Newspapers)					0.473 (0.02)**					0.446 (0.05)**	0.383 -0.44	0.366 -0.45
δ						0.931 -0.49	4.288 (0.00)***	34.329 (0.00)***	2.037 -0.37	0.496 -0.73		
θ												0 -0.5
α	-14.505 (0.00)***	-12.624 (0.00)***	-13.486 (0.00)***	-12.174 (0.00)***	-13.996 (0.00)***	-14.888 (0.00)***	-12.871 (0.00)***	-46.312	-12.269 (0.00)***	-14.127 (0.00)***	-13.784 (0.00)***	-13.845 (0.00)***
Log-Likelihood	4,676.90	4,673.95	4,673.22	4,673.95	4,674.10	4,677.18	4,677.24	4,672.13	4,674.02	4,674.14	4,679.50	4,674.39
Prob>chi2	0	0	0	0	0	0	0	0	0	0	0	0
Number of obs	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,484

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 16 Hypothesis 2 Test on DJIA Data

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance Equation

Models (A) to (E) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t}$

Model (F) to (J) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda_i Total_{i,t} + \delta Backlog_t$

Model (K) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5} \lambda_i \, Total_{i,t}$

Model (L) $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \sum_{i=1}^{5} \lambda_i \, Total_{i,t} + \theta V_{t-1}$

Equation	Variable	A	B	C	D	E	F	G	H	I	J	K	L
Returns	ω	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
		(0.01)***	(0.00)***	(0.01)***	(0.01)***	(0.00)***	(0.01)***	(0.00)***	(0.01)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***
	μ	-0.082	-0.082	-0.082	-0.083	-0.082	-0.081	-0.078	-0.082	-0.083	-0.081	-0.081	-0.081
		(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***	(0.01)***
Conditional Variance	β	0.088	0.09	0.089	0.088	0.09	0.088	0.09	0.09	0.089	0.088	0.088	0.078
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.898	0.899	0.899	0.899	0.897	0.898	0.905	0.899	0.899	0.902	0.901	0.906
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	λ (Total-Wires)	0.146					0.142					0.162	0.164
		(0.00)***					(0.01)**					-0.11	-0.14
			-0.18					-0.622				-0.472	0.108

λ (Total-Tradejournal)		-0.4					(0.01)**				(0.09)*	(0.02)**
λ (Total-Reports)			0.167					0.162			0.305	0.297
			-0.23					-0.25			-0.89	-0.3
λ (Total-Magazines)				-3.608					-2.62		0.409	-0.825
				-0.3					-0.33		(0.08)*	-0.4
λ (Total-Newspapers)					0.363					0.283	0.338	-0.229
					(0.07)*					(0.07)*	-0.51	-0.64
δ						0.622	3.614	0.277	1.62	27.377		
						-0.66	(0.00)***	-0.88	-0.42			
θ												0
												-0.17
α	-14.367	-12.888	-13.495	-12.522	-13.9	-14.594	-12.941	-13.556	-12.817	-39.932	-13.796	-14.305
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	4,771.97	4,770.07	4,769.52	4,770.14	4,770.08	4,772.07	4,771.46	4,769.52	4,770.24	4,768.73	4,774.15	4,766.76
Prob>chi2	0	0	0	0	0	0	0	0	0	0	0	0
Number of obs	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,486	1,484

* $p<0.1$;
**
$p<0.05$;
***
$p<0.01$

1.6.3 Temporal Relationship between Information and Volatility

In this section, the third hypothesis of this essay is tested to explore whether the relation between the flow of public information and stock market volatility is contemporaneous or could be characterised by a delay. The first part of testing this hypothesis is linked to the role of the backlog of information during non-trading days. The second part will review the role of information published during previous days (t-n) through the use of lags.

1.6.3.1 The Role of Information Backlog

Table 16 presents the results of Eq. (10) where *Total-All*, *Backlog-All*, and *Backlog* (*Dummy*) are used as explanatory variables in the conditional variance equation. The results are presented for the FTSE, S&P 500, and DJIA in the same table. Using this specification, the coefficient associated with *Backlog-All* represents the explanatory power of the backlog of all information on volatility upon the markets reopening, while the coefficient associated with the *Total-All* actually represents the explanatory power of information released on that day only.

The results suggest that for the S&P 500, the *Backlog* of information appears to matter in explaining volatility (significant at the 1% level). In contrast the *Backlog-All* variable is not significant in either the FTSE or the DJIA. This can be further investigated using the disaggregated data variables to explore whether different media types matter more in terms of *Backlog* or in terms of new information. The results of Columns B, D, and F in Table 16, suggest that backlog of information is an important explanatory variable when looking at *Magazines* (significant at the 1% level), *Trade Journals* (at least at the 10% level), and *Reports* for FTSE only (at the 10% level).

Table 17 Information Backlog

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \lambda\, Total\_All_t + \tau Backlog\_All_t + \delta Backlog_t$

Equation	Variable	FTSE	S&P 500	DJIA
Returns	ω	0.001	0.001	0.001
		(0.01)***	(0.01)**	(0.01)***
	μ	-0.056	-0.095	-0.083
		(0.05)**	(0.00)***	(0.01)***
Conditional Variance	β	0.126	0.073	0.089
		(0.00)***	(0.00)***	(0.00)***
	γ	0.865	0.913	0.897
		(0.00)***	(0.00)***	(0.00)***
	λ	0.07	0.033	0.021
		(0.03)**	-0.46	-0.51
	τ	-0.042	5.387	0.026
		-0.46	(0.00)***	-0.55
	δ	0.763	-323.5	0.495
		-0.6		-0.77
	α	-14.511	-13.654	-13.935
		(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		4,693.37	4,678.71	4,770.39
Prob>chi2		0	0	0
Number of obs		1,486	1,486	1,486

* p<0.1; ** p<0.05; *** p<0.01

The fit of the model has improved when including the *Backlog* of these publications, which empirically supports the hypothesis posited from the literature that information might accumulate during non-trading days and be incorporated in prices when markets re-open. The impact of information *Backlog* is an important consideration in the information volatility relation and will be further explored in a subsequent essay in this thesis.

1.6.3.2 The Role of Information Lags

Table 18 Lagged *Total-All* Variable

Returns Equation $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Conditional Variance $\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \lambda\, Total\_All_t + \lambda_1\, Total\_All_{t-1} + \cdots + \lambda_n\, Total\_All_{t-n}$

Equation	Variable	FTSE	S&P 500	DJIA
Returns	ω	0.001	0.001	0.001
		(0.01)***	(0.01)**	(0.01)***
	μ	-0.056	-0.095	-0.082
		(0.05)**	(0.00)***	(0.01)***
Conditional Variance	β	0.122	0.088	0.09
		(0.00)***	(0.00)***	(0.00)***
	γ	0.867	0.897	0.894
		(0.00)***	(0.00)***	(0.00)***
	λ	0.033	-0.029	-0.021
		-0.49	-0.74	-0.8
	λ_1	0.04	0.048	0.046
		-0.51	(0.00)***	(0.02)**
	$\lambda 2$	-0.022	-0.04	-0.052
		-0.86	-0.78	-0.76
	λ_3	0.045	0.045	0.048
		-0.33	-0.17	(0.06)*
	λ_4	-0.041	-0.05	-0.057
		-0.79	-0.71	-0.7
	λ_5	0.032	0.055	0.047
		-0.53	(0.01)***	(0.05)**
	α	-14.657	-13.986	-13.779
		(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		4,675.05	4,660.03	4,755.54
Prob>chi2		0	0	0
Number of obs		1,481	1,481	1,481

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 18 presents the results of Eq. (11) where 5 lags of the flow of public information are tested as explanatory variables for the conditional variance, in addition to the contemporaneous information flow. The results show that the *Total-All* variable (contemporaneous) loses its significance, which is a clear sign of multi-collinearity amongst information flow and its lagged variables. This suggests caution in interpreting the results for each individual coefficient, however, a number of general conclusions can be drawn from the table. In particular the first and fifth lag of information flow appear significant for the two US indices (at least at the 5%) while the third lag is significant only for the DJIA (at the 10% level). It is important to note that any time a coefficient is found significant, it also displays the expected positive sign.

The significance of the lags can be further tested using a joint deletion test. The results are summarised in Table 18. For the FTSE the joint deletion of the 5 lags confirms that I fail to reject the null that their coefficients are statistically significantly different from zero. This seems to point towards the notion that the relation between the flow of information and volatility is contemporaneous. In other words, information released on previous trading days does not significantly impact volatility on consecutive days. However, for the S&P 500 and DJIA, the joint deletion test for all five lags, allows for the rejection of the null, suggesting some of the lags are significant as shown in the results for Model B and C. Further, joint deletion-tests for the lags which are not found significant in the previous regression (lag 2-4 in S&P 500 and lags 2&4 for DJIA), alongside the Tot*al-All* variable, confirm that for the US indices the relation between volatility and information might not be only contemporaneous. In particular, information released during the previous trading day or day t-5 (1 week before) appears to significantly impact today's volatility. This is an interesting result particularity given the differences between the US and European indices. These results suggest that the contemporaneous or sequential hypothesis are not necessarily mutually exclusive. The Mixture of Distribution Hypothesis (Clark, 1973) which suggests that traders receive and process information simultaneously looks to be supported by the FTSE results, while the Sequential Information Hypothesis (Copeland, 1976) which suggests that information is processed by traders and incorporated into prices sequentially, is supported by the S&P 500 and DJIA results. Given the daily aggregated characteristics of the datasets used in this essay, it is impossible to further explore the SIA or MDH in the level of detail required. A formal analysis of this nature could be accomplished with higher frequency intra-day data, where intra-day volatility patterns could be related to intra-day information patterns which would provide more information about the types of traders as well as their trading behaviour in regard to processing of information.

1.6.4 Temporal Relation by Media Type

Table 19 present the results of further analysis of the temporal relation of volatility and each media type[24]. The results show a uniform pattern across the different indices, with the strongest results associated with *Total-Trade Journals*, and *Total-Wire-Feeds*. However, once again, there appears to be evidence of multi-collinearity between the information flows by media type, which reduces the validity of the coefficient and their significance. This suggests a potential area for future research, in examining the added value of each information type, and their lags in explaining stock market volatility.

[24] Due to issues of convergence several coefficients are missing for the S&P 500, as well as Reports from the DJIA.

Table 19 Hypothesis 3 Results Temporal Relation by Media Type

Hypothesis 3 Results

Equation	Variable	FTSE	S&P 500	DJIA
Returns	ω	0.001	0.001	0.001
		(0.00)***	(0.00)***	(0.00)***
	μ	-0.053	-0.08	-0.077
		(0.05)*	(0.00)***	(0.00)***
Conditional Variance	β	0.099	-0.024	0.044
		(0.00)***	(0.00)***	(0.00)***
	γ	0.869	0.997	0.943
		(0.00)***	(0.00)***	(0.00)***
	λ0 (Total-Newspapers)	1.637	-15.234	6.074
		(0.01)**		(0.00)***
	λ1 (Total-Newspapers)	-1.615	-2.183	24.631
		(0.03)**		(0.00)***
	λ2 (Total-Newspapers)	2.943	-1.139	-6.078
		(0.00)***		(0.00)***
	λ3 (Total-Newspapers)	-0.959	6.041	-12.663
		-0.11		(0.00)***
	λ4 (Total-Newspapers)	-2.764	-2.168	-7.866
		(0.02)**		(0.01)***
	λ5 (Total-Newspapers)	-0.199	3.932	-11.011
		-0.63	(0.00)***	(0.00)***
	λ0 (Total-Magazines)	-0.98	2.907	0.546
		(0.01)***		(0.02)**
	λ1 (Total-Magazines)	-0.316	-5.702	-1.268
		-0.57		(0.02)**
	λ2 (Total-Magazines)	-1.524	0.054	5.546
		(0.00)***		(0.00)***
	λ3 (Total-Magazines)	0.904	1.074	-4.012
		(0.00)***	(0.00)***	(0.00)***
	λ4 (Total-Magazines)	-2.792	-0.625	5.513
		(0.05)**	(0.00)***	(0.00)***
	λ5 (Total-Magazines)	-0.559	-0.157	-65.795
		-0.34		(0.00)***
	λ0 (Total-Reports)	0.455	-0.337	
		(0.01)***	-0.16	
	λ1 (Total-Reports)	-0.513	4.547	
		-0.32	(0.00)***	
	λ2 (Total-Reports)	-0.975	-5.214	
		(0.03)**		
	λ3 (Total-Reports)	1.768	10.45	
		(0.00)***		
	λ4 (Total-Reports)	-1.581	-20.016	
		(0.02)**		
	λ5 (Total-Reports)	-0.023	2.003	
		-0.98	(0.00)***	

	λ0 (Total-Trade Journals)	0.184	-0.809	-0.462
		(0.00)***	(0.00)***	(0.06)*
	λ1 (Total-Trade Journals)	-0.069	0.854	0.666
		-0.52	(0.00)***	(0.00)***
	λ2 (Total-Trade Journals)	0.409	0.716	-1.354
		(0.00)***	(0.00)***	(0.00)***
	λ3 (Total-Trade Journals)	-0.207	-0.301	1.251
		(0.00)***	(0.00)***	(0.00)***
	λ4 (Total-Trade Journals)	0.14	0.528	-1.44
		(0.01)***	(0.00)***	(0.00)***
	λ5 (Total-Trade Journals)	-0.06	0.057	0.872
		-0.47	(0.00)***	(0.00)***
	λ0 (Total-Wire)	-0.029	3.036	-0.062
		-0.86	(0.00)***	-0.87
	λ1 (Total-Wire)	0.391	-0.985	-4.995
		(0.01)***		(0.00)***
	λ2 (Total-Wire)	-1.265	-6.691	1.709
		(0.00)***		(0.00)***
	λ3 (Total-Wire)	0.314	0.443	2.057
		(0.01)***	(0.00)***	(0.00)***
	λ4 (Total-Wire)	-0.066	2.141	2.598
		-0.66		(0.00)***
	λ5 (Total-Wire)	0.494	1.607	-0.696
		(0.00)***	(0.00)***	-0.22
	α	-12.375	-23.298	-19.316
		(0.00)***		(0.00)***
Log-Likelihood		4,704.29	4,728.38	4,797.34
Prob>chi2		0	0	0
Number of obs		1,481	1,481	1,481

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

1.7 Conclusion

This essay has examined the relation between stock market volatility and the flow of public information, building onto two inter-related parts of the financial economics literature: First, past research has demonstrated the need for improving proxies to capture public information arrival in the stock market. Second, the empirical literature on stock market volatility has established the primary role of information arrival, however, several questions remain relating to the timing of this relation, as well as the way different categories of information may impact stock market volatility. In order to address these questions, the essay has used returns for three different indices across two countries (the UK FTSE and the U.S. Dow Jones, and S&P 500).

In order to explicitly relate information flow to index returns volatility, the ProQuest search engine is used to derive a proxy for the flow of public information. This is done by recording the daily volume of publications in the business-focussed database. This data collection exercise in itself presents the first innovation in this essay, as it expands the definition of public information, through a novel use of ProQuest ABI database. Further this essay adds to the literature, as to the author's knowledge, it is the first attempt at disaggregating the flow of public information by media type. The daily flow of public information is thus recorded for all publication types and disaggregated to account for publications in the forms of *Wire-Feeds*, *Newspapers*, *Magazines*, *Trade Journals*, and *Reports*. Where previous contributions have classified information by types of announcement, this is the first time that the impact of information flow can be tested in relation to the type of media it is carried in. This enables further analysis into public information's impact on volatility persistence.

The relation between this newly created information flow proxy variables and volatility in index returns is then analysed using an augmented GARCH model, aiming to test for three hypotheses: (a) whether, using the newly created proxy, the flow of information can contribute to explaining volatility in the three index returns, (b) whether different types of information media impact volatility differently and (c) whether the relation between information flow and volatility is contemporaneous or whether volatility can be affected by information released in previous days. The final hypothesis is threefold, as it explores the role of information backlog (from previous non-trading days) on volatility when the market re-opens, the role of lags in total information flow, and also explores whether these relation vary across media types.

Overall, the results confirm that an increase in the flow of public information (as measured by the information proxy created with ProQuest) has a positive impact on the volatility of index returns. For all indices considered, the coefficient associated with the *Total-all* variable is always positive and highly statistically significant in all cases (at the 1% level), further confirmed after controlling for the effects of trading volume. These findings are in line with previous contributions and suggest that magnitude of index return volatility is dependent on the rate of information arrival.

Further, when examining the role of information by media type on volatility (i.e. whether the type of media information is carried in can affect the way investors digest this information into prices), the results are generally mixed for the various media types. The results suggest that the media which deliver the most current information, i.e. *Wire-Feeds*, tend to have the largest impact on returns volatility. In contrast, information carried through media with often less frequent publication dates, (i.e. *Reports*, *Magazines*) which potentially contain less up-to-date information to market participants, shows a lesser impact on index returns. Interestingly, these results appear in line with previous contributions to the literature (Kalev et al., 2004; Cousin & Launois, 2008) examining the relation between information and volatility: the publication with the highest frequency of publication (in this case *Wire-Feeds*), is more likely to have a significant impact on stock market volatility. It is likely that this is explained by the timeliness of the information released (i.e. more frequently released information is likely to be more relevant to daily market activity) and the digestibility of the medium (e.g. short *Wire-Feeds* are easier to be incorporated into trading decisions than long magazine articles). Overall, the results of hypothesis 2 tests suggest that wire-feeds show the strongest contemporaneous relation to stock market volatility, while trade journals could also be a significant explanatory variable, and for the latter particularly through a backlog of information.

Finally, through examining the temporal relation between volatility and information flow, several conclusions can also be drawn. Firstly, through the construction of a backlog variable which records the number of publications during non-trading, it is possible to explicitly tests for the impact of information backlog on volatility when markets re-open. The results suggest that for the S&P 500, the *Backlog* of information appears to matter in explaining volatility, it is not significant in either the FTSE or the DJIA. This issues will be further explored using a larger number of indices in the third Essay of this thesis. Secondly, the inclusion of lagged values for the flow of information in the conditional volatility equation allow for further testing of the temporal relation. The significance of certain lags is confirmed through the analysis, but only for the U.S. indices. In the FTSE, there is no evidence of past lags affecting current volatility in the sample period analysed. There are a number of possible explanations for these results, but in general, the information captured using the ProQuest proxy could be of less relevance to the UK than to the U.S. markets (e.g. most of the information could be related to US traded stocks).These differences could also be due to differences between the indices themselves, both in terms of the composition of the index, and the behaviour of market participants, as explained by cultural differences. Thirdly, the use of lags is extended to the various media types, the findings are less conclusive. While the results suggest issues of multi-collinearity in explanatory variables, lags of Wire-feeds and Trade-Journals are still found to be the most significant explanatory variables of index volatility. Although the results for Wire-feeds and Trade-Journals appear consistent across the indices, it is not necessarily the case with lags of other media types. Overall these results provide interesting insight by the disaggregation of information by media type, and suggest a strong venue for future research using intra-day data. If the timing or type of publication of information has an impact on volatility, it is likely that given the speed at which information is published in the digital age, data providing detail about the time of day information is published might provide additional insights into the relation. This is of particular interest when comparing indices operated across different time zones. It is likely that information released at certain time on a given day is incorporated in different ways in these indices, depending on whether it is published during trading or non-trading hours in that location.

The research presented in this essay could be extended in a number of ways. First, the construction of the information proxy could be improved through continuing the intensive data collection process: the daily count of publication could be gathered through a wider range of databases; more numerous tests could be run to determine which publications should be included and could bias the analysis; etc. The preliminary results of this essay are conclusive in relating the information flow and volatility and suggest that further work on the proxy itself could further contribute to this literature.

Ultimately, if the data collection process could be automated, it would also be useful to gather this information proxy at higher frequency. While daily data provide valuable insight in the relation with volatility, it is possible that intra-day data would present more conclusive results, particularly when looking at the impact of information by media type. For example, more drastic differences could be highlighted between wire-feeds, which are released continuously at various times of day, and newspapers, which are generally published at regular time before the market opens.

These findings could also be tested further by looking at volatility in returns for individual stocks linked to targeted publications by media type, and see whether the results also hold in a security specific context. Another interesting research area would be to extend these tests to other markets, such as currency, commodities and bonds markets, which are likely to be impacted by different types of news, and potentially by different media types as well.

2 Essay 2: Day of the Week Patterns in International Stock Market Indices in Light of the Financial Crisis.

2.1 Introduction

Since Fama's (1965) seminal article on market efficiency (EMH), market anomalies have been an area of considerable research interest. One of these anomalies which has generated a great deal of both empirical and theoretical examination is the Day of the Week Effect (French, 1980). According to French (1980), prices should rise higher on Mondays than on other days because of the amount of information released between market close on Friday and Monday's opening bell. This represents three days of information backlog, compared to the normal one day period between trading days. Consequently, expected returns for Monday should in theory" be three times the expected return for other days of the week" (French, 1980). However, empirical findings have documented just the opposite (Agrawal & Tandon, 1994; Mills & Coutts, 1995), with returns lower on Monday and abnormally high on Friday. These contributions (further explored in the literature review section) all refer to an anomalous market activity called "Day of the Week Effect": a recurring daily event, where securities tend to behave differently from the other days of the week. This anomalous activity has been referred to as the most puzzling phenomenon in finance (Jaffe & Westerfield, 1985; Keim & Stambaugh 1984).

Many empirical contributions have attempted to identify a Day of the Week effect across various markets and timeframes. These have differed with regard to choice of modelling methods (linear regression or conditional volatility, e.g. GARCH) and types of dataset (equal-, value-, or price-weighted indices). The literature review presented in this essay shows that the choice of methods and dataset is the crucial determinant of whether a Day of the Week effect is identified. Equal- and (and in some cases) price-weighted indices tend to exhibit a Day of the Week effect in comparison to value weighted indices. Similarly, the use of more advanced conditional volatility modelling such as GARCH tend to reveal the existence of a Day of the Week Effect in index volatility rather than on index returns (as proposed by earlier contributions using linear models).

This essay contributes to this literature in several ways. First, I aim to examine whether the Day of the Week Effect exists in four indices across two countries (FTSE, NASDAQ, S&P 500, and Russell 2000). I follow the literature in using a GARCH model, as established as an adequate modelling tool for Day of the Week Effects on indices (Kiymaz and Berument, 2003; Choudhry, 2010; Al-Jafari, 2012). Further, as the dataset used in this essay encompasses the most recent financial crisis, it is possible to explore whether any Day of the Week Effect identified is found to change during the crisis period. Finally, building on the previous essay (Essay 1) in this thesis, the Day of the Week Effect is further explored through explicitly testing for the role of public information release. In particular, using interaction variables between the daily flow of public information and Day of the Week dummies, it is possible to investigate the direct link between information release, information backlog, and any Day of the Week Effect. This can be seen as a complementary test for the information release hypothesis proposed in Essay 1.

This essay is organized in six sections as follows: Section 2.2 provides a selected review of the literature on the Day of the Week Effect, including a discussion on the methods and datasets used in past contributions. Section 2.3 presents the datasets used in this analysis and some general descriptive statistics relating to the Day of the Week Effect. The methods of analysis and hypotheses tested are discussed in Section 2.4.

The empirical results are presented in Section 5.5. Finally, Section 2.6 provides concluding comments.

2.2 Literature Review

2.2.1 The Day of the Week Effect

The term "market anomaly" is used widely throughout the literature to refer to a variety of events, but it is commonly associated with a period of time during which security market prices are hard to justify based on classical theories of finance. Because their existence appears to conflict with classical theories, anomalous activity on the stock market is an area of much research in the finance literature. Researchers attempt to empirically illustrate the existence of anomalies using various methods, detailing how price deviates from classical expectations. Theoretical researchers also study anomalies in an attempt to explain price deviations by explaining why anomalies appear. A market anomaly can be defined in the following way:

> "Financial market anomalies are cross-sectional and time series patterns in security returns that are not predicted by a central paradigm or theory. This sense of the term 'anomaly' can be traced to Kuhn (1970). Documentation of anomalies often presages transitional phase towards a new paradigm."
>
> Keim, (2008)

This definition is used to frame the concept of Day of the Week Effect explored in this essay.

The Day of the Week Effect refers to an empirically documented market anomaly, in which the behaviour of index or stock returns on one specific day varies from the other days of the week. The origins of the Day of the Week Effect in the literature are linked to empirical and theoretical observations of stock market behaviour on one specific day of the week: originally the Monday Effect, or Weekend effect. First documented by Clark (1973), and further tested in a seminal paper by French (1980), the hypothesised existence of singular stock market returns behaviour on Mondays originates in the trading process itself. As proposed by French (1980), "since most stocks are traded only from Monday through Friday, if returns are generated continuously in calendar time, the distribution of returns for Monday will be different from the distribution of returns for other days of the week" (i.e. Monday returns should be three times higher than other days), whereas if "returns are generated in trading time, the distribution of returns will be the same for all five days of the week" (Monday returns should not be significantly different from other days). In order to test for these hypotheses, French (1980) examines mean returns for each day of the week using the Standard &Poor's 500 from 1953 through 1977. The results show that neither of the two hypotheses (calendar time or trading time) holds, as the Monday returns are found to be significantly negative, while the average for the four other days was positive. This is a seminal paper, which subsequently generated a great deal of interest in the topic, particularly reviewed in the 1980s and 1990s as new statistical methods evolved.

Subsequent research confirmed that the distribution of securities returns varies with the day of the week (Keim & Stambaugh, 1984; Rogalski, 1984; Aggarwal & Rivoli, 1989, Kiymaz & Berument, 2003; Hui, 2005). These studies find the average return for Monday to be significantly lower than the average return for the other days of the week. Interestingly, these findings are not restricted to the US equity markets, and evidence has documented the effect in international markets as well (Jaffe & Westerfield, 1985; Solnik & Bousquet, 1990; Barone, 1990; Kiymaz & Berument, 2003). The Day of the Week Effect or Monday Effect has also been documented in different types of markets including: gold markets (Ball, Torus, & Tschoegl, 1982; Ma, 1986), debt markets, (Gibbons & Hess, 1981; Flannery & Protopapadakis, 1988; Singelton & Wingender, 1994), currency markets, (Coats, 1981; McFarland, Pettit, & Sung, 1982; Thatcher & Blenman, 2001), futures markets (Cornell, 1985, Dyl and Maherly, 1986) and Real Estate Investment Trusts, (Redman, Manakyan, & Liano, 1997). The review presented in this paper focused on stock market research, and empirical documenting of the Day of the Week anomaly. The review does not aim to be exhaustive and report all the numerous contributions to the literature on the Monday Effect or Day of the Week Effect. Instead the review focuses on selected contributions, which have influenced the development of a more sophisticated modelling method, and identified the importance of assumptions when empirically documenting the existence of a daily anomaly in stock market pricing.

The next sections of the review attempt to identify the differences in methods and dataset which have led to variations in findings over time. Explanations for Day of the Week Effects are also presented in Section 2.4.

2.2.2 Linear Models (OLS) – Contributions and Findings

Following the work of French (1980), Gibbons and Hess (1981) further examine the existence of the Day of the Week Effect on the S&P 500 (and extend tests to the 30 stocks in the Dow Jones Industrial Average, and Treasury Bills) by using a linear regression model for examining the statistical significance of the anomaly. A linear regression is run, which includes dummy variables for each day of the week[25]. Results support French's (1980) findings and confirm the presence of a Monday Effect, with persistently negative mean stock returns on Mondays (and below average Treasury Bill returns). These results are found to hold after correcting for the potential heteroskedasticity, and for market effects (i.e. testing for mean-adjusted returns, and for value-weighted index). Subsequently, a number of contributions have used similar methodologies, with varying datasets, and have further confirmed the existence of days of the week effects in a large number of countries (Rogalski, 1984; Jaffe & Westerfield, 1985; Condoyanni et al., 1987; Ziemba, 1991; Agrawal & Tandon, 1994). In summary, the literature featuring linear regression methods seem to confirm these findings. Further, it appears that the majority of these studies confirming the existence of a Day of the Week Effect have used return data from equal-weighed indices. This is a point of major importance, which to my knowledge has not previously been explicitly addressed in literature reviews of the Monday Effect.

Indeed, there are difference in results associated with the choice of index construction (equal-, value-, or price-weighted). Value-weighted indices place a higher weight on the largest companies in an index, as it is constructed by placing weights based on market capitalization. In contrast, the construction of the other two methods does not. For example, the S&P 500 is a value-weighted index by way of construction, so that price changes in the largest companies would have the most impact on the value of the index. Thus, if one were to examine the Day of the Week Effect on the S&P 500 using an alternative construction method (than what exists in the market), for example an

[25] Except Wednesdays, which is arbitrarily excluded to avoid the dummy variable trap.

equal-weighted index, one would expect index returns to vary drastically, and thus Day of the Week Effects as well.

This distinction of index construction method is not always explicit in the literature, but is explored in more detail here. A value-weighted index (also called market capitalisation weighted index) is constructed by multiplying the number of shares outstanding for each security by the unit price of the security. It is shown in Eq. (12):

$$Value\ Weighted\ Index_t = \frac{\sum_{i=1}^{N} S_i\, P_i}{D} \tag{12}$$

where N is the number of constituent securities in the index, S is number of shares outstanding for each security i, P is the unit price of each security i, and D is the divisor for the chosen index, which is a number initially determined at the inception of the index, to ensure that it has a convenient value (e.g. 1,000 or 100). The value of the divisor can be adjusted as necessary to reflect changes in the index (unrelated to price changes in its components).

In a price-weighted index, the weight of each security in the index is determined only by its unit price, as shown in Eq. (13). A property of such indices is that a stock split for one of the securities in the index changes the relative weights of all other securities in the index, and thus the divisor must be adjusted accordingly.

$$Price\ Weighted\ Index_t = \frac{\sum_{i=1}^{N} P_i \left(\frac{P_i}{\sum_{i=1}^{N} P_i} \right)}{D} \tag{13}$$

Finally, in an equal-weighted index, the value of each security in the index is weighted equally, so if there are N securities in the index, each security is assigned a weighted of 1/N, as shown in Eq. (14) below.

$$Equal\ Weighted\ \ Index_t = \left(\frac{1}{N}\right)\frac{\sum_{i=1}^{N} P_i}{D} \tag{14}$$

The literature on the Day of the Week Effect can be divided into several distinct groups in terms of which type of indices they use (equal-weighted indices, versus value-weighted, and to a lesser extent price-weighted). This directly impacts the results and this distinction is rarely explicitly addressed in the literature (For example if I found a Day of the Week Effect in the "S&P 500" without specifying if it was equal- or value-weighted the result would be somewhat misleading). Gibbons and Hess (1981), use both value- and equal-weighted indices finding a more pronounced Monday Effect in the equal-weighted indices. This distinction directly impacts the results and in the case of confirm that small size stocks are more likely to exhibit Day of the Week effects, as suggested by previous uses of equal-weighted indices. Table 20 below presents a non-exhaustive list of major contributions to the day of the week empirical literature, with details methods and index types used in each paper.

Table 20 Selected Contributions on Day of the Week Effect - Data and Methods

Paper	Index	Value-, Equal-, or Price-Weighted	Econometric Method	Day of Week Finding
French, 1980	S&P 500, 1953-1977	Value-weighted	Descriptive statistics and linear regression	Significant negative returns for Mondays
Gibbons and Hess, 1981	S&P 500, and DOW 30 and Treasury Bills, 1962-1978	Value- and equal-weighted (CRSP)	Linear regression	Negative mean returns on Mondays, more pronounced in equal-weighted portfolio.
Keim and Stambaugh, 1984	S&P Composite, NYSE data from 1928 to 1982	Does not specify	Linear regression	Negative mean returns on Mondays
Rogalski, 1984	DJIA, and S&P 500, 1974-1984	Value-weighted (implied)	Linear regression	Negative mean returns on Mondays
Jaffe and Westerfield, 1985	UK, US, Japan, Australia, Canada; 1973-1982	Both value- and equal-weighted	Linear regression	Negative mean returns on Mondays
Condoyanni et al., 1987	DJIA, France CAC, Japan, Canada, Singapore, UK,1969-1984	Does not specify	Linear regression	Negative mean returns on Mondays
Jaffe and Westerfield, 1989	US S&P 500 (1930-1962); S&P 500 (1962-1981); Japan (1970-1983); Canada (1976-1983); Australia (1973-1982); UK (1950-1983).	Japan price-weighted; Canada value-weighted; Australia price-weighted; UK equal-weighted US not specified	T-value testing on hypothesis that average Monday return is statistically significant from zero, when following weeks of markets rising	Mean Mondays returns are negative but only true when the market has declined in the previous week.
Connolly, 1989			GARCH	No Monday Effect

Connolly, 1990	S&P 500, DJIA, 1963-1983	Both value- and equal-weighted	GARCH	No Day of the Week Effect, minimal evidence for weekend, suggests previous literature results were driven by poor stats (OLS)
Ziemba, 1991	Japan 1949-1988, equal and value weighted	Both value- and equal-weighted	Linear regression	Some periods support Mondays Effect
Chang, Pinegar and Ravichandran, 1993	24 countries, 1985-1992	Does not specify (except US- value-weighted)	OLS regression correcting for Heteroskedasticity using White's 1980 Standard errors, and Bayesian critical value t-statistics.	Some Mondays and Weekend Effects in approximately half of the indices
Agrawal and Tandon, 1994	18 countries, 1971-1987	All value-weighted, except Japan price-weighted, and UK equal-weighted.	OLS	Cite several Mondays Effects, strongest in UK
Kiymaz and Berument, 2001	S&P 500, 1973-1997	Does not specify	OLS, GARCH	Significant Day of Week Effect with OLS highest price on Wednesday and lowest on Monday, and in lowest return for Monday during 1987 and 1997 crisis periods with GARCH
Kiymaz and Berument, 2003	Canada, DAX, NIKKEI, UK, US, 1988-2002	Does not specify	GARCH	Mixed results by index, but claim extensive support for Monday Effect in mean and variance equation

Gregoriou et al., 2004	UK, 1986-1997	Does not specify	GARCH model for Variance OLS for mean	Monday Effect significant in variance.
Bhattacharya et al. (2010)	The Bombay Stock Exchange, 1991-2000	Value-weighted	OLS,GARCH	Significant day of the week effects for Thursday and Friday in the return equation, and Monday and Thursday in the variance equation.
Choudhry, 2010	South East Asian markets (India, Indonesia, Malaysia, Philippines, South Korea, Taiwan and Thailand) 1990-1995	Does not specify	GARCH model	Significant Day of Week Effects in both/either mean and variance equations, for various days across indices. Not a clear pattern across indices.

2.2.3 GARCH Modelling and the Day of the Week Effect

In reviewing the literature on the Day of the Week and Weekend Effect in Table 20, it can be seen that until the 1990s, all or most contributions have used linear regression models (OLS) to document the existence of the anomaly; and that correspondingly, these studies all tend to find evidence of the Day of the Week Effect. Connolly (1989) examines the robustness of the Day of the Week Effect to model specifications, as well as sample size, time periods, and returns measure (equal- or value-weighted) and finds that "the evidence of a weekend anomaly is clearly dependent" on these assumptions. In particular, a number of OLS assumptions are unsuitable to the stock market data used in the previously cited studies, as returns tend to be auto-correlated, and exhibit leptokurtosis and heteroskedasticity. Connolly (1989) distinguishes itself from previous contributions in the literature, as it first proposes to test for the Day of the Week Effect in a Generalised Autoregressive Conditional Heteroskedasticity model (GARCH), and compares results to those of linear regressions.

As presented in details in the first essay of this thesis, the AutoRegressive Conditional Heteroskedasticity (ARCH) model proposed by Engle (1982), and later generalised by Bollerslev (1986) in the GARCH model, were developed to model the time-varying and clustering of volatility in stock returns (Jondeau et al., 2007). The GARCH (1,1) model can be described as below:

$$r_t = \omega + \mu r_{t-1} + \varepsilon_t \quad (15)$$

$$\varepsilon_t |\, \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 \quad (16)$$

where r is the return, μ is a constant, ε_t is the serially uncorrelated error terms with a mean of zero and a conditional variance of σ_t^2. In Eq. (18), explicitly modelling the conditional variance of returns, the coefficients β and γ represent the persistence (or clustering) of returns volatility, i.e. $\beta + \gamma$ embodies the extent to which volatility is determined by its past level.

Connolly (1989) compares results for three indices: the S&P 500, the equal- and value-weighted CRSP indices (Centre for Research in Security Prices), over 7 three-year sub-periods (from 1963 to 1983). After accounting for potential bias in previous linear regression methods, the results of statistical tests show that evidence of a Day of the Week Effect is actually weak, and disappears after 1975. Further, Connolly (1989) finds strong evidence of ARCH effects in the data. The paper compares the results of constant mean and constant variance models with GARCH (1,1), where the return equation is augmented with day of the week dummies. The results are relatively indecisive, as different models are found to better perform in each time period for each index. However, the paper does illustrate that evidence of the Day of the Week Effect is indeed highly dependent on the choice of modelling method, sample, and type of index data.

In light of Connolly's (1989) proposal to use GARCH modelling, a number of subsequent studies have further investigated the Day of the Week Effect using this more sophisticated category of models. However, instead of (or in addition to) augmenting the return equation with day of the week dummies, the variance equation can also be augmented to examine whether there could be a Day of the Week Effect in returns volatility. Berument and Kiymaz (2001) first test for variations in stock market volatility across days of the week. In addition to an OLS regression of S&P 500 (1973 to 1997) returns (assuming constant variance), a GARCH (1,1) is run, followed by an augmented GARCH (1,1), where both the constant terms of the return and the time-varying variance equations are allowed to change for each day of the week. In other words, the returns equation and conditional variance equation of the GARCH (1,1) are augmented with dummies representing each trading day. The results of the OLS estimation confirm previous literature findings that Monday returns are the lowest, however Engle's (1982) test for Autoregressive Conditional Heteroskedasticity (ARCH effects) rejects the Null hypothesis of constant conditional variance and thus indicates a preference for GARCH modelling. In the augmented GARCH specification with dummy variables, the results show that a Day of the Week Effect is indeed present in S&P 500 returns volatility, i.e. volatility varies significantly across trading days. The lowest returns are observed on Mondays (consistent with previous linear regression findings), while the highest and lowest volatility are respectively found on Friday and Wednesday. The analysis also disaggregates the data in two sub-periods (pre- and post- 1987 crisis), and confirms Day of the Week Effects in both returns and volatility, although the day with highest volatility pre-crisis is Tuesdays instead of Fridays (post-crisis).

Kiymaz and Berument (2003) extend these findings to international markets, using GARCH modelling to explore day of the week patterns in volatility in Canada, Germany, Japan, the US, and the UK from 1988 to 2002. While their results confirm that a Day of the Week Effect is largely present in returns across indices, they also find that the days with highest and lowest volatility (Day of the Week Effect in volatility) varies across countries:

"the highest volatility of returns on Mondays for Germany and Japan, on Fridays for Canada and the US, and on Thursdays for the UK. The lowest volatility of returns occurs on Mondays for Canada and Tuesdays for Germany, Japan, the UK, and the US."

Kiymaz and Berument (2003), pp. 377-378

Gregoriou et al. (2004) also confirm the existence of a Day of the Week Effect in the 1986-1997 FTSE returns and volatility, although they only use a dummy for Mondays in the variance equation. Monday volatility is found to be significant and positive, albeit their results are mitigated by the construction of "transaction-costs-corrected" returns which are adjusted with bid-ask spreads.

A number of subsequent studies have used similar GARCH methodologies to extend the day of the week returns volatility findings to other markets. For example, Bhattacharya et al. (2010) examine Day of the Week Effects, through comparing an OLS regression with a GARCH model augmented with day of the week dummies in both the returns and variance equations in the Bombay Stock Exchange from 1991 to 2000. Differentiating between reporting weeks[26] and non-reporting weeks, the results show the presence of a Day of the Week Effect in returns, which is significantly positive on non-reporting Thursdays and Fridays, whereas conditional variance is significantly affected by both reporting and non- reporting Mondays and reporting Thursdays[27]. Choudhry (2010) also applies the augmented GARCH (1,1) model to

[26] When the banking sector reports to the Reserve Bank of India every other Friday.

[27] The sample is also divided into two sub-periods (1991-1995 and 1996-200) to reflect changes made to the Indian Index in the mid-1990s and the results show evidence of changes in Day of the Week Effects between the two periods.

examining Day of the Week Effects in emerging Asian markets (India, Indonesia, Malaysia, Philippines, South Korea, Taiwan, and Thailand) from 1990 to 1995, and finds that a Day of the Week Effect is present in both returns and variance equations but differs between the seven indices.

Overall, research has shown that when accounting for time-varying conditional variance, GARCH should be a preferred methodology compared to OLS which assumes constant variance, evidence largely confirms the existence of Day of the Week Effects in the majority of indices. However, this Day of the Week Effect is not only found in index returns, but also in index volatility when it is explicitly modelled in GARCH. The international and time period comparisons have shown that Day of the Week Effects are not restricted to Mondays. Instead Day of the Week Effects vary with each index, and sub-period analysed, suggesting that the underlying factors might differ. A number of potential explanations for the Day of the Week Effect have been proposed in the literature. The major proposals are summarised in the next section.

2.2.4 Potential Explanations for Day of the Week Effect

As previously highlighted, many early contributions in the 1980s found strong evidence of a Monday Effect (lower returns on Mondays than any other day of the week). These contributions tend to suggest that the persistence of day of week anomalies is a sign of market inefficiencies (French & Roll, 1980, Gibbons & Hess, 1981). Indeed, if such a pattern is found to exist in the market, then arbitrage theories suggest it could not persist over time.

In line with inefficient market theories, Keim and Stambaugh (1984) propose a "high Friday returns" hypothesis, which suggests that the price of the last trades of the week tend to be high, so that when market reopens this artificially high price is corrected. Through their analysis of NYSE data from 1928 to 1982, they are able to compare two periods: 1928-1952 (during which the NYSE was open on Saturdays) and 1953-1982 (after the market was closed on Saturdays). They find that during the first 28 year period, Saturday returns tended to be higher. Upon the close of the Saturday market, Fridays become the last trading day of the week and thus prices tended to be higher. Keim and Stambaugh (1984) suggest that this upward price bias on Friday is corrected on Mondays resulting in the Monday Effect.

Alternative explanations in the late 1980s tend to be linked to individual trader's decision making process (Miller, 1988, Dyl & Holland, 1990; Lakonishok & Maberly, 1990). Miller (1988) suggests two factors which could impact individual investor trading patterns during the week. The first factor is the investor's state of mind: Miller (1998) argued that on the weekend, the investor has time to think about his/her portfolio, without distractions. Consequently, this creates a higher amount of trades being placed on Monday. The second factor addresses information individuals receive from brokerage houses[28.] Miller (1988) proposes that the information investors receive during the week is biased towards buy recommendations[29], while on the weekend small investors are less likely to receive recommendations from brokerage houses. The consequence of this cyclic change in the source of information is that individuals place a higher percentage of sell orders on Mondays than any other day of the week, creating downward pressure on security prices on Mondays.

---

[28] This information could include analyst reports and brokerage "buy" lists.
[29] For further evidence of this phenomenon see: Groth et al. (1979), Diefenbach, (1972), and Dimson and Fraletti, (1986).

A further theory based on investor psychology, proposed by Rystrom and Benson (1989), is known as the Blue Monday Syndrome. The argument is that investors sometimes act irrationally and that this inherent irrationality can be affected by mood. From evidence in the psychology literature (Christie & Venables, 1973; Stone et al., 1985), they find a large portion of investors are inherently pessimistic on Mondays relative to other days of the week. Their argument is largely based on psychological findings that people are the most depressed on Monday mornings and the most optimistic on Friday evenings (Christie & Venables, 1973; Stone et al., 1985). This pessimism leads investors to place an unusually high percentage of sell orders, creating downward pressure on prices resulting in the Monday Effect.

However, the more recent empirical contributions to the evidence on the Day of the Week Effect are actually more in line with efficient market theories, as they relate the existence of daily price movements to the arrival of the information. For example, Patell and Wolfson (1982) and Penman (1987) suggest the timing of corporate announcements as a cause for the lower average returns found on Mondays. They find companies release good news during the week and bad news after the market close on Friday. They argue that corporations are strategic with their information releases, aiming for the negative impact of bad news on share prices to be mitigated over the weekend as investors are unable to immediately react and are forced to take additional time in processing data. They find that due to a larger percentage of bad news being released after the market close on Friday, a higher percentage of sell orders are produced on Monday.

Recent evidence across markets and time periods has more recently shown that the Day of the Week Effect is not restricted to Mondays or stock returns, but instead can manifest itself through stock market volatility, and on different days of the week. Accordingly, the link to information release has become a more prominent argument in the literature, as more empirical studies of volatility emerged. The public information release hypothesis, first proposed by French and Roll (1986) suggests that the flow of public information is linked to volatility. Since their seminal paper, a broad literature has established the link between the flow of public information and stock market volatility[30], especially following the development of GARCH modelling. Accordingly, a Day of the Week Effect in volatility could presumably be linked to a different rate of information arrival. Harvey and Huang (1991) suggest that more information is released on Fridays, such as important macroeconomic news. This could for example help explain the results of Kiymaz and Berument (2003) which report high volatilitys on Fridays in the US, or Gregoriou et al. (2004) in the UK. However, the link between a Day of the Week Effect in stock market volatility and the release of information has, to my knowledge, not been explicitly tested in past contributions. Thus a contribution to the literature would be to empirically test the information release hypothesis as an explanation for the Day of the Week Effects. This is one of the contributions of this essay.

2.2.5 Questions Addressed in This Essay

This essay aims to contribute to the literature in several ways. First, the essay aims to establish whether a Day of the Week Effect is detected in both US and UK stock markets in recent years. This essay uses daily data from 2005 through 2010, as the most recent analysis of daily effect on UK and US stock markets. In addition to the FTSE and S&P 500 which are the most commonly used indices in the literature, the

[30] This literature is reviewed in Essay 1 of this thesis, and will not be covered in this section.

essay also investigates the existence of a Day of the Week Effect in two other indices: the NASDAQ, a composite of the common securities exchanged on the NASDAQ, one of the world's largest exchanges by market capitalization, and the RUSSELL 2000, as the most common index of small-capitalization securities. Using recent advances in methods to represent time-varying volatility, I attempt to explore whether a Day of the Week Effect is still present in these indices. Through the explicit modelling of time-varying conditional variance, I build on previous literature to explore whether a Day of the Week Effect could be present in either stock returns and/or volatility. Further, given the dataset covers a period of the most recent financial crisis, the analysis also explores whether a Day of the Week Effect (if it exists) could change before or after the financial crisis. Finally, the analysis will attempt to explicitly test for the role of the arrival of public information in explaining potential Day of the Week Effects, by using a proxy for the daily flow of public information (described in detail in Essay 1).

2.3 Data

The essay uses daily index returns in the period from January 2005 through December 2010. Four indices are considered: the UK FTSE and the US S&P 500, NASDAQ, and RUSSELL 2000. The data was obtained through Datastream. It should also be noted that all the indices used in the essay are market-weighted indices, as opposed to equal-weighted indices, in order to consider stocks according to their importance in index construction. Daily returns are generated from closing prices, as the natural logarithm of first difference.

As shown by Schwert (2011), stock market return volatility has been found to be historically high during the last financial crisis (2007-2008). Because this period is included in the sample data used in this essay, the data is also presented into two sub-periods: a pre-crisis period from January 1, 2005 through August 8, 2007 and a second period representing the crisis and period immediately following the crisis, from August 9, 2007 through December 31, 2010[31]. Using this disaggregation, the 1,486 observations are divided between 643 pre-crisis and 843 post-crisis observations.

2.3.1 General Descriptive Statistics

Table 21 presents some general descriptive statistics of index returns. In addition to the mean, standard deviation, kurtosis, variance, and skewness of the full sample, the descriptive statistics are also shown pre- and post-crisis separately. While the mean return is positive (and between 0.01 and 0.02%) for the full sample across the 4 indices, it is higher and positive during the period leading to the crisis, and lower (and negative for S&P 500 and FTSE) post August 9, 2007. Similarly, the standard deviation and variance in returns is lower pre-crisis than post-crisis. There is also a general pattern of slightly less kurtosis and skewness to the right in pre-crisis period. In terms of index comparison, the S&P 500 appears to have the most leptokurtic distribution during the crisis, and the RUSSELL 2000 has the lowest.

[31] The latest financial crisis is said to have begun August 9, 2007, when BNP Paribas froze three investment funds, first bank to acknowledge its exposure risk to the sub-prime mortgage markets (Federal Reserve Bank Of St. Louis, 2012).

Table 21 Descriptive Statistics Index Returns

Returns	Mean	S.D.	Kurtosis	Variance	Skewness	Observations
Pre-Crisis						
Russell 2000	0.0003	0.0106	0.6107	0.0001	-0.1391	643
S&P 500	0.0003	0.0070	2.0448	0.0000	-0.4248	643
NASDAQ	0.0005	0.0082	3.1277	0.0001	-0.3846	643
FTSE	0.0003	0.0073	1.9256	0.0001	-0.4328	643
Crisis						
Russell 2000	0.0001	0.0226	2.7253	0.0005	-0.1092	843
S&P 500	-0.0001	0.0185	6.1612	0.0003	0.0564	843
NASDAQ	0.0000	0.0167	5.8394	0.0003	0.1247	843
FTSE	-0.0001	0.0169	5.3674	0.0003	-0.0101	843
Total						
Russell 2000	0.0002	0.0184	4.5647	0.0003	-0.1369	1486
S&P 500	0.0001	0.0004	10.2775	0.0002	0.0177	1486
NASDAQ	0.0002	0.0137	8.4074	0.0002	0.0628	1486
FTSE	0.0001	0.0136	8.4589	0.0002	-0.0682	1486

2.3.2 Day of the Week Effect - Descriptive Statistics

It is also useful, before going into regression analysis, to examine the existence of Day of the Week patterns through descriptive statistics. Table 22 below presents the descriptive statistics for index returns, again for both the full sample and the crisis-disaggregation for each day of the week.

First, looking at the overall sample, there is no striking day of the week pattern in returns common across the four indices, other than in standard deviation: it is found to be the highest for Monday returns compared to all other days. However, mean returns are found to be lowest on Mondays for the Russell 2000, on Fridays for the S&P 500 and NASDAQ, and on Wednesdays for the FTSE. In contrast, the highest mean return is on Wednesdays for the Russell 2000, Tuesdays for the S&P 500, Mondays for the FTSE, and Mondays/Thursdays for the NASDAQ. This seems to suggest that if a Day of the Week pattern is indeed found in stock returns, this effect will fall on a different day of the week depending on the index considered, questioning previous findings of a consistent Monday effect in returns).

Some general observations can also be made when examining the pre-crisis and post-crisis data. During the pre-crisis period, there is again no clear pattern in mean returns across the four indices. Lowest mean returns are found on Tuesdays for the Russell, Wednesdays for the NASDAQ, and Thursdays for the FTSE and S&P 500. However, interestingly, these days also appear to be the ones with highest standard deviations (except for the Russell 2000 where it is on Fridays).

During and after the crisis, Mondays do present interesting patterns: they are on average the day with lowest returns for the Russell 2000 and S&P 500, but highest returns for NASDAQ and FTSE. However, in all cases, Mondays exhibit the highest standard deviation across the four indices. In contrast, Fridays exhibits the lowest standard deviation for the Russell 2000 and S&P 500, and the lowest mean returns for the NASDAQ.

Overall, the post-crisis period returns seem to indicate that Mondays are more likely to display a Day of the Week pattern across indices than other days, if not in mean returns, potentially in return volatility. This is an interesting finding, which relates to most recent contributions to the Day of the Week Effect literature. Indeed, as explicit models of conditional volatility are increasingly used, evidence of the Day of the Week Effect is less and less focused on actual returns, and more and more on index volatility. We would then expect our results to confirm such findings, particularly in the post-crisis sub-sample. This is explored in more detail in empirical results in Section 2.5.

Table 22 Price Descriptive Statistics Separated by Day of the Week, Pre-Crisis and Crisis Data

Returns	Mondays		Tuesdays		Wednesdays		Thursdays		Fridays	
Pre-Crisis	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Russell 2000	0.0003	0.0102	-0.0005	0.0111	0.0005	0.0119	-0.0004	0.0170	0.0004	0.0175
S&P 500	0.0009	0.0060	-0.0004	0.0076	-0.0005	0.0067	-0.0006	0.0076	-0.0003	0.0066
NASDAQ	0.0004	0.0072	0.0018	0.0085	-0.0002	0.0078	0.0005	0.0089	0.0005	0.0084
FTSE	0.0004	0.0063	-0.0006	0.0077	-0.0001	0.0079	-0.0010	0.0081	0.0014	0.0063
Crisis	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Russell 2000	-0.0017	0.0249	0.0011	0.0228	0.0006	0.0217	-0.0001	0.0244	0.0006	0.0191
S&P 500	-0.0009	0.0220	0.0012	0.0198	-0.0004	0.0173	-0.0003	0.0187	-0.0003	0.0144
NASDAQ	0.0020	0.0196	0.0009	0.0138	-0.0004	0.0165	0.0006	0.0159	-0.0028	0.0173
FTSE	0.0016	0.0209	0.0004	0.0153	-0.0004	0.0165	-0.0012	0.0146	-0.0006	0.0170
Total	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Russell 2000	-0.0008	0.0201	0.0003	0.0187	0.0012	0.0177	0.0001	0.0197	0.0002	0.0157
S&P 500	-0.0001	0.0172	0.0005	0.0158	0.0004	0.0137	-0.0001	0.0148	-0.0003	0.0117
NASDAQ	0.0013	0.0156	0.0013	0.0118	-0.0003	0.0134	0.0004	0.0133	-0.0013	0.0142
FTSE	0.0011	0.0164	0.0000	0.0126	-0.0004	0.0134	-0.0003	0.0122	0.0003	0.0134

2.3.3 Information Data

The second category of data used in the analysis is a proxy for the flow of public information, gathered using ProQuest as described in Essay 1 of this thesis. This dataset is a proxy for the flow of public information, gathered specifically from business related sources, using filters to create daily information flows. The dataset is composed of the recorded number of publications, collected daily from January 1, 2005 through December 31, 2010. This six year timeframe represents a total of 2,192 daily observations. These observations correspond to 1,486 trading days Table 23 below presents the definitions of the information flow variables used in this essay for convenience.

Table 23 Information Variables and Descriptions

New-All	Number of publications released on the current trading day.
Backlog-All	Number of publications released while the markets were closed.
Backlog	Dummy variable which equals 1 on a trading day with an information backlog and equals 0 on days without.
Total-All	Total number of publications on a given trading day (covering all media types) including, when it exists (i.e. Backlog Dummy = 1) the backlog of publications that were issued over the weekends/holidays. Total-All= New-All + Backlog-All

In Essay 1, this dataset was used to document the general relation between volatility and the flow of public information. The analysis also showed that the existence of a backlog of information during non-trading days (such as weekends or holidays) has a major impact on stock market volatility. It is used in this essay as well, to proxy for the flow of public information to explore whether a Day of the Week Effect can be explained through patters of information arrivals explicitly.

The disaggregation by media type is not used in this essay, but could constitute an interesting avenue for future research on the Day of the Week Effect. The fact that this flow of information proxy is wide-ranging and generally a major advantage of the dataset in Essay 1: given the focus on documenting a general relation between information flow and volatility, it is better to use a proxy for the flow of information which is as general as possible.

In the case of this essay, this could potentially represent a drawback: if it is shown that different day of the Week Effects are found on different indices, the argument could be that different flows of information are the explanatory cause behind it. Thus, when comparing Day of the Week Effect across indices and countries, the proxy of information flow is not disaggregated to establish whether the information is likely to impact one index over another. There are major difficulties in attempting such a disaggregation. For instance, one could think of disaggregating the information flow data to account for its country of origin (i.e. where the information was published), but this could be argued to be misleading given the increasingly global context of information release. Another solution would be to identify which index it is likely to impact by analysing the information content, however this requires highly sophisticated methods of content analysis, and poses the risk of excluding a crucial part of information which is not directly linked to specific stocks. Given these difficulties, the ProQuest proxy is deemed acceptable in this first attempt at examining daily patterns of information flows and their link to returns and volatility. Table 24 presents some descriptive statistics for each day of the week, for the three variables used here, namely: *Total-All*, *Backlog-All,* and *New-All.*

Table 24: Descriptive Statistics - Information Flow (000s)[32]

Information Flow	Mean	S.D.	Kurtosis	Variance	Skewness	Obs.
Pre-Crisis						
Total-All	11.9482	8.1094	9.0660	65.7618	2.6556	643
New-All	10.2223	6.1112	8.7951	37.3465	3.0996	643
Backlog-All	1.7375	5.0318	51.9450	25.3193	6.1210	643
Crisis						
Total-All	15.9941	9.8385	7.6279	96.7967	2.4772	843
New-All	13.1119	5.7210	7.2266	32.7304	2.5994	843
Backlog-All	2.8821	7.1119	16.2763	50.5789	3.5877	843
Total						
Total-All	14.2450	9.3471	8.0658	87.3682	2.4958	1486
New-All	11.8621	6.0645	6.6828	36.7784	2.5587	1486
Backlog-All	2.3880	6.3214	24.3641	39.9598	4.3118	1486

[32] Note that the information flow is scaled by 1,000 to enable a clearer interpretation of modelling results.

Some general observations can be drawn out. Firstly, the average total number of publications is found to be higher during the crisis than before, both when including or excluding the backlog of information (for example, almost 16,000 for *Total-All* post-crisis compared to less than 12,000 pre-crisis) . Similarly, the backlog of information is also larger during the crisis (approximately 2,900 post-crisis compared to 1,700 pre-crisis. Thirdly, the standard deviation and variance are also found to be larger for all variables during the crisis compared to the pre-crisis sample.

Following a similar exercise as for returns, the information flow data is now presented for each of the five days of the week, in Table 25. The mean and standard deviation are shown for each day of the week for each of the information variables, and separates the data into pre- and crisis periods as well. As expected, the total information flow is largest on Mondays, compared to any other day of the week, reflecting the fact that *Total-All* for Mondays incorporates the information backlog of the weekend. However, it also appears that *New-All* is largest on Mondays as well, so that there is more information published on that day.

Across the five days of the week, the mean number of publications is still larger during the crisis, however, the standard deviation is found to be higher during the end of the week (Thursdays and Fridays) before the crisis, and higher during the beginning of the week in the crisis sample (Mondays and Tuesdays). Interestingly, the lowest standard deviation is found to be on Wednesdays both pre- and post-crisis.

Table 25 Descriptive Statistics for the Information Variable Proxies: Pre-Crisis and Crisis Data

Information flow	Monday		Tuesday		Wednesday		Thursday		Friday	
Pre-Crisis	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total-All	20.4195	9.3881	10.5717	7.4624	10.5262	6.1691	10.2626	6.4242	9.3664	6.1512
New-All	12.1072	5.0957	9.2331	6.2110	10.5456	6.0997	10.2237	6.4245	9.2945	6.1641
Backlog-All	8.3123	8.4107	1.3386	4.3159	0.0363	0.4220	0.0389	0.4503	0.0718	0.5874
Crisis	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total-All	28.8818	11.5331	13.8698	9.5416	13.0384	5.4346	12.8681	5.0709	12.8494	5.5532
New-All	16.0835	7.3364	11.6205	4.7727	12.8832	5.2351	12.8681	5.0709	12.4612	5.1627
Backlog-All	12.7984	7.9997	2.2493	8.3112	0.1551	2.0581	0.0000	0.0000	0.3882	2.9153
Total	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Total-All	25.3103	11.4542	12.4471	8.8451	11.9479	5.8889	11.7343	5.8469	11.3219	6.0653
New-All	14.4053	6.7670	10.5906	5.5581	11.8685	5.7359	11.7172	5.8514	11.0725	5.8305
Backlog-All	10.9050	8.4564	1.8565	6.8837	0.1036	1.5721	0.0171	0.2985	0.2494	2.2215

2.4 Methodology

2.4.1 GARCH Modelling

As well documented in the financial markets literature, as well as the Day of the Week Effect literature, the existence of time-varying volatility in index returns suggests the need to use more sophisticated models of returns than simple OLS. Accordingly, the essay employs GARCH modelling (Bollerslev, 1986) to explicitly allow conditional variance to be dependent on past returns and volatility[33]. The first essay in this thesis has confirmed the good fit of the GARCH (1,1) model for the FTSE and the S&P500. Further tests in the RUSSELL 2000 and NASDAQ confirm the presence of ARCH effects in these indices as well[34]. The GARCH (1,1) model was shown in Eq. (18) and (19) in the literature review. The simple GARCH (1,1) results will be shown alongside the augmented models proposed below.

2.4.2 Day of the Week Effect

In order to test for Day of the Week Effects, I augment the GARCH model to include dummy variables into the mean and variance equations, as proposed by several authors in the literature (e.g. Baker et al. 2008; Choudhry, 2010). With this method, dummy variables

[33] A number of recent contributions have explored the value-added in the use of the exponential GARCH model (EGARCH) developed by Nelson (1991) in the context of Day of the Week Effects (see for example Berument et al., 2007; Al-Jafari, 2012; Monteiro, 2012; Srinivasan & Kalaivani, 2013). However, the evidence of superiority of this model is not clearly established. In this essay, I compare the GARCH and EGARCH specifications through the Shwartz's AIC and BIC tests, and confirm the need to use GARCH. The results are presented in Appendix 1.
[34] The results of the tests are reported in Appendix 2.

are created for each of the five trading days (D_1 for Mondays, D_2 for Tuesdays, D_3 for Wednesdays, D_4 for Thursdays and D_5 for Fridays). For example, for Mondays:

$$D1_t = \begin{cases} 1 & \text{if } t \text{ is a Monday} \\ 0 & \text{Otherwise} \end{cases} \tag{17}$$

These dummy variables are then incorporated as explanatory variables in both the returns and variance equations of the GARCH model. The augmented GARCH model is shown in Eq. (18) and (19).

Model 1: The Augmented Day of the Week Model

$$r_t = \omega + \mu r_{t-1} + \sum_{i=1}^{5} \kappa_i D_{i,t} + \varepsilon_t \tag{18}$$

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \sum_{i=1}^{5} \chi_i D_{i,t} \tag{19}$$

However, in order to avoid the dummy variable trap, one of the dummies must be arbitrarily excluded[35] (Kiymaz & Berument, 2003; Al-Jafari, 2012). The dummy variable for Thursdays is excluded in most regressions, but difficulties in convergence mean that occasionally, another dummy was dropped instead.

[35] Another possibility is to exclude the constant.

In terms of interpreting the results, if the coefficients associated with the dummy variables are found to be significant, this would indicate a Day of the Week Effect in either stock return or volatility for the index under consideration. This will allow for investigating the presence of a Monday Effect in the data. As documented in past contributions (Keim & Stambaugh, 1984; Rogalski, 1984), we would expect that the coefficient of the Monday dummy in the returns equation would be negative and statistically significant. Similarly, other contributions have shown the opposite Day of the Week Effect can be true for Fridays, with generally higher returns on Fridays (Keim & Stambaugh, 1984; Agrawal & Tandon, 1994). In this case, we would expect the coefficient associated with the Friday dummy in the returns equation to be positive and statistically significant. However, more recent contributions have also shown the existence of Day of the Week Effects in stock return volatility with the explicit modelling of conditional variance in GARCH models. This is explored through the coefficients of the dummy variables in the conditional variance equation.

2.4.3 The Flow of Public Information and the Day of the Week

One of the first objectives of this essay is to empirically test for French's (1980) hypothesis of a Monday Effect in index returns. If a Monday Effect is detected through testing of Model 1, a possible explanation in the literature has been that information accumulated over the weekend is incorporated into prices when the market reopens. Essay 1 has already suggested that information is a significant determinant of index return volatility.

We can further extend this finding by exploring whether a Day of the Week Effect can be attributed to the release of information, in other words, if a day of the week is found on an index in either return or volatility, can it be explained by the fact that information flow has changed? This can be tested explicitly through the use of interaction variables. In this section, the returns and variance from Model 1 shown in Eq. (18) and (19) are augmented to include interaction variables between the total information and the day of the week dummies. Model 2 is shown in Eq. (20) and (21) below:

Model 2: Day of the Week Effect and Information Backlog

$$r_t = \omega + \mu r_{t-1} + \kappa_i D_{i,t} + \Phi_i (D_{i,t}.Total\_All_t) + \varepsilon_t \qquad (20)$$

$$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \chi_i D_{i,t} + \Psi_i (D_{i,t}.Total\_All_t) \qquad (21)$$

In Model 2, the impact of the day of the week dummy and its interaction variables have to be explored separately for each day of the week, because of the difficulties to converge that augmented GARCH models face[36].

With this model, the coefficients Φ_i and Ψ_i are the coefficients of the interaction variables between day of the week dummies and information flow. In other words, they represent the different impact of information flow on volatility on that specific day of the week

[36] Optimization was attempted for a model incorporating 4 dummies and 4 interaction variables using a very large number of maximization techniques in Stata without success. Thus the results are presented for each day of the week separately.

compared to the others. With this model, it is possible to establish whether information flow is the determinant of the Day of the Week Effect. With a joint deletion test of κ_i and Φ_i allows us to reject the null at the 5 % level (if the F-test p-value is lower than 0.05), we can reject the null that there is a remaining Day of the Week Effect after controlling for the impact of information flow. A p-value higher than 0.05 would therefore be indicative that information flow seems to be the explanatory variable behind changes in volatility and returns. This finding would be in line with the empirical literature on volatility, and arguments in favour of market efficiency.

2.5 Empirical Results

This section presents the results of the models described in the previous section. For each set of results, the data is presented in the two sub-samples (pre- and during crisis).

2.5.1 GARCH (1,1)

First, the results of the simple GARCH (1,1) model are presented in Table 26 and Table 27 below for the four indices covered in the essay, for the pre- and post-crisis data respectively.

Table 26 GARCH: (1,1) Pre-Crisis

Returns: $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Variance: $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

	Coeff.	FTSE	NASDAQ	S&P 500	Russell 2000
Returns	ω	0.001	0.001	0	0.001
		(0.02)**	(0.00)***	(0.07)*	-0.22
	μ	-0.044	-0.026	-0.061	-0.003
		-0.3	-0.55	-0.18	-0.95
Variance	β	0.125	0.157	0.053	0.048
		(0.00)***	(0.00)***	(0.02)**	(0.03)**
	γ	0.818	0.801	0.893	0.896
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	α	0	0	0	0
		(0.01)**	(0.01)**	-0.11	-0.13
Log-Likelihood		2,304.23	2,266.31	2,306.11	2,023.25
Prob>chi2		0	0	0	0
Number of obs		643	643	643	643

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 27 GARCH (1,1) Crisis Data

Returns: $r_t = \omega + \mu r_{t-1} + \varepsilon_t$

Variance: $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$

	Coeff.	FTSE	NASDAQ	S&P 500	Russell 2000
Returns	ω	0.001	0.001	0.001	0.001
		-0.1	(0.08)*	(0.08)*	(0.02)**
	μ	-0.067	-0.03	-0.115	-0.11
		(0.08)*	-0.44	(0.01)***	(0.01)***
Variance	β	0.107	0.105	0.102	0.097
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.879	0.884	0.888	0.894
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	α	0	0	0	0
		(0.01)**	(0.02)**	(0.01)**	(0.07)*
Log-Likelihood		2,398.22	2,404.10	2,378.80	2,162.37
Prob>chi2		0	0	0	0
Number of obs		843	843	843	843

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

The results of the GARCH (1,1) model confirm the results of the test for ARCH effects. The β and γ coefficients are significant at least at the 5% level for all indices, and their sum is very close to 1, confirming the presence of conditional time-varying heteroskedasticity in index returns. This is the case in both pre-crisis and crisis periods. By looking at the sum of the ARCH and GARCH terms of the volatility equation ($\beta + \gamma$), it is also possible to examine the extent of volatility clustering in the data, as shown in Table 28 below. This shows that volatility clustering was more pronounced during the crisis, than during the pre-crisis period.

Table 28 Volatility Clustering Pre- and Crisis Data

$\beta + \gamma$	FTSE	NASDAQ	S&P 500	Russell 2000
2005-2007	0.943	0.958	0.946	0.944
2007-2010	0.986	0.989	0.990	0.991

2.5.2 Day of the Week Effect

The results of Table 29 and Table 30 below show the augmented GARCH (1,1) model with the day of the week dummies in both the mean and variance equation, in the pre- and post-crisis periods respectively.

It can be seen that across all four indices, there is no clear Day of the Week Effect in the returns equation, either pre- or post-crisis. Only the coefficients of the Tuesdays and Fridays dummy are significant at the 10% level for the Russell 2000 in the pre-crisis sample, and only the Friday dummy (at 10% again) for the NASDAQ after the crisis. These results suggest that the Day of the Week Effect in returns disappears when using a more sophisticated modelling method such as GARCH, as first argued by Connolly (1999).

Table 29 Augmented GARCH with Day of Week Dummies Pre-Crisis

Returns: $r_t = \omega + \mu r_{t-1} + \kappa_1 D_{1,t} + \kappa_2 D_{2,t} + \kappa_3 D_{3,t} + \kappa_5 D_{5,t} + \varepsilon_t$

Variance: $\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2 + \chi_1 D_{1,t} + \chi_2 D_{2,t} + \chi_3 D_{3,t} + \chi_5 D_{5,t}$

	Coeff.	FTSE	NASDAQ	S&P 500	Russell 2000
Returns	κ1	0	0	0	-0.002
		-0.77	-0.93	-0.69	-0.14
	κ2	-0.001	0.001	0	-0.002
		-0.15	-0.25	-0.79	(0.08)*
	κ3	-0.001	-0.001	0.001	
		-0.23	-0.51	-0.14	
	κ4				-0.002
					-0.19
	κ5	0.001	0	0	-0.002
		-0.45	-0.88	-0.62	(0.09)*
	ω	0.001	0	0	0.002
		(0.08)*	-0.24	-0.61	(0.02)**
	μ	-0.046	-0.024	-0.052	-0.002
		-0.28	-0.57	-0.21	-0.96
Variance	β	0.148	0.17	0.075	0.053
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.757	0.771	0.815	0.885
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	χ1	-9.883	0.58	-2.076	-6.792
		-0.92	-0.78	-0.99	-0.76
	χ2		0.843	10.214	
			-0.67	-0.9	
	χ3	-0.093	-9.68		-5.683
		-0.9	-0.93		-0.83
	χ4	-9.402		0.178	-6.072
		-0.9		-1	-0.97
	χ5	-8.25	-0.281	-1.304	-8.269
		-0.92	-0.93	-	-0.9
	α	-11.33	-12.611	-20.82	-10.325
		(0.00)***	(0.00)***	-0.81	(0.00)***
Log-Likelihood		2,311.15	2,269.37	2,317.61	2,028.75
Prob>chi2		0	0	0	0

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 30 Augmented GARCH with Day of Week Dummies During the Crisis

Returns Equation: $r_t = \omega + \mu r_{t-1} + \kappa_1 D_{1,t} + \kappa_2 D_{2,t} + \kappa_3 D_{3,t} + \kappa_5 D_{5,t} + \varepsilon_t$

Conditional Variance: $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \chi_1 D_{1,t} + \chi_2 D_{2,t} + \chi_3 D_{3,t} + \chi_5 D_{5,t}$

	Coeff.	FTSE	NASDAQ	S&P 500	Russell 2000
Returns	κ1	0.002	0.001	0.001	0.001
		-0.17	-0.56	-0.37	-0.44
	κ2	0	-0.001	0	0.001
		-0.83	-0.47	-0.84	-0.48
	κ3	0.002	-0.001	0.001	0.002
		-0.25	-0.42	-0.46	-0.28
	κ4				
	κ5	0	-0.003	-0.001	0
		-0.79	(0.04)**	-0.65	-0.89
	ω	0	0.002	0	0
		-0.91	(0.07)*	-0.73	-0.87
	μ	-0.059	-0.024	-0.113	-0.109
		-0.11	-0.51	(0.00)***	(0.01)***
Variance	β	0.108	0.111	0.101	0.092
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.875	0.876	0.889	0.899
		(0.00)***	(0.00)***	(0.00)***	(0.00)***
	χ1	8.029	-1.213	-8.006	-0.688
		-0.81	-0.99	-0.87	-0.95
	χ2	6.302	-0.281	-2.048	0.178
		-0.85	-1	-0.8	-0.98
	χ3		-0.18	-7.602	-0.258
			-1	-0.84	-0.98
	χ4	1.154			
	χ5	1.053	8.017	-6.027	-124.982
		-0.98	-	-0.83	(0.00)***
	α	-18.856	-18.846	-11.31	-12.038
		-0.57	(0.00)***	(0.00)***	(0.01)***
Log-Likelihood		2,400.69	2,408.70	2,380.95	2,163.33
Prob>chi2		0	0	0	0

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

In the variance equation, there are stark differences in the two sub-samples. Mondays appear to have the lowest volatility in the FTSE, S&P 500 and Russell 2000 before the crisis, while it is Wednesdays for the NASDAQ and the S&P 500. After August 9, 2007, volatility for the FTSE is highest on Mondays, while it is at its lowest on the same day for the NASDAQ and S&P 500. Volatility is at its lowest on Fridays for both the FTSE and the Russell 2000.

A more surprising finding is the lack of significance of the dummy coefficients in the volatility equation. Only the Friday coefficient of the Russell 2000 during the crisis is significant at the 1% level. Previous contributions have illustrated the presence of Day of the Week Effects in some of the indices used in this essay. For example Berument and Kiymaz (2001) find evidence of Wednesday and Friday effects in S&P return volatility between 1973 and 1997, both pre- and post- 1987 crisis. Similarly, Gregoriou et al. (2004) also find evidence of a Monday Effect in the FTSE from 1986 through 1997. The results of this essay suggest that there does not appear to be strong evidence of a Day of the Week Effect remaining in the indices considered during the period 2005 through 2010, either in returns or their volatility.

Overall, the results suggest very few significant coefficients. However, Fridays appear to be the most significant Day of the Week Effect overall. Both the Russell 2000 and NASDAQ have a significant coefficient for the Friday dummy in the returns equation before the crisis, and the Russell 2000 also exhibits a significant Friday dummy in the volatility equation after the crisis. Interestingly, the Russell index is composed of 2000 stocks amongst the smallest market capitalisation stocks in the U.S. The presence of a Day of the Week Effect in this index suggests that smaller stocks might be more subject to it, in line with the previous observation that studies using equal-weighted indices tend to find more evidence of Day of the Week Effects.

Given these results, there might be a limited scope to add to the explanation with the information variable. This is with the exception of exploring whether the Friday Effect in Russell 2000 returns before the crisis, and Russell 2000 volatility after the crisis, or the Friday Effect in the NASDAQ returns pre-crisis, can be fully explained by the information release hypothesis presented in the first essay. Therefore, the results of the model using the interaction variables) are presented in the next section.

2.5.3 Information and Day of the Week Effect

Table 31 and Table 33 present the results (pre- and post-crisis) of the augmented GARCH model to include a day of the week dummy and an interaction variable (day of week dummy * *Total-All*) for the Russell 2000. By including the interaction variable and the day of the week dummy, the results show that the individual day effect patterns in price and volatility are in part determined by the flow of information. Using the joint deletion f-tests (interaction variable and day of the week dummy), I can reject the null hypothesis that there is a remaining Day of the Week Effect after controlling for the impact of information flow as shown in Table 31 below. As Table 33 shows these results are uniform across all indices, during both pre- and post-crisis periods. Consequently, these results support findings in the literature which suggest price and volatility change are reflective of the flow of information (French and Roll, 1986; Clark, 1973). They refute the notion that there is a Day of the Week Effect once you control for the flow of information. The results of the joint deletion tests for all indices are shown in Table 33 below and the remaining results for hypothesis three (S&P 500, FTSE, and NASDAQ) are shown in Appendix 4 (all results confirm the findings shown below).

Table 31: Augmented GARCH with Interaction and Day of Week Dummies Pre-Crisis in the Russell 2000

Returns: $\boldsymbol{r_t = \omega + \mu r_{t-1} + \kappa_i D_{i,t} + \Phi_i(D_{i,t}.Total\_All_t) + \varepsilon_t}$

Variance: $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \chi_i D_{i,t} + \Psi_i(D_{i,t}.Total\_All_t)$

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
Returns	Φ	0.000	0.000	0.000	0.000	00
		-0.93	-0.8	-0.13	-0.22	-0.55
	κ	-0.001	-0.001	0	-0.002	-0.001
		-0.83	-0.76	-0.88	-0.27	-0.44
	ω	0.001	0.001	0	0.001	0.001
		-0.2	-0.14	-0.79	-0.21	-0.16
	μ	-0.004	-0.001	0	-0.007	-0.008
		-0.93	-0.98	-1	-0.87	-0.85
ARCH	β	0.049	0.053	0.048	0.051	0.043
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***
	γ	0.896	0.886	0.899	0.89	0.901
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HET	Ψ	-0.007	-0.016	-0.433	0.009	0.278
		-0.96	-0.5	-0.99	-1	-0.52
	χ	-0.587	13.255	-6.03	-5.766	-7.902
		-0.9	-0.88	-0.96	-0.94	-0.59
	α	-11.889	-23.391	-11.809	-11.685	-11.817
		(0.00)***	-0.79	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,023.33	2,027.33	2,026.50	2,024.18	2,024.69
Prob>chi2		0	0	0	0	0
Number of obs		643	643	643	643	643

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 32 Augmented GARCH with Interaction and Day of Week Dummies During the Crisis

Returns: $\boldsymbol{r_t = \omega + \mu r_{t-1} + \kappa_i D_{i,t} + \Phi_i(D_{i,t}.Total\_All_t) + \varepsilon_t}$

Variance: $\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2 + \chi_i D_{i,t} + \Psi_i(D_{i,t}.Total\_All_t)$

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
Russell return	Φ	0.000	0.000	0.000	0.000	0.000
		(0.06)*	-0.67	-0.42	-0.29	-0.57
	κ	-0.005	0	-0.001	0.002	0.001
		-0.12	-0.87	-0.72	-0.55	-0.85
	ω	0.001	0.001	0.001	0.001	0.001
		(0.04)**	(0.06)*	-0.1	(0.01)**	(0.01)**
ARMA	μ	-0.107	-0.11	-0.108	-0.108	-0.109
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
ARCH	β	0.101	0.083	0.097	0.095	0.097
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	γ	0.889	0.906	0.894	0.895	0.894
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HET	Ψ	-0.326	0.525	0.364	-0.245	-0.407
		-0.33	-0.48	-0.77	-0.65	-0.53
	χ	7.327	-31.391	-9.661	4.614	4.945
		-0.21	-0.54	-0.81	-0.14	-0.33
	α	-12.25	-12.331	-12.3	-12.995	-12.499
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,164.34	2,164.94	2,163.45	2,163.83	2,163.17
Prob>chi2		0	0	0	0	0
Number of obs		843	843	843	843	843

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 33 Joint Deletion Tests for Day of the Week Dummies and Interaction Variable

Index	Day of Week F-test				
FTSE	Monday	Tuesday	Wednesday	Thursday	Friday
CHI	12426.180	6.650	4.120	1.370	10.19
PROB	0.000	0.156	0.390	0.712	0.0374
NASDAQ					
CHI	4.710	5.240	3.280	0.390	1.500
PROB	0.319	0.263	0.350	0.983	0.826
S&P 500					
CHI	0.240	0.850	5.450	0.790	0.720
PROB	0.972	0.932	0.244	0.939	0.948
Russell 2000					
CHI	0.130	1.150	6.280	1.530	2.700
PROB	0.998	0.886	0.179	0.821	0.610

The final section below provides concluding remarks and suggestions for future research.

2.6 Conclusion

This essay examines the existence and potential explanations through information patterns of Day of the Week Effect in US and UK stock market indices before, during and after the latest financial crisis. A review of the literature on Day of the Week Effect suggests that early contributions found evidence of recurring daily patterns in stock price movements (i.e. market anomaly). In particular, index returns were found to be consistently lowest on Mondays (French and Roll, 1987). However, it is worthwhile to note these contributions were mainly based on linear regression of stock returns, and often based their findings on equal-weighted indices, which place an equal emphasis on all component stocks and ignore their relative importance through market-capitalisation.

More recent evidence using sophisticated models of time-varying conditional volatility (e.g. GARCH) have shown more mixed results. Following the work of Connolly (1989), GARCH models have been increasingly used to detect daily anomalies in returns or volatility. While these contributions do not confirm the overwhelming evidence of Monday effects in returns, a number of papers do show the existence of other day of the week patterns, especially in volatility, varying across indices and countries. This essay built on this literature by examining whether the previously identified day of the week patterns in UK and US indices are still present using more recent financial data, whether they might have changed since the latest crisis, and whether they can be explained by the flow of information, building on the first essay of this thesis.

The method used for analysis of the Day of the Week Effect is an augmented GARCH(1,1) model. Day of the week dummies are constructed for both the return and volatility equations. Further, an interaction variable is also created to examine whether day of the week effects can be explained by the flow of information, using the *Total-All* variable from essay 1 (*Total-All* *day of week dummy). The dataset is divided into two sub-samples to examine pre-crisis and crisis period.

Weak results are found for the significance of the day of the week dummies across the indices tested. We find no strong evidence of day of the week patterns over the sub-samples in the FTSE or S&P, either in returns or return volatility, which were identified in older data from previous contributions (e.g. Berument and Kiymaz, 2001; Gregoriou et al., 2004). Overall, the results suggest very few significant coefficients. However, Fridays appear to be the most significant Day of the Week Effect overall. Both the Russell and NASDAQ have a significant coefficient for the Friday dummy in the returns equation before the crisis, and the Russell 2000 also exhibits a significant Friday dummy in the volatility equation after the crisis. Interestingly, the Russell index is composed of 2000 stocks amongst the smallest market capitalisation stocks in the U.S. The presence of a Day of the Week Effect in this index suggests that smaller stocks might be more subject to it, in line with the previous observation that studies using equal-weighted indices tend to find more evidence of Day of the Week Effects.

Finally, by introducing an interaction variable between the day of the week dummy and the total flow of public information (as defined in the previous essay), it is possible to examine whether the information flow is a determinant of the day of the week patterns identified. Overall, the joint deletion test of the interaction variable and the dummy suggest that the information flow satisfactorily explains the da of the week effects in the Russell.

Although overall the evidence for a Day of the Week Effect seems to disappear when using GARCH model and controlling for the flow of information, there are a number of potential extensions for future research. For example, recent evidence in favour of Day of the Week Effects is primarily focused on emerging markets such as in India, Thailand and Indonesia (Choudhry, 2010; Bhattacharya et al. 2010). Further research could attempt to examine the difference between emerging and more established indices, and attempt to explain these differences through the flow of information.

3 Essay 3: The Predictive Power of Investor Sentiment in Early-Warning Systems Of Stock Market Crises And Recoveries

3.1 Introduction

There has been a trend in recent academic literature examining the empirical link between investor sentiment and stock returns. A strong link between the two variables has been documented, for example Brown and Cliff (2005) demonstrate (in US markets) that periods of higher optimism tend to be followed by periods of lower returns. This finding suggests an inverse relationship between the two variables. Fisher and Statman (2000) find that individual investors are easily influenced by sentiment. Baker and Wurgler (2006) argue sentiment has varying effects on different types of stocks particularly securities which are hard to value or arbitrage. Schmeling (2009) and Chang et al. (2012) research country specific factors, finding institutional and cultural differences with regard to interpretation and role of sentiment in price setting. This recent trend of empirical analysis is based on a more historic trend with research attempting to theoretically link investor sentiment to stock returns such as Black (1986) and De Long et al. (1990). Since a large body of empirical and theoretical research supports the notion that sentiment does play a significant role in price setting, the role sentiment plays in detecting market crises or market recoveries holds great economic and academic significance, although until recently, remains largely under-explored. Several studies have anecdotally mentioned sentiment as a possible cause in market crises (De Long & Shleifer, 1991; Shiller, 2000). Yet, not a great deal of research has analysed this link[37].

[37] This literature is explored in details in Section 3.2 below.

The goal of this essay is to specifically examine the link between investor sentiment and stock market crises and recoveries. In this objective, it builds on existing research in three major ways.

First, the essay aims to assess sentiment's role in predicting the probability of the occurrence of crises (and recoveries)[38] in a cross-country comparison (six indices in five countries) through the use of an Early Warning System (EWS) model. Using literature-based indicators, stock market crises and early signs of recovery are detected across the six indices. Using logit models, the probabilities of stock market crises and early signs of recovery are estimated using both a set of leading macroeconomic variables (informed by the literature) and investor sentiment. This method allows for identifying the value-added of investor sentiment in predicting the probability of the occurrence of stock market recoveries. Given the difficulties with measuring investor sentiment, various proxies are used and compared in this essay. Three common proxies are examined to establish the effectiveness of sentiment as a crisis/recovery predictor: the Baker and Wurgler (2006) Sentiment Index (BWSI), the put-call ratio (PCR) and the volatility index (VIX).

The results confirm the significance of investor sentiment in predicting the probability of the occurrence of crises, as the model performance in accuracy of prediction is improved by the addition of the sentiment index variable. More specifically, the results find the commonly used orthogonalised Baker and Wurgler (2008) sentiment index to perform best. Secondly, although less significant, the results suggest that investor sentiment can

[38] The methods used predict the probability of the occurrence of a crisis in the following 12 months.

also contribute in predicting early signs of the occurrence of market recoveries following crises.

This essay differs from previous research in a variety of ways. First, it compares stock market data from three distinct regions, through six indices in total from North America, Europe, and Asia. Second, it applies methods of EWS models to stock market crises using the leading macroeconomic indicators are selected from recent crises literature. Third, it explicitly test three different proxies for sentiment to determine the most accurate choice for EWS models. Fourth, it also tests a model predicting the probability of the occurrence of early signs of market recovery.

The remainder of this essay is organized as follows. Section 3.2 provides more detail into the literature on investor sentiment and its impact on stock market returns, both through theoretical and empirical contributions of behavioural finance. Section 3.3 describes the data used in the analysis, and focuses on the identification of crises in the six indices. Section 3.4 details the method for analysis, including a description of basic principles behind early warning signals models of crises, and the econometrics methods used in this paper. Section 3.5 presents the empirical results of the model for predicting the probability of the occurrence of stock market crises, and discusses the significance of the sentiment indices as predictors, while also comparing results across countries and indices. Section 3.6 proposes a new method for detecting early signs of market recoveries based on the same methodology as early warning signals models, and further tests the value-added of sentiment in predicting the probability of the occurrence of these early signs of recovery. Finally, Section3.7 presents the conclusions of the essay.

3.2 Investor Sentiment in the Financial Literature

The process by which security prices change with the onset of new information has been studied extensively in the financial literature (See Essay 1). However, empirical studies of stock market events have led to further research developments in the drivers of stock price changes. Advances in the literature have focused on the role of investor psychology in explaining that price changes occur due to more than just fundamentals. This literature review summarises the major contributions to this line of research, which have provided evidence of the importance of sentiment in stock market prices.

3.2.1 Origins in the Literature

The origins of sentiment and its role in market price setting dates back to Adam Smith (1776), who argued prices can drift from their natural price to its market price due to market mood. Later, in the 20th Century, John Maynard Keynes also suggested that sentiment could play a role in market prices: “the market is subject to waves of optimistic and pessimistic sentiment, which are unreasoning and yet in a sense legitimate where no solid basis exists for sound calculation” (Keynes, 1936, p. 154). This same notion has been the focus of growing attention in more modern research on financial markets (Shiller, 2000; Baker & Wurgler 2007), and has led to a wide range of specific interests. For example, Smidt (1968) links sentiment to speculative bubbles, while Zweig (1973) focuses on biased expectations and Black (1986) discusses noise. Because of the variety in the literature, investor sentiment has been defined numerous times depending on the focus of the study considered.

3.2.2 Defining Investor Sentiment

Formal definitions of sentiment have been offered in a variety of papers over the course of recent years. Baker and Wurgler (2006) can be seen as building on Keynes' earlier statement. Looking at explaining cross-sectional variation in mispricing, they propose that sentiment can either be "the propensity to speculate," driving relative demand for investments, or be simply defined as "optimism or pessimism about stocks in general", and it is relatively weaker arbitrage that can explain cross-sectional variation (p. 1648-1649). For De Long et al. (1990a) market anomalies are linked to some unsophisticated investors misguided (and unpredictable) beliefs about future cash flows and risks, which cannot be fully corrected by arbitrageurs. Shleifer (2000, p.12) suggests that sentiment "reflects the common judgment errors made by a substantial number of investors, rather than uncorrelated random mistakes." A similar definition is used in Shiller (2000) to at least partially explain the cause for the formation of the technology bubble of the late 1990s. Overall, sentiment can be seen to be classified as the irrational (or not fully rational) portion of asset valuation (Shleifer, 2000). Following these definitions, the literature review aims to summarise the advancements both in the theory and empirical evidence of the importance of investor sentiment in financial markets, starting with its evolution in response to the Efficient Market Hypothesis (EMH).

3.2.3 Excess Volatility: Questioning of Classical Theories

An efficient market is one in which asset prices always fully reflect all available information (Fama, 1970). If this were the case and financial markets were perpetually in an efficient state, then sentiment would have no influence on prices and consequently be of little consequence. According to the Efficient Market Hypothesis, in an efficient market, when new information is released it is absorbed instantaneously into the price, which is then set at the new fundamental value. In consequence, prices should only change in light of new information and investors should not be able to systematically obtain abnormal returns, since only new information related to the fundamental value of the security should influence the price. There are three principles underlying the theories of efficient markets: a small part of investors can be considered fully rational and correspond to arbitrageurs or institutional investors, who value securities according to its true fundamental value, as the net present value of future cash flows discounted to account for risks. Second, unsophisticated investor (or "retail" or individual" investors) can be considered "irrational" by comparison, but the results of their trade does not necessarily create divergence in efficient pricing, because their trades can sometimes offset each other. Finally, even in the event of unsophisticated investors all trading in a similar way, the behaviour of arbitrageurs will actively trade away the potentially inefficient impact of unsophisticated investors on pricing. Advances in behavioural finance have proposed challenges to these notions, both theoretical and empirical, as shown in the next two sections.

3.2.4 Practical and Theoretical Questions

This section presents arguments questioning the theoretical foundations of the EMH. Miller (1977) challenges the first principle underpinning the EMH, arguing that it is improbable to think that investors make matching judgements for each security in regard to return and risk, all in light of uncertainty and difficulties related to forecasting. In reality, given the context of uncertainty, it is more likely that rational people would actually make different forecasts for returns of the same security. Further, it is likely that some investors would usually use a rule of thumb approach to value assets instead of following the axioms of rationality (Tversky & Kahneman, 1974; Kahneman & Tversky, 1979).

Secondly, on the third principle of the EMH described above, the role of arbitrageurs in efficient price setting can in fact be complicated by the reality of financial markets. If investor sentiment could lead to mispricing through excess optimism, the theory of efficient markets would dictate that this would be limited by actions from rational arbitrageurs. Yet, arbitrageurs may face several problems in markets: first, there might not be adequate substitutes available for existing positions. Second, mispricing (deviation from fundamental value) can actually worsen before it is corrected, increasing arbitrageurs exposure in the short term. For example, institutional investors are often acting as managers on behalf of other investors and might be evaluated often, which would reduce their investment horizon, and might face liquidity constraints before the mispricing is fully corrected. More specifically, Shleifer and Vishney (1997) contribute to this research by exploring the limits of arbitrage theory (one of the main aspects of the EMH). They argue that arbitrageurs face significant constraints such as risk and financing costs, and at times these constraints thwart their ability to take offsetting price positions against irrational investors; this void of successful arbitrage activity allows prices to further deviate from their fundamental values over time.

Further to this point, constraints on short sale might limit arbitrageurs ability to be reflected into market prices (Brunnermeier & Pederson, 2009), which could lead to arbitragers being able to prevent assets from severe undervaluation, however, not from severe overvaluation. These notions suggest that even in a market formed of rational agents, investors may not possess the tools to limit extreme asset price rises such as bubble formation. A number of papers document the absence of short-selling in mutual funds (see for example, Diether et al., 2002; Almazan et al., 2004; Hong & Stein, 2007). The absence of short selling is also reflected by retail investors, as research has shown they rather trade in times of increasing stock prices (Miller, 1977; Harris & Raviv, 1993; Daniel et al., 1998; Duffie et al., 2002).

These observations suggest in actual markets, in light of limits to arbitrage, there are limits to efficient market theories, and investor sentiment may lead to mispricing even over longer time horizons (Shleifer & Summers, 1990; Houge et al., 2001; Chen et al., 2002; Scheinkman & Xiong 2003; Hong et al., 2006). These findings suggest that sentiment can play a role in asset mispricing, and particularly in over-pricing.

These objections to the theoretical underpinnings of the EMH particularly in light of limited arbitrage (Merton 1987; Lee et al., 2002) paved the way for a new school of thought examining human fallibility in competitive markets. This field of behavioural finance has resulted in several new theories of market pricing. One of the most notable contributions is the model proposed by De Long et al. (1990) who define three types of investors: (i) Positive feedback traders who follow pro-cyclic patterns (ii) rational informed speculators who maximise their future pay-out in light of the expected shock, and (iii) uninformed traders who trade only on currently available information. The model suggests that arbitrageurs, in certain situations, can be completely rational and have incentives to increase a present mispricing pattern, as opposed to eliminating the mispricing pattern. The possibility of arbitrageurs engaging in pro-cyclical trading behaviour to take advantage of the pricing pattern is at the origins of the positive feedback effect.

Barberis et al. (1999) and Daniel et al. (1998) present models where investors are subject to psychological biases. Daniel et al. (1998) distinguishes between informed risk-neutral agents and uninformed risk-averse agents and posits that self-attribution bias would lead to short-term positive autocorrelation of returns while overconfidence would lead to long-term negative autocorrelation. Barberis et al. (1999) suggest that investors place a higher emphasis on wealth losses than gains. Hong and Stein (1999) rely on the notion of bounded rationality of investors, which could lead to under-reaction of short-term traders, while strategies trading on momentum are based on following trends which in turn can result in long-term overreaction. Regardless of their emphasis on investors' bias or bounded rationality, these behavioural models of investors are able to present arguments to explain underreaction and overreaction patterns from a theoretical standpoint.

3.2.5 Empirical Questions

A great deal of research has also been conducted which empirically challenges some of the foundations of the EMH, and notions of sentiment have been tested through these challenges. For example, the EMH posits that asset prices fully reflect all information, and thus investors should be unable to achieve greater than market returns. In consequence, a reoccurring pattern in market returns should theoretically not exist (such as an anomaly), nor should a trading strategy be capable of performing above average market returns particularly using variables to forecast returns (Shleifer 2000). This section explores some of the major empirical findings which question the EMH through sentiment.

If one could design a portfolio trading strategy based on reoccurring patterns, then one could in theory earn abnormal profits. One application of this idea is the existence of contrarian trading (buying previous losing stocks and selling previous winners) originated with Levy (1967), who claimed one could obtain abnormal profits using such a strategy. This idea builds on the notion that long-term trends tend to reverse and short-term trends tend to continue. Jegadeesh and Titman (1993, 2001) examine both over- and underreaction and their results suggest the presence of momentum [i.e. following trend which can in turn lead to overreaction model as described above (see Hong & Stein, (1999)] in asset prices. Specifically, their findings suggest that in the short term, prices tend to follow previous directional movements (i.e. in the short term, if prices have been increasing, they tend to continue to increase), which presents evidence in favour of under and overreaction in stock markets. A great deal of research has been conducted documenting the presence of these concepts empirically in markets see: (DeBondt & Thaler 1985, 1987; Cutler et al., 1991; Hong & Stein 1999; Barberis & Shleifer, 2003; Stambaugh et al., 2012).

An approach suggested by Baker and Wurgler (2007) is to take a bottom-up approach and explore the individual biases of investor psychology. These biases include (but are not limited to) overconfidence, overreaction, underreaction, feedback trader models, etc. How these behaviours occur across markets is based on psychological concepts.

For example, people tend to underreact to the onset of new information, suggesting that they can be slow to incorporate it into their buy-sell decision-making process. This concept is further supported by the idea that, in some cases, the actual diffusion of information can be slow [for example, (Hong & Stein, 1999) argue that bad news travels slowly]. Similarly, there are psychological explanations for the occurrence of overreaction such as positive feedback trading[39] (DeLong et al., 1990b; Hong & Stein, 1999), herding behaviour[40] (Shiller, 2001), heuristics[41] (Shleifer, 2000), and self-attribution biases[42].

Further research into asset price movements and the existence of trading strategies capable of earning abnormal returns led to work examining *the explicit role of sentiment* in asset prices. This research builds on these individual investor principles and behavioural foundation to attempt to explain these perceived weaknesses in the EMH through measurements of investor sentiment.

[39] Refers to a situation where investors purchase securities when prices rise, and sell securities when prices fall.
[40] Suggests that investors can have financial incentives to behave in accordance with a group. Suggesting that investors are afraid of missing opportunities (by not following a group) or failing alone (if we choose our own path which proves to be less prosperous).
[41] When investors follow a set of rules be self-prescribed or common place.
[42] When an individual attributes successes to their above average skills, and failures to bad luck.

3.2.6 Testing for the Role of Investor Sentiment in the Stock Market.

Recently, a large volume of research has explored sentiment's role in price setting (Chung et al., 2012; Chi et al., 2012; Zou & Sun, 2012; Baker & Wurgler, 2006, 2007; Tetlcok, 2007).

Brown and Cliff (2005) find that when investor sentiment increases with the market prices of securities, the build-up in optimism leads to a prolonged period of overvaluation in price levels. Additionally, when investor sentiment is increasingly pessimistic and with security prices falling, this can lead to prolonged periods of undervaluation.

Chung and Hung (2012) find investor sentiment levels have predictive power on asset prices in economic expansion periods. Chi et al. (2012) find further support for investor sentiment's predictive power in the Chinese stock market; they also find that stocks with high levels of consumer sentiment have higher returns than securities with low levels of consumer sentiment. Baker and Wurgler (2006, 2007) expand this literature in several notable ways. They find that large size stocks are less likely to be affected by consumer sentiment levels[43]. Thus value-weighted returns may overshadow the predictive effect of sentiment (which is one of the reasons this thesis uses equal-weighted returns). They further contribute to this area arguing that investor sentiment levels are mean reverting. They find that waves of increasing sentiment during periods of economic expansion create periods of over-pricing. This is particularly pronounced with stocks which are difficult to value or to arbitrage. They argue that technology stocks are hard to value and suggest that

[43] It is important to note that here "consumer sentiment" levels are referring to a level of consumer sentiment in a general sense per the questions from the survey questionnaire. This is in contrast to sentiment levels about a specific company.

the Technology Bubble of the late 1990s was caused by extraordinary levels of investor sentiment. Tetlock (2007) explores media and stock prices using a popular Wall Street Journal column, chiefly finding that high levels of pessimism in the media predict downward pressure on stock prices.

3.2.7 Investor Sentiment and Country Specific Factors (Cultural)

Several contributions have also shown that cultural differences can have a significant impact on the role of sentiment in investment behaviours (Sapienza et al., 2006). As such, several documented phenomenon can be explained through country-specific culture. One example of this is herding behaviour among institutional investors in specific regions, as shown by Hong et al. (2004). Otoo (1999) finds that investors from different cultural backgrounds even interpret information differently. This idea can help explain the popularity of momentum trading strategies in the US in comparison to Asia (Chui et al., 2010). Further, Odean (1998) finds that overconfident investors trade more. Consequently, an examination of international evidence on consumer confidence and stock returns can help empirically explain the previous culturally unique relationships.

International evidence explicitly examining consumer confidence and expected stock returns across markets suggests sentiment negatively forecasts aggregate stock market returns across countries (Schmeling, 2009). Through a study over 18 countries, they find that on average, sentiment negatively forecasts aggregate market returns. Additionally, sentiment is found to tend to have a larger impact on returns in countries which have less market integrity and are culturally prone to herd-like behaviour and overreaction[44]. Further research by Baker, Wurgler, and Yuan (2009) compile indices of sentiment for six market indices and determine that sentiment is both statistically and economically a significant contrarian predictor of market returns (periods of high sentiment are followed by periods of low returns) across countries. Dong and Guo (2014) examine how global sentiment impacts domestic (over 50 indices across the globe in total) prices, as well as how domestic sentiment impacts domestic prices. They find that US sentiment exhibits more extreme volatility patterns than other regions and more specifically, that changes in global sentiment can have a Granger causal relationship on US prices and US sentiment levels. Corredor et al. (2013) examine four major indices in Europe and find sentiment to be significant in influencing returns. They also find varying intensity across markets suggesting a causal link in both stock characteristics and cross country cultural and institutional differences, further highlighting the sensitivity in results to the choice of proxy for sentiment. Chang et al. (2012) attribute the varying impact factor of sentiment on stocks to country specific factors with importance focused on the differences with information quality, corporate governance, and even legal systems. Upon reviewing past research, there appears to be inconsistency on whether sentiment affects prices.

3.2.8 Sentiment Proxies

While there are a large number of academic studies examining the role of sentiment in stock market pricing, a wide range of indicators have also been used to proxy for

[44] See Chui et al. (2010), herd-like behaviour relates to high correlation of noise traders based on overly optimistic or pessimistic expectations.

investor sentiment. Bandopadhyaya and Jones (2006) propose an interesting classification of these indicators into five categories:

(i) Measures of general economic optimism (e.g. consumer surveys such as the University of Michigan Consumer sentiment index). In past studies, these have been found to correlate with investor sentiment surveys (Fisher and Statman, 2003) and with returns of small stocks, which are held disproportionately by retail investors (Qui and Welsh, 2006).

(ii) Measures of stock market investors' optimism (e.g. investor surveys, or Put-to-Call ratios)[45]. For example. Brown and Cliff (2005) use Investor Intelligence data and show that it is an effective predictor of stock market returns over longer time horizon. Lee et al. (2002) also find that excess returns are positively correlated with Investor Intelligence measures of sentiment.

(iii) Measures of stock market riskiness (e.g. liquidity, turnover or first day returns on the IPO markets). Baker and Wurgler (2006) construct) a sentiment index based on these variables, and shows that a high value tend to generate subsequently low returns for stocks which are not attractive to arbitrageurs (small stocks, non-dividend paying stocks, etc.)

(iv) Measures of riskiness of a particular stock i.e., BETA. This is used in Capital Asset Pricing Models to embody the risks of a particular stock or portfolio in comparison with the markets.

[45] Put/call ratio, measuring the ratio of trading volume of put options to call options, is considered a good bearish indicator (Brown and Cliff, 2004).

(v) Finally, measures of risk aversion e.g. Volatility Index (VIX), the Chicago Board Options Exchange's measure of markets' expectations for volatility, also called the "investor fear" or risk aversion indicator, has been shown to be related to periods of market "turmoil" (Whaley, 2000).

Although these indices have been used extensively in the academic literature or by practitioners, they have not been tested in the context of stock market crises. In this essay, it is useful to conduct the analysis using a selection of these proxies, to be able to compare their performance in predicting stock market crises.

3.2.9 Sentiment and Stock Market Crises

Although much research has been conducted on the explicit relation between investor sentiment and stock market returns, there is relatively little research that has focused on financial crises. A few studies have anecdotally mentioned sentiment as a possible cause in market crises (De Long & Shleifer, 1991; Shiller, 2000), yet there has been remarkably little research explicitly examining this link. One recent paper is an example of this research: the findings of Zouaoui et al. (2011) suggest that investor sentiment is a significant predictor of stock market crises.

Building on this paper, this essay aims to further examine this relation through the use of a model based on the literature of early warning signals (see Methods section for a review of *this* literature). The essay compares the predictive value of investor sentiment across indices and regions. Further, the essay will expand this research by relating sentiment and early signs of recoveries, to investigate whether this link holds.

3.3 Financial Crises in the Data

3.3.1 Datasets

This essay focuses on stock market crises in five countries in total: three European countries (France, Germany, and the UK), Japan, and the United States. The dataset consists of monthly index values for the CAC, DAX, FTSE, NIKKEI, DJIA, and NASDAQ indices. The time period covers 1995 through 2014, which should include the two most recent financial crises (the technology bubble crash and the sub-prime crisis). In this essay, the link between investor sentiment and stock market crises is examined. In particular, the essay aims to determine whether investor sentiment can be used to predict the probability of occurrence of crises and specifically which proxy for sentiment performs the best. Therefore, it is necessary to turn to the literature to identify stock market crises (and their components, e.g. crash, trough, recovery) in the dataset.

3.3.2 Identifying Crises

In this essay, the identification of crises in each of the indices over the sample period is accomplished by following the model used in Patel and Sakar (1998), and Zouaoui et al. (2011). An indicator is created called CMAX, which is designed to detect extreme price variations in the indices. The CMAX compares the current index value with the maximum value over a given period T. It is defined in Eq. (22):

$$CMAX = \frac{P_{i,t}}{max(P_{i,t-T} \ldots \ldots, P_{i,t})} \tag{22}$$

Where $P_{i,t}$ is the monthly value of a stock market index for country i, at time t. The denominator represents the maximum value of the index i over a rolling period of T months. In this essay, following Patel and Sakar (1998), a period T of 24 months is chosen[46]. The CMAX is thus equal to 1 when the current period experiences the highest monthly index value in a two-year period, which would reflect a long period of price increases. In contrast, CMAX gets closer to zero during a period of falling prices.

The CMAX can then be used to identify a threshold, indicative periods of abnormal price falls (i.e. when CMAX is abnormally low). Following the literature, a threshold is defined as the mean of the CMAX over the sample period, minus two standard deviations. The threshold is used to create a new dummy variable which is equal to 1 when a crisis is identified (CMAX falls under the threshold) or zero otherwise. This is shown in Eq. (23) below:

$$C_{i,t} = \begin{cases} 1 & if\ CMAX < \overline{CMAX} - 2\,\sigma_{i,t} \\ 0 & otherwise \end{cases} \tag{23}$$

Where $C_{i,t}$ is the crisis dummy, indicating whether a crisis is identified in country i at time t, $\overline{CMAX}$ and $\sigma_{i,t}$ are respectively the mean and standard deviation of CMAX over the entire sample. Because it is likely that falling prices will continue after the crisis is detected, it is possible that this indicator would detect the same crisis a number of times. In order to avoid double-counting, a crisis is automatically eliminated (i.e. is allocated a $C_{i,t}$ value of zero instead of 1), if it is detected twice during the same twelve-month period.

A number of concepts can be defined with regard to crises following the work of Patel and Sarkar (1998), and Zouaoui et al. (2011). These are summarised in Table 34.

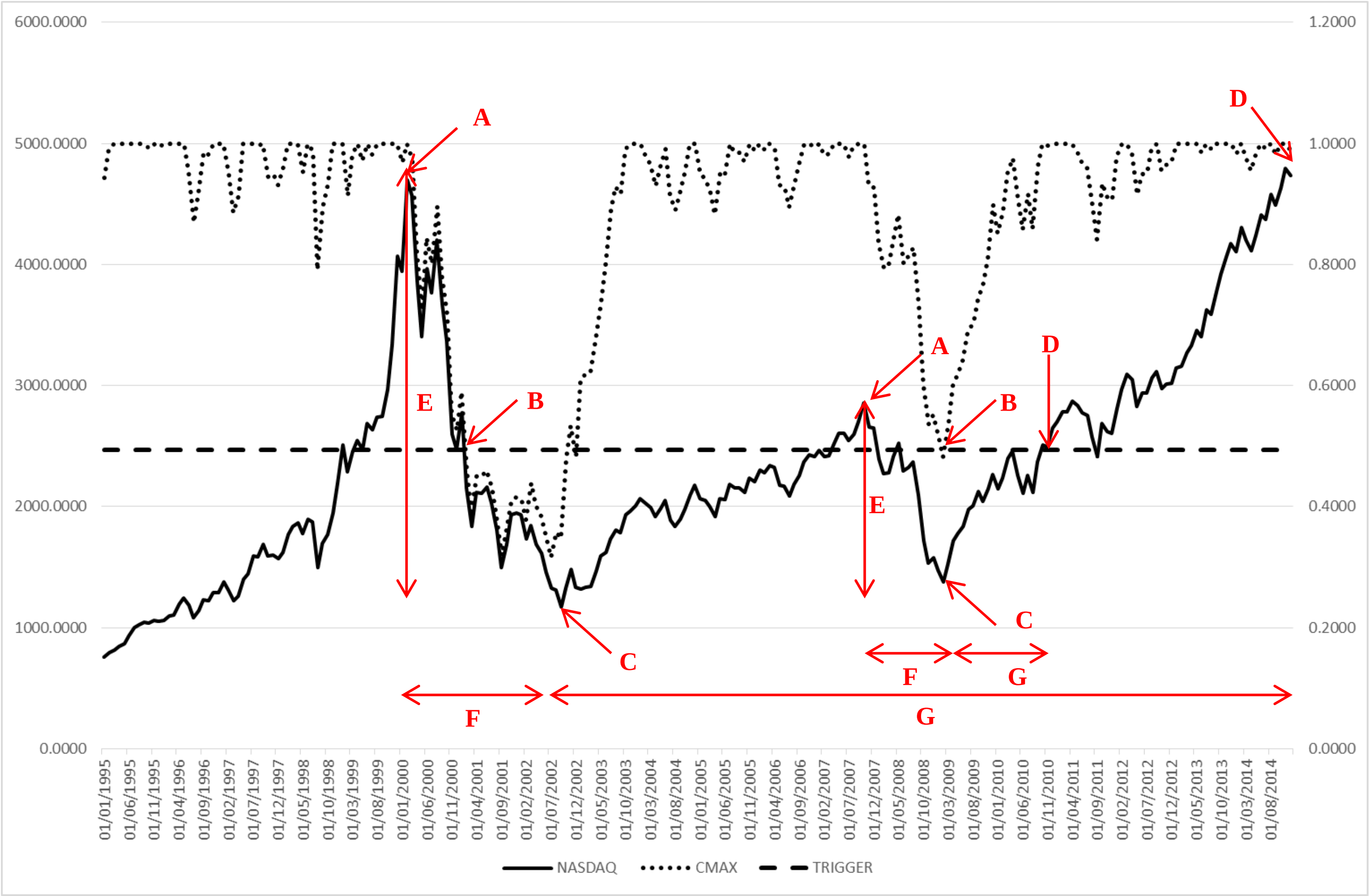
[46] Because the CMAX is backward looking, a time frame longer than 24 months would necessarily involve dropping a relatively large number of observations in the sample.

Table 34 Defining Crisis Concepts

Concept number	**Concept**	**Definition**
A	Beginning of Crises Marker	The month during which the price index hits the maximum over the two year window, prior to the month when the crash is triggered.
B	Beginning of the Crash Marker	The month the price falls below the threshold level.
C	The Crash Trough	The month when the crash reaches its minimum.
D	The Recovery	First month post-crash, when the index reaches its pre-crash maximum value.
E	Magnitude of Crises	The difference between the value of the index at the maximum and minimum prices.
F	Length to Trough	Number of months between the month that triggered the beginning of the crisis and the month of the trough.
G	Recovery Period	The number of months for the index to return to the maximum.

These concepts are illustrated using the NASDAQ from January 1995 through December 2014 in Figure 16. There are two crises detected for the NASDAQ over the sample period. These are identified where the CMAX value drops below the threshold (mean CMAX minus two standard deviations), shown as points B on the graph.

Figure 16 Illustrative Concepts in the NASDAQ



The first crisis occurs in February 2000, where the monthly index starts falling (point A). The crash is detected a year later in February of 2001 (point B). The trough is not reached until the end of September 2002 (point C), and did not recover until late 2014 (point D). This crisis is considered particularly unique because the magnitude of the crash was so large (75% loss in index value, represented by the interval E) that the index pre-crisis value was not reached again until nearly the end of the sample (November 2014). There were 31 months to the trough (F), and 177 months until the recovery (G). This crisis corresponds to the crash of the information technology bubble, or dot com bubble. Its impact was particularly felt on the NASDAQ, as a composite index of primarily technology companies. Prior to this crisis, the NASDAQ exhibited extremely high returns (1 year return of 105%, and 3 year of nearly 57%)[47].

The second crisis in the NASDAQ over the sample period begins in October 2007, with a crash detected by the CMAX in February 2009. The trough was reached in February 2009, the same month where the indicator detects the crash. Because the pre-crisis index value is lower in this second crisis than during the technology crash, the index recovers sooner (April 2011). This second crisis corresponds to the sub-prime financial crisis: it took 25 months to reach the trough of the crisis, and 162 months until recovery, while the index lost 72% of its value.

[47] Because of the magnitude of the loss of value on the NASDAQ during the technology crash, a second crisis is detected using the CMAX in March 2002. However, following the literature, this crisis is rejected as it constitutes double-counting of the technology crash.

3.3.3 Crises in the Data

Following the methods and indicators described above, crises can be identified in the five remaining indices in the dataset. The index and CMAX indicators and thresholds are represented graphically for the CAC, DAX, FTSE, Nikkei, and DJIA in Appendix 5.

Table 35 presents a summary of all the crises identified in the data for the six indices and five countries. During the sample period, twelve crises are identified, with two identified in each index. These correspond to the two well-documented events in the technology bubble and the sup-prime financial crises.

Table 35 Crises Detected in the Data

					Duration of Crises			Annual Returns before Crises	
Index	Beginning of Crises	Beginning of Crash	Date of Trough	Date of Recovery	Months to Trough	Month to Recovery	Price Decline to Trough	1 Year	3 Years
NASDAQ	29/02/2000	28/02/2001	30/09/2002	28/11/2014	31	177	-75.04%	105.27%	56.59%
	31/10/2007	27/02/2009	27/02/2009	29/04/2011	16	42	-51.81%	20.81%	13.26%
DJIA	29/12/2000	30/09/2002	28/02/2003	31/12/2004	26	48	-33.67%	3.21%	8.43%
	31/05/2007	31/10/2008	27/02/2009	31/01/2013	21	68	-47.42%	20.39%	16.13%
FTSE	31/08/2000	31/07/2002	31/01/2003	31/10/2007	29	86	-46.54%	6.82%	11.59%
	31/10/2007	31/10/2008	27/02/2009	31/10/2013	16	72	-43.02%	9.66%	13.31%
CAC	31/08/2000	31/07/2002	31/03/2003	N/A	31	N/A	-60.48%	44.37%	33.95%
	31/05/2007	28/11/2008	27/02/2009	N/A	21	N/A	-55.73%	23.81%	18.58%
DAX	31/08/2000	31/07/2002	31/03/2003	30/04/2007	31	80	-66.41%	36.91%	23.24%
	31/12/2007	30/01/2009	27/02/2009	31/05/2013	14	65	-52.35%	22.29%	23.78%
Nikkei	31/03/2000	28/09/2001	30/04/2003	N/A	37	N/A	-61.49%	28.42%	5.35%
	29/06/2007	31/10/2008	27/02/2009	N/A	20	N/A	-58.27%	16.98%	16.17%

The months during which the crash is detected varies in the different markets. For the technology bubble, the crash is felt on the NASDAQ well ahead of the other indices in February 2001. The NIKKEI followed as the CMAX reached the threshold in September 2001. The three European markets (CAC, DAX, and FTSE) follow with the crash being detected simultaneously in July 2002. Finally, the DJIA is the last index to enter the crash, as it is mostly composed of industrial stocks, and less exposed to technology market valuations. The DJIA is also the index which recovers relatively fast (2004). For other indices, this crisis has been particularly difficult to recover from. Although the FTSE and DAX recovered in 2007, the NASDAQ recovered only in 2014, while the NIKKEI and the CAC had not recovered at the end of the sample. This crisis has been particularly long, with an average of 30.8 months from beginning to trough, and 56.8% loss in monthly index value (largest for the NASDAQ and smallest for the DJIA).

In contrast, for the sub-prime crisis, the DJIA is one of the first to experience the crash (October 2008), alongside the NIKKEI and FTSE. They are followed by the CAC, DAX, and NASDAQ respectively. The crisis was less significant than the previous one, in that it took on average 18 months to reach the trough and 51.3% value was lost on average.

3.4 Methods

This essay aims to explore both the optimal proxy for sentiment and its performance as a predictor of financial crises. Specifically, this question is best addressed in the context of crises forecasting and, therefore, this essay chooses to follow the literature on Early Warning Systems.

3.4.1 Early Warning Systems of Financial Crises

Early Warning System (EWS) indicators have gained attention in recent years, especially in light of the sub-prime financial crisis. Logit models are suggested as the preferred model in developing EWS, as suggest by (among others) the European Central Bank (Bussiere & Fratzscher, 2002). Research has paid particular attention to advanced economies (Barrell et al., 2010; Babecky et al., 2013). These analyses attempt to develop a model using a number of independent variables to create signals or warning signs of future crises. This research can be of interest to researchers, as well as policy makers.

The findings from the most recent literature show a better fit of model using multinomial logit with three outcomes in comparison to binomial logit. Multinomial logit reduces the number of errors associated with the model (e.g. Type 1 and Type 2 errors), Bussiere and Fratzscher (2006) (multinomial regression based EWS). These models are designed to create a warning indicator when variables change, which have historically shown patterns of similar change before economic crises. This is then used to predict the probability of occurrence of a crisis. These patterns can then be used by policy makers in their decision-making to attempt to curtail the underlying causes of economic decline.

Early Warning System models are used to predict the probability of occurrence of a multitude of crises across varying socio-economic country samples. For example, including financial/economic crises in OECD countries (Barrell et al., 2010; Alessi & Detken, 2011), Asian Pacific countries (Wong et al., 2010), EU and Organization for Economic Cooperation and Development (OECD) countries (Bebecky et al., 2013), and sub-Saharan African countries (Caggiano et al., 2014). However, EWS is also used for predicting the probability of occurrence of crises in a variety of markets, including currency markets (Comelli, 2014), emerging markets (Cumperayat & Kouwenberg, 2013). EWS have become an area of growing interest to researchers, with potential for future research as EWS models are ever improving and needed.

3.4.2 The Model

Following this literature, the analysis used in this essay follows a discriminant analysis method with a logit model. The rationale behind this choice of model is that the essay attempts to identify specific behaviour in a number of selected independent variables before a crisis is detected, which could help in identifying a future crisis. In other words, the probability of occurrence of a crisis over a given period is estimated using a number of independent variables, including sentiment variables.

The logit model is used to predict the probability of the outcome of a dependent variable (here linked to the occurrence of a crisis within a given time period). In logistic regression, the dependent variable is categorical (i.e. it is not a continuous variable). In simple logit models, the dependent variable is binary, it can only represent two outcomes. In multinomial logit models, the dependent variable can take more than two values.

3.4.2.1 The Dependent Variable

The construction of dummy variables to embody the occurrence of a crisis over a given time period is essential in the formulation of the logit model. I follow the methodology in Bussiere and Fratzscher (2006) and Zouaoui (2011) to construct the dependent variable.

Using the crises identified in Table 35, I define a binary dummy equal to one if the current month is considered during a crisis, as well as during the 12 months[48] preceding the crisis, and equal to zero in all other periods. This variable is used as a dependent variable in the binary logit estimation, to predict the probability of occurrence of the crisis:

$$A_{i,t} = \begin{cases} 1 & if\ x \in [1, \dots, 12]\ such\ that\ C_{i,t+x} = 1 \\ n.a. & if\ x \in [1, \dots, 11]\ such\ that\ C_{i,t-x} = 1 \\ 0 & otherwise \end{cases} \tag{24}$$

Following Bussiere and Fratzscher (2006) and Caggiano et al. (2014), the 11 months following the crisis are excluded from the dummy variable A. This newly created variable is then used in a logit regression to estimate the predicting power of a number of independent variables (these are detailed in the next section).

Additionally, Bussiere and Fratzscher (2006); Borrio and Drehmann (2009); and Caggiano et al. (2014) suggest that instead of eliminating post-crisis periods from the analysis, the model forecasting accuracy can be improved by creating a dependent variable with more than two outcomes and using the multinomial logit model. Based on these findings, a new dependent variable B is created in this essay as well, which is similar to variable A, but takes the value 2 during the 11 months following the crisis, as follows:

$$B_{i,t} = \begin{cases} 1 & if\ x \in [1, \dots, 12]\ such\ that\ C_{i,t+x} = 1 \\ 2 & if\ x \in [1, \dots, 12]\ such\ that\ C_{i,t+x} = 1 \\ 0 & otherwise \end{cases} \tag{25}$$

[48] 12 months and 24 months are commonly used in the literature. Both were tested in this essay, but as the results appeared similar, the essay only shows results for12 months.

Variables A and B are then used as dependent variables in the logit and multinomial logit models described below.

3.4.2.2 The Logit Model

The model used for analysis in this essay is a logit model which attempts to explain the dependent variables $A_{i,t}$ and $B_{i,t}$ with a number of predictors or independent variables. With this methodology, I attempt to estimate the probability that the dependent variable is equal to 1 given the value of the independent variables. By using the dependent variables described above which are created by considering the 12 months preceding the crisis, the model thus attempts to predict the probability of whether a crisis is likely to occur within a 1-year period, but is not designed to be able to predict the exact month a crisis would occur. The basic logit model used in this essay with variable A is defined as follows:

$$\Pr(A_{i,t} = 1) = f\left(\alpha_0 + \sum_{k=1}^{n} \alpha_k E_{i,t}^{k}\right) \tag{26}$$

With a logistic function: $f(w) = \frac{e^w}{1+e^w}$

Where the probability that the crisis variable A is equal to 1 is determined by a vector of explanatory variables E^k, associated with a vector of coefficients α_k. For the multinomial model, variable A is replaced by B in Eq. (26).

In this essay, two types of explanatory variables are included in the model. First, standard macro-economic variables are included following findings from the empirical literature. Second, new variables embodying investor sentiment are tested. Three models are actually estimated to establish the added contribution of sentiment in predicting the probability of occurrence of a crisis over a 1-year period. The three models are shown in Eq. (27), (28), and (29) below.

Model 1 $$\Pr(A_{i,t} = 1) = f\left(\alpha_0 + \sum_{k=1}^{n} \alpha_k M_{i,t}^{k}\right) \tag{27}$$

Model 2 $$\Pr(A_{i,t} = 1) = f(\alpha_0 + \beta\, SENT)^{49} \tag{28}$$

Model 3 $$\Pr(A_{i,t} = 1) = f\left(\alpha_0 + \sum_{k=1}^{n} \alpha_k M_{i,t}^{k} + \beta\, SENT\right) \tag{29}$$

With a logistic function: $f(w) = \frac{e^w}{1+e^w}$

3.4.2.2.1 EWS Model Choices and Logit Review

Although logit has been established in the literature as the baseline model for EWS, alternative models are available such as Trait Recognition (Kolari & Wagner, 1996). Candelon et al. (2009) propose a framework for statistically evaluating the performance of two models used in EWS. They compare data from 12 countries over the span of 20 years 1985-2005. They compare a Markov switching model, and logit model, using the following criteria: the Kuiper scores, Bayesian error rate, and ROC curves, finding the logit model to outperform the Markov switching model.

However, I also acknowledge the potential downsides of the logit approach, such as the sensitivity of dates with regard to the crisis, thresholds used for identification of crises, and lack of theoretical reasoning for logit.

[49] There are three types of investor sentiment tested in Model 2, and likewise three added to Model 3 resulting in seven models in total.

3.4.2.3 Macroeconomic Variables

Following the literature, the well-established link between macroeconomic variables and market prices is recognised in the methodology. I use a set of macro-economic variables gathered from a review of the literature. According to Boucher (2004) and Coudert and Gex (2008), the variables which exhibit the strongest link in explaining stock market crises are the variables listed in Table 36. The variables include the cyclically adjusted price to earnings ratio (*CAPE*), interest rates (*INT*), inflation index (*IFR*), and the term spread between long-term and short-term government interest rates (*TS*); also gross domestic product (*GDP*), and industrial production (*PROD*). The sources of data for all these variables are summarised in Appendix 6.

Table 36 Macro-economic Variables

Variable code	Variable
CAPE	Price to earnings ratio reported by Shiller[50]
INT	Interest Rates
TS	Term spread
GDP	Gross domestic product
PROD	Industrial Production
IFR	Inflation Rate

Following mean reversion theory, Poterba and Summers (1988), the price to earnings ratio is also included (*CAPE*), as proposed by Campbell and Shiller (1998): Mean reversion theory suggests after prolonged periods of high price to earnings ratios, periods where

[50] Cyclically Adjusted Price to Earnings Ratio as reported on Robert Shiller's website "http://www.econ.yale.edu/~shiller/data.htm"

prices revert to more historic means will follow. Industrial production (*PROD*)[51] is included as it is a leading indicator in EWS models, Frankel and Sarvelos (2010). Also included is a measure of inflation (*INF*), as markets are historically negatively correlated with high inflation, Fama and Schwert (1977). Term spread acts as a leading indicator as an inverted yield curve has preceded nearly every financial crisis since the 1960s, Zaloom (2009). Finally, *GDP* is included as a measure of the stage of economic development (Frankel and Saravelos, 2012).

3.4.2.4 The Sentiment Variable

There has long been an ongoing debate amongst researchers as to the appropriate measure of investor sentiment. However, there seems to be some agreement that a perfect proxy for investor sentiment does not exist, with each proxy having benefits and problems. In the absence of such agreement, I explicitly test three commonly used proxies from three different categories of proxies described in the literature review section. The selection was based jointly on findings from the existing literature linking sentiment and stock returns, and availability of data. These three proxies seem to be commonly used as proxies, for the US, but the EU (Fratzscher, 2002; Martikainen & Puttonen, 1996) and Japanese markets as well, (Park & Lee, 2003). The three proxies used in the essay are: The Baker and Wurgler (2008) Orthogonalised Sentiment Index (*BWSI)*, the Volatility Index (*VIX*), reported by the Chicago Board of Exchange (CBOE), and the Put to Call Ratio, (*PCR*)

[51] Industrial production is included in addition to *GDP* in EWS forecasting following Frankel and Sarvelos (2010).

reported by the CBOE. More details about these sentiment proxies are provided in the Table below.

Table 37 Sentiment Proxies

Sentiment Proxy	**What it measures**	**How it is constructed**	**Source of data**	**Example previous studies**
Put/Call Ratio (PCR)	Stock market optimism	The ratio of the trading volume of put options to the call options which trade on the CBOE.	Chicago Board Options Exchange (CBOE)	Bandopadhyaya and Jones (2006, 2008); Park & Lee (2003); Xing, Zhang & Zhao (2010)
Baker and Wurgler (2007) Sentiment Index (BWSI)	Stock market riskiness	An index constructed by the CBOE, which shows the markets expectation of the coming 30 days volatility. A measure of expected volatility calculated as 100 times the square root of the expected 30-day variance of the S&P 500 rate of return. The variance is annualized and VIX expresses volatility in percentage points.	Wurgler's website	Baker and Wurgler (2007)

Volatility Index (VIX)	Investor risk aversion	A composite index of sentiment, based on the common variation in six underlying components: the closed-end fund discount, NYSE share turnover, the number and average first-day returns on IPOs, the equity share in new issues, and the dividend premium	Chicago Board Options Exchange (CBOE)	Bandopadhyaya and Jones (2006, 2008); Ang et al. (2006); Fisher & Statman (2000)

Although commonly used, these proxies are not without problems of their own. Previous studies have posited that sentiment indicators could reflect not only investor sentiment, but potentially also a component reflecting changes in fundamentals (e.g. business cycle, macroeconomic conditions. Baker and Wurgler (2006, 2007) explicitly address this possibility by orthogonalising the data to macroeconomic variables (discussed later in the essay). Similarly, the *PCR* and *VIX* proxies could be measuring changes in actual market risk itself rather than investors' attitude towards risk (Kumar and Persaud, 2002), but is unfortunately impossible to distinguish between the two.

There exist many other choices to proxies for sentiment; to name a few: the University of Michigan Consumer Sentiment Index, the net cash flow into mutual funds (Randall, Suk & Tully, 2003), Barron's Confidence Index (Lashgari, 2000), the Risk Appetite Index (Kumar & Persaud, 2002). However, all these measures face potential benefits and problems.

Consequently, this essay has chosen three major proxies for investor sentiment to measure their performance in a EWS model predicting the probability of the occurrence of stock market crises and recoveries. These sentiment variables are added as explanatory variables in the logit models, in addition and in comparison with macroeconomic variables. This has not been done previously and contributes both to the literatures on EWS of stock market crises and on the relation of sentiment proxies to asset prices in domestic and international markets.

3.4.3 Model Performance

Following the literature (Kaminsky et al., 1998; Bussiere & Fratzscher, 2006), the fit of the model can be assessed by examining its efficiency in predicting the probability of occurrence of a crisis. In each time period, the model produces a probability that a crisis will occur within the next 12 months, based on the values of the independent variables. This generated probability is a continuous variable. In order to assess the performance of the model, the probability of occurrence of a crisis generated by the model can be compared with the actual incidence of crises in the data.

Following the literature, a bounded threshold probability method is used. A probability threshold is defined: If the generated probability is above the threshold in a given period, then the model is considered to predict a crisis within the next 12 months. The threshold is effectively a limit above which the probability produced from the logit models is a signal of a future crisis. Then a test is performed to compare the model predictions with the actual occurrence of the crisis. The effectiveness of the model is measured in how well it can forecast future crises.

Table 38 outlines the possible outcomes of the test. Positive signals which support the quality of the model can be the result of only two outcomes: either a correct signal (i.e. the model predicts a crisis which does actually happen) or a correct, no signal (i.e. the model does not predict a crisis and indeed no crisis is present in the data).

Table 38 Logit Model Performance - Predicting Errors

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Logit Model</td></tr>
<tr><td>Crisis signal generated by the model</td><td>Crisis signal not generated</td></tr>
<tr><td rowspan="2">Actual Crises</td><td>The variable indicates a crisis</td><td>Correct signal</td><td>Missing signal (error 1)</td></tr>
<tr><td>The variable does not indicate a crisis</td><td>False signal (error 2)</td><td>No signal and no crises</td></tr>
</table>

In contrast, two types of errors can occur. Type 1 error occurs when the model fails to predict a crisis that occurs, and Type 2 error occurs when a crisis was wrongly predicted by the model. The proportion of correct signals produced by the models is largely influenced by the chosen thresholds. This is what Bussiere and Fratzscher (2006) describe as “the trade-off problem”: the lower the threshold, the more signals will be produced by the model, but also the more likely it is to produce false signals (Type 2 error). In contrast, a higher threshold is likely to reduce the number of crises signals, but would result in more frequent missing signals errors (Type 1). However, in practice the costs of failing to produce a signal for an actual crisis are larger than the cost of predicting a crisis which does not materialize. Following Zouaoui et al. (2011) and Bussiere and Fratzscher (2006), in this essay, two thresholds are used: 25% and 50%.

3.5 Results

Three main questions are addressed in this section. First, through the methods described above, the contribution of each independent macroeconomics variable in predicting the probability of occurrence of stock market crises in the various indices can be assessed.

Second, and more importantly as a contribution of this essay, the value-added of including investor sentiment as an independent variable to predict the probability of occurrence of crises can be evaluated by comparing the fit and performance of the model with and without each sentiment proxy. Finally, the results can be compared across the various indices, to determine which of the sentiment variables proves best as a crisis predictor. These results may vary across regions and countries (US, EU, and Asia). In this section, the results are presented in summary table forms for the multinomial logit model only.

3.5.1 Macroeconomic Variables

Table 39 presents the results of the binomial and multinomial logit model of crises probability prediction using macroeconomic variables only (Model 1), as a summary for all indices. Presenting the results of the multinomial logit model enable it to separate pre- and post-crisis periods. This avoids post-crisis periods biasing the results. Overall, the coefficients associated with macroeconomic variables are generally statistically significant and present the expected sign[52] during the pre-crisis period. Because we are interested in the value of the independent variables in predicting the probability of occurrence of crises, we focus on the analysis of results for periods preceding crises, and do not present the results for the post-crisis period.

Industrial production; used as a leading economic indicator, shows a positive sign associated to its coefficient and is statistically significant at the 1 or 10% levels in four of the six markets with the exception of the FTSE and CAC. This supports findings in the

[52] With several exceptions presented in the next paragraphs.

literature that industrial production is positively related to stock market crises (Frankel & Saravelos, 2012). Frankel and Saravelos (2012) argue industrial production may even provide better results in predicting crises than GDP as the composition of GDP changes across countries. The results suggest that periods of strong industrial production are often followed by a sharp decline in stock prices (in this case stock market crisis). Similarly, *GDP* generally presents a positive coefficient, which is statistically significant at the 1% level in the three European indices (FTSE, CAC, and DAX). This supports frequent findings in the literature (Frankel & Saravelos, 2012), which has shown that GDP levels are a leading indicator in market crises. This supports the notion that periods of high economic growth are generally directly preceding crises (Kaminsky et. Al., 1998).

For inflation, the coefficient presents a negative sign and is statistically significant at the 1% level for the DJIA[53], FTSE and CAC. In contrast, the coefficient has a positive sign for the DAX and NIKKEI, and is statistically significant at the 1% level. However, for the NASDAQ, the coefficient shows a positive sign but is not found significant even at the 10% level. The NASDAQ presents a very different pattern during this time periods from the other markets, it is generally in decline from the beginning of the sample. These findings are interesting given that a negative relation between inflation and probability of financial crises would appear counter-intuitive with policy makers' objectives of maintaining stable, positive inflation rates. This would suggest attempts to maintain price stability could actually positively influence the probability of financial crises. Interestingly, this result is also found in the literature (Zouaoui et al., 2011), who propose

[53] The coefficient for the DJIA shows a positive sign but is not statistically significant.

that the finding supported in theory, with a" paradox of credibility." Goodfriend (2001), Borio and Lowe (2002), and Borio and Zhu (2012) argue that as inflation is contained, periods of economic slowdowns are rather due to instability in the financial sector. In other words, if economic growth is relatively stable and inflation stays at a moderate range, then central banks may not have incentive to increase discount rates, which may lead to unchecked growth in financial markets leading to imbalances. This is a very timely topic and similar problems are presented in emerging markets, Montes and Peixoto (2014) use data from Brazil to highlight the dilemma of central banks facing a "paradox of credibility" and attempting to limit bubble creation in credit and stock markets.

Table 39 Results of Model 1

NASA and NASB represent the results of the binomial and multinomial logit model respectively.

Model 1	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
IPROD	0.606	0.512	0.942	0.942	0.059	0.058	0.008	0.006	0.062	0.071	0.098	0.093
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.18	-0.19	-0.75	-0.79	(0.07)*	(0.05)**	(0.00)***	(0.00)***
IFR	0.24	0.071	-0.23	-0.23	-0.192	-0.201	-0.687	-0.734	1.865	1.705	0.65	0.64
	-0.47	-0.83	-0.54	-0.54	(0.08)*	(0.08)*	(0.01)**	(0.01)**	(0.00)***	(0.00)***	(0.00)***	(0.00)***
INT	-4.932	-5.042	-5.022	-5.022	0.39	0.374	-	-	-	-	1.736	1.729
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.24	-0.25	-	-	-	-	(0.04)**	(0.04)**
TS	5.212	5.32	4.635	4.635	0.37	0.361	-0.469	-0.464	1.84	1.733	-0.52	-0.492
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.33	-0.34	-0.22	-0.24	(0.00)***	(0.00)***	-0.52	-0.55
GDP	0	0	0	0	0	0	0	0	0	0	0	0
	-0.13	-0.38	-0.3	-0.3	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.00)***	(0.00)***	-0.15	-0.17
CAPE	-0.313	-0.28	-0.729	-0.729	-	-	-	-	-	-	-	-
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-	-	-	-	-	-	-	-
Constant	-46.559	-35.563	-32.266	-32.268	-4.624	-3.959	26.351	28.392	-110.62	-101.76	-81.763	-80.189
	(0.00)***	(0.02)**	(0.07)*	(0.07)*	-0.65	-0.7	(0.05)**	(0.04)**	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	-54.03	-102.31	-37.36	-37.36	-68.97	-108.16	-62.42	-124.93	-57.29	-99.84	-61.74	-106.4
Prob>chi2	.		.		.		.		.		.	

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Interest rates[54] display an expected negative coefficient sign in the U.S. markets, and are found statistically significant at the 1% level for the DJIA and NASDAQ. This result corresponds to central banks' rhetoric that interest rates are traditionally cut during crises to limit the severity of their impacts, and re-stimulate the economy.

For term spreads, the coefficient is positive and statistically significant for three of the six markets for which data is available. This finding is particularly striking for the U.S. indices and is strongly supported by previous literature, which identifies the variable as a key early warning indicator for financial crises in U.S. markets (Alessi & Detken, 2011). The coefficient is not found significant for the NIKKEI, FTSE, or CAC. This result is somewhat intuitive for the NIKKEI as the term spread in Japan was reasonably flat over the sample period.

The results for the *CAPE* (cyclically adjusted price to earnings ratio) variable are only available for the U.S. markets, and present negative and statistically significant coefficients.

As explained in section 4.3, one of the best performance measures for a model of early warning of crises is its efficacy in predicting the probability of occurrence of crises. The findings from the Type 1 and Type 2 error tests are presented in Table 40 and Table 41

[54] Interest Rates are not presented for France and Germany as the variable is co-integrated with Term Spreads. This could be likely due to the somewhat unique relationship in the sample data corresponding to the introduction of the Eurozone.

below. The errors are presented for the two threshold levels of 25% and 50%[55]. The results for both Dummy A and B show an acceptable performance of the model: crises are predicted correctly over 55% of the time for all indices (25% threshold) and false alarms are relatively low (less than 11% of the time) for both thresholds.

Table 40 Model 1 Dummy A Probability Errors

Dummy A	Model 1					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	35.71%	23.08%	61.54%	65.38%	42.31%	50.00%
0.5	53.57%	38.46%	96.15%	88.46%	73.08%	73.08%
Type II						
0.25	16.98%	9.81%	8.41%	9.81%	14.02%	9.35%
0.5	8.02%	2.80%	0.00%	3.27%	6.54%	0.47%

Table 41 Model 1 Dummy B Probability Errors

Dummy B	Model 1					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	42.86%	23.08%	65.38%	65.38%	61.54%	50.00%
0.5	64.29%	38.46%	96.15%	88.46%	80.77%	76.92%
Type II						
0.25	7.04%	5.58%	8.37%	9.77%	5.12%	7.44%
0.5	0.94%	0.47%	0.00%	3.26%	0.47%	0.00%

[55] Type 1 errors correspond to the case when a crisis occurs, yet there was not a signal produced by the model. Type 2 errors are detected when no crisis occurs in the data, but the model predicts one.

3.5.2 Investor Sentiment

Table 42, Table 43, and Table 44 present the results of the binomial and multinomial logit models (Model 2) for each of the sentiment variables across all indices. As one of the objectives of this essay is to examine both the potential incremental benefit of sentiment in predicting the probability of occurrence of market crises and which proxy for sentiment performs the best, Model 2 examines the role of each of the three sentiment variables in absence of the macroeconomic variables. In general with all three proxies, the results show a very strong role of sentiment in predicting the probability occurrence of stock market crises, across the indices. However, results are mixed across sentiment indices. The coefficients are positive and statistically significant at the 1% level for both the *VIX* proxy and *BWSI* proxy across all markets, while with the *PCR* proxy, significance is not found in either the NIKKEI or the NASDAQ.

As the first analysis covering two major financial crises and recovery periods, these results are promising. First, the positive coefficient is satisfactory. It suggests that a high sentiment today is likely to predict the occurrence of a crisis within the next twelve months. This corroborates previous findings of a negative relation between investor sentiment and future index returns. Further, there could be further support in the behavioural literature looking at explanation of stock market bubbles, with the concepts of herding and irrational exuberance. Irrational exuberance suggests that investors can be temporarily overly optimistic about the future of the economy, creating unprecedented demand for securities, causing prices to rise, in turn confirming investors' beliefs of price rises and further increasing demand, leading to periods of artificially high prices. Herding, as proposed by Shiller (2001) suggests that investors can have financial incentives to behave in accordance with a group, which could increase this effect towards creating bubbles. In the stock market, investors are afraid of missing opportunities (by not following a group) or failing alone (if we choose our own path which proves to be less prosperous). Investors thus tend to follow the group in trading decisions, and could exacerbate trends in sentiment. This process, also exacerbated by the media coverage can in some ways create a self-fulfilling prophecy, as empirically documented by Griffin et al. (2003). Following this large increase in price driven by collectively encouraged optimism (bubble), stock prices can suddenly collapse, and result in a crisis. This would explain the positive relation between sentiment and the prediction of future crisis.

In terms of model performance presented in Table 45 through Table 50, the results of the error tests do not conclude the ability of sentiment alone to predict the probability of occurrence of stock market crises. The relatively large Type 1 error rates (around 75% in most cases at the 25% threshold), suggest that the model requires more explanatory variables to predict the actual crises. The very low Type 2 errors (close to >10% in most cases) suggest that although the model generates very few false signals, the combinations of the Type 1 and Type 2 error results suggest that the model simply appears to generate very few probabilities higher than 0.25 and 0.5, so that almost no crises are detected at all.

Table 42 Model 2A Results Summary Table

Model 2A	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
VIX	0.132	0.187	0.13	0.144	0.118	0.14	0.127	0.16	0.14	0.174	0.113	0.126
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Constant	-4.774	-6.048	-4.863	-5.188	-4.554	-5.046	-4.8	-5.569	-5.153	-5.961	-4.474	-4.762
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	-67.04	-145.23	-68.42	-126.24	-70.57	-123.3	-67.95	-122.75	-64.61	-123.49	-70.78	-129.27
Prob>chi2	.	.	.	.	.	.	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 43 Model 2B Results Summary Table

Model 2B	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
PCR	1.162	1.35	3.974	4.42	3.593	3.99	3.759	4.206	3.421	3.831	1.793	1.96
	-0.33	-0.29	(0.00)***	(0.00)***	(0.01)***	(0.00)***	(0.01)***	(0.00)***	(0.01)***	(0.01)***	-0.16	-0.14
Constant	-0.594	-0.441	-5.101	-5.482	-4.761	-5.098	-4.913	-5.294	-4.616	-4.965	-3.202	-3.342
	-0.54	-0.67	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	-73.32	-160.95	-67.86	-135.29	-68.65	-136.09	-68.29	-135.82	-68.96	-136.65	-71.53	-139.63
Prob>chi2	.	.	.	.	.	.	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 44 Model 2C Results Summary Table

Model 2C	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
BWSI	4.419 (0.00)***	2.179 (0.00)***	0.544 (0.10)*	0.546 (0.10)*	0.994 (0.00)***	0.997 (0.00)***	0.943 (0.00)***	0.946 (0.00)***	0.878 (0.01)***	0.882 (0.01)***	3.052 (0.00)***	2.89 (0.00)***
Constant	-2.551 (0.00)***	-2.072 (0.00)***	-1.876 (0.00)***	-1.877 (0.00)***	-2.055 (0.00)***	-2.056 (0.00)***	-2.033 (0.00)***	-2.033 (0.00)***	-2.005 (0.00)***	-2.005 (0.00)***	-2.804 (0.00)***	-2.796 (0.00)***
Log-Likelihood	-48.92	-142.48	-71.44	-108.2	-68.03	-111.07	-68.51	-112.13	-69.08	-113.84	-43.93	-112.3
Prob>chi2	.	.	.	.	.	.	.	.	.	.	.	.

* p<0.1; ** p<0.05; *** p<0.01

Table 45 Model 2A Dummy A Predictive Probabilities Errors

Dummy A	Model 2 A					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	71.43%	73.08%	80.77%	76.92%	69.23%	88.46%
0.5	78.57%	88.46%	92.31%	88.46%	80.77%	96.15%
Type II						
0.25	12.26%	9.35%	9.81%	9.81%	8.88%	9.81%
0.5	2.83%	3.74%	4.21%	3.74%	2.80%	4.21%

Table 46 Model 2B Dummy A Predictive Probabilities Errors

Dummy A	Model 2 B					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	100.00%	69.23%	73.08%	69.23%	73.08%	100.00%
0.5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Type II						
0.25	0.00%	7.94%	7.01%	6.54%	6.54%	0.93%
0.5	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 47 Model 2C Dummy A Predictive Probabilities Errors

Dummy A	Model 2 C					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	35.71%	96.15%	80.77%	80.77%	80.77%	50.00%
0.5	60.71%	100.00%	100.00%	100.00%	100.00%	50.00%
Type II						
0.25	15.09%	6.07%	6.07%	6.54%	5.61%	5.14%
0.5	5.66%	0.47%	0.93%	1.40%	0.47%	2.34%

Table 48 Model 2A Dummy B Predictive Probabilities Errors

Dummy B	Model 2 A					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	75.00%	73.08%	84.62%	76.92%	69.23%	96.15%
0.5	96.43%	100.00%	100.00%	100.00%	100.00%	100.00%
Type II						
0.25	4.23%	3.26%	3.26%	3.26%	3.72%	2.79%
0.5	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 49 Model 2B Dummy B Predictive Probabilities Errors

Dummy B	Model 2 B					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	100.00%	73.08%	80.77%	76.92%	84.62%	100.00%
0.5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Type II						
0.25	0.00%	5.12%	4.65%	4.65%	4.65%	0.00%
0.5	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 50 Model 2C Dummy B Predictive Probabilities Errors

Dummy B	Model 2 C					
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI
Error						
Type 1						
0.25	60.71%	96.15%	80.77%	80.77%	80.77%	50.00%
0.5	85.71%	100.00%	100.00%	100.00%	100.00%	50.00%
Type II						
0.25	0.94%	6.05%	6.05%	6.51%	6.05%	2.33%
0.5	0.00%	0.47%	1.40%	0.93%	0.47%	0.93%

3.5.3 Sentiment and Macro Variable Results

These findings are further investigated by observing the results of Model 3, which combines the explanatory variables of Models 1 and 2 (macroeconomic control variables with the investor sentiment variables). Table 51, Table 52, and Table 53 summarise the results of Model 3, which give an indication of the value added by including sentiment in predicting the probability of occurrence of stock market crises. In Model 3, the coefficients of the sentiment variables continue to displays a positive sign, and are statistically significant at the 1% level for the *VIX* in all markets. For the *PCR* proxy the coefficient displays a positive sign and is statistically significant at least at the 5% level for all markets except the DJIA and NIKKEI. For the *BWSI* proxy, the coefficient is statistically significant at the 1% level for all markets expect the DJIA (not found significant) and the NIKKEI (10% level).

Further, examination of the three sentiment proxies in relation to the macroeconomic variables reveals no notable changes with the significance of other variables' coefficients.

Table 51 Model 3A Results Summary Table

Model 3A	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
IPROD	0.72	0.46	1.101	1.101	0.058	0.05	0.012	0.016	0.083	0.082	0.101	0.091
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.24	-0.3	-0.64	-0.55	(0.04)**	(0.05)**	(0.00)***	(0.00)***
IFR	0.093	0.108	-1.488	-1.488	-0.834	-0.781	-1.055	-0.988	2.34	2.17	0.54	0.455
	-0.79	-0.76	(0.01)**	(0.01)**	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.00)***	(0.00)***	(0.03)**	(0.06)*
INT	-6.59	-5.763	-7.184	-7.184	0.398	0.282	-	-	-		1.893	1.765
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.34	-0.49	-	-	-		(0.07)*	(0.06)*
TS	7.4	6.738	6.849	6.849	1.069	0.925	-0.223	-0.36	1.95	1.952	0.177	-0.039
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.05)*	(0.08)*	-0.61	-0.42	(0.00)***	(0.00)***	-0.84	-0.96
GDP	0	0	0	0	0	0	0	0	0	0	0	0
	-0.39	-0.73	-0.17	-0.17	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.06)*	-0.14
VIX	0.196	0.206	0.243	0.243	0.263	0.243	0.163	0.185	0.182	0.204	0.138	0.096
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***
CAPE	-0.382	-0.242	-1.037	-1.037	-	-	-	-	-	-	-	-
	(0.00)***	(0.01)***	(0.00)***	(0.00)***	-	-	-	-	-	-	-	-
Constant	-59.917	-50.29	3.848	3.849	23.018	22.771	36.809	32.255	-142.62	-133.46	-76.469	-64.243
	(0.00)***	(0.00)***	-0.86	-0.86	-0.12	-0.12	(0.03)**	(0.06)*	(0.00)***	(0.00)***	(0.01)***	(0.02)**
Log-Likelihood	-37.52	-82.03	-28.58	-28.58	-50.71	-70.58	-49.27	-96.42	-45.09	-80.1	-55.01	-88.69

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 52 Model 3B Results Summary Table

Model 3B	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
IPROD	0.804	0.623	0.87	0.87	0.064	0.061	0.007	0.004	0.036	0.038	0.046	0.042
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.18	-0.2	-0.76	-0.88	-0.31	-0.31	-0.2	-0.24
IFR	0.211	-0.029	-0.792	-0.792	-0.109	-0.119	-0.545	-0.575	1.809	1.709	0.8	0.611
	-0.58	-0.94	-0.1	-0.1	-0.38	-0.35	-0.11	-0.1	(0.00)***	(0.00)***	(0.05)**	(0.09)*
INT	-7.706	-7.606	-7.28	-7.28	0.487	0.452	-	-	-	-	2.682	3.248
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.21	-0.23	-	-	-	-	-0.13	(0.06)*
TS	8.204	8.163	6.866	6.866	0.627	0.589	-0.48	-0.489	1.406	1.376	-0.428	-0.5
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.17	-0.19	-0.28	-0.28	(0.00)***	(0.00)***	-0.63	-0.58
GDP	0	0	0	0	0	0	0	0	0	0	0	0
	-0.31	-0.89	-0.48	-0.48	-0.36	-0.35	-0.15	-0.13	(0.00)***	(0.00)***	-0.26	-0.38
PCR	8.967	7.697	1.517	1.517	5.634	5.633	5.49	5.408	6.61	7.435	1.264	0.812
	(0.05)**	(0.08)*	-0.75	-0.75	(0.03)**	(0.03)**	(0.04)**	(0.04)**	(0.03)**	(0.02)**	-0.65	-0.77
CAPE	-0.592	-0.493	-0.628	-0.628	-	-	-	-	-	-		

	(0.00)***	(0.00)***	(0.01)***	(0.01)***	-	-	-	-	-	-		
Constant	-52.587	-36.943	-20.017	-20.018	-10.674	-9.412	21.397	22.732	-108.63	-103.11	-94.686	-73.416
	(0.01)***	(0.04)**	-0.38	-0.38	-0.33	-0.39	-0.18	-0.17	(0.00)***	(0.00)***	(0.04)**	(0.06)*
Log-Likelihood	-37.83	-77.69	-28.78	-28.78	-63.14	-101.07	-58.64	-119.47	-49.45	-88.71	-54.71	-93.83
Prob>chi2	.	.	.	.	.	.	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 53 Model 3C Results Summary Table

Model 3C	NASA	NASB	DJIAA	DJIAB	FTSEA	FTSEB	CACA	CACB	DAXA	DAXB	NIKA	NIKB
IPROD	0.465	0.383	1.041	1.041	0.049	0.052	0.009	0.01	0.041	0.041	Doesn't converge	0.07
	(0.10)*	(0.01)***	(0.00)***	(0.00)***	-0.34	-0.3	-0.71	-0.68	-0.32	-0.31		-0.72
IFR	-1.097	-0.57	-0.708	-0.708	-0.303	-0.295	-0.283	-0.267	2.607	2.135		11.939
	-0.2	-0.35	-0.32	-0.32	-0.21	-0.23	-0.45	-0.48	(0.00)***	(0.00)***		-0.21
INT	-7.5	-7.331	-7.268	-7.268	0.666	0.63	-	-	-	-		26.191
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.09)*	-0.11	-	-	-	-		-0.46

TS	6.031	6.538	5.908	5.908	1.249	1.216	-0.044	-0.055	2.114	1.811		-17.64
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.04)**	(0.05)*	-0.92	-0.9	(0.00)***	(0.00)***		-0.16
GDP	0	0	0	0	0	0	0	0	0	0		0
	-0.31	-0.34	-0.55	-0.55	(0.03)**	(0.03)**	-0.23	-0.24	(0.00)***	(0.00)***		-0.13
BWSI	12.934	6.32	1.59	1.59	2.214	2.209	1.36	1.373	2.442	2.236		37.415
	(0.01)***	(0.00)***	-0.19	-0.19	(0.00)***	(0.00)***	(0.01)***	(0.01)***	(0.00)***	(0.00)***		(0.08)*
CAPE	-0.401	-0.239	-0.416	-0.416	-	-	-	-	-	-		
	(0.02)**	(0.04)**	(0.09)*	(0.09)*	-	-	-	-	-	-		
Constant	15.002	-9.856	-51.884	-51.884	-5.196	-5.661	1.564	0.539	-157.75	-130.29		-1,410.51
	-0.71	-0.73	-0.15	-0.15	-0.72	-0.69	-0.93	-0.98	(0.00)***	(0.00)***		-0.19
Log-Likelihood	-16.27	-43.85	-21.94	-21.94	-55.4	-88.58	-56.1	-93.78	-38.94	-66.86	0	-31.96
Prob>chi2	.	.	.	.	.	.	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

In terms of model performance, the error tests point to the superior predicting ability of Model 3. In general all models' predictive capacity is improved with the inclusion of the sentiment variable, the model improves as measured in the percentage of error in crises predicting probability of occurrence of crisis across all indices. As one of the objectives of this essay is to determine which sentiment proxy performs the best, the probability of occurrence of crises prediction is explored by model in each subsection below.

3.5.3.1 Model 3 A – VIX Proxy

The number of missed signals (error 1) is lower than 33% on average in the six indices (at the 25% threshold for Dummy A), while the number of false signals (error 2) is lower than 12% (at the 25% threshold level Dummy A). An improvement in both error tests is strong evidence in favour of the value added of investor sentiment in models of stock market crises probability.

Table 54 Model 3A Dummy A Predictive Probabilities Errors

Dummy A Index	Model 3 A NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error Type 1							
0.25	14.29%	19.23%	50.00%	46.15%	34.62%	38.46%	33.79%
0.5	35.71%	30.77%	65.38%	61.54%	61.54%	61.54%	52.75%
Type II							
0.25	14.62%	11.21%	13.08%	13.55%	15.42%	7.01%	12.48%
0.5	8.49%	6.54%	7.94%	5.14%	10.28%	1.87%	6.71%

Table 55 Model 3A Dummy B Predictive Probabilities Errors

Dummy B	Model 3 A						
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	28.57%	19.23%	53.85%	46.15%	42.31%	46.15%	39.38%
0.5	42.86%	30.77%	69.23%	76.92%	73.08%	65.38%	59.71%
Type II							
0.25	5.63%	6.05%	4.65%	7.91%	5.12%	5.58%	5.82%
0.5	1.88%	1.40%	0.93%	2.33%	1.40%	0.93%	1.48%

3.5.3.2 Model 3 B- PCR Proxy

In general the results for the *PCR* proxy are slightly worse compared to the *VIX* proxy (as measured by the model's crisis occurrence probability prediction error terms). The number of missed signals (error 1) is lower than 38% on average in the six indices (at the 25% threshold for Dummy A), while the number of false signals (error 2) is lower than 12% (at the 25% threshold level Dummy A). One noteworthy difference is in the mean of Type I errors at the 50% threshold level, with the *PCR* proxy there is over a 10% increase.

Table 56 Model 3B Dummy A Predictive Probabilities Errors

Dummy A	Model 3 B						
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	21.43%	11.54%	61.54%	53.85%	26.92%	50.00%	37.55%
0.5	35.71%	30.77%	88.46%	84.62%	69.23%	53.85%	60.44%
Type II							
0.25	16.98%	7.01%	11.21%	10.28%	14.02%	7.48%	11.16%
0.5	10.85%	3.27%	0.93%	2.34%	8.41%	0.47%	4.38%

Table 57 Model 3B Dummy B Predictive Probabilities Errors

Dummy B Index	Model 3 B NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error Type 1							
0.25	32.14%	11.54%	61.54%	61.54%	38.46%	50.00%	42.54%
0.5	53.57%	30.77%	88.46%	84.62%	80.77%	61.54%	66.62%
Type II							
0.25	5.16%	3.26%	10.23%	9.30%	6.98%	6.51%	6.91%
0.5	1.41%	0.93%	0.93%	2.33%	1.40%	0.00%	1.16%

3.5.3.3 Model 3 C- BWSI Proxy

Model 3 C outperforms both of the prior proxies for sentiment as measured by errors. Dummy A slightly outperforms Dummy B in terms of Type 1 errors, while the reverse is the case in terms of Type 2 errors. The number of missed signals (error 1) is considerably lower at 16% on average in the six indices (at the 25% threshold for Dummy A), while the number of false signals (error 2) is lower than 10% (at the 25% threshold level Dummy A). One noteworthy difference is in the mean of Type 1 errors at the 50% threshold level, with the *PCR* proxy there is over a 10% increase.

Table 58 Model 3C Dummy A Predictive Probabilities Errors

Dummy A Index	Model 3 C NASDAQ	DJIA	FTSE	CAC	DAX	Mean
Error						
Type 1						
0.25	3.57%	11.54%	30.77%	38.46%	15.38%	19.95%
0.5	17.86%	23.08%	80.77%	100.00%	53.85%	55.11%
Type II						
0.25	14.15%	6.07%	9.35%	9.81%	12.15%	10.31%
0.5	10.85%	3.74%	2.34%	1.40%	7.94%	5.25%[56]

Table 59 Model 3C Dummy B Predictive Probabilities Errors

Dummy B Index	Model 3 C NASDAQ	DJIA	FTSE	CAC	DAX	Mean
Error						
Type 1						
0.25	14.29%	11.54%	30.77%	38.46%	23.08%	23.63%
0.5	28.57%	23.08%	80.77%	96.15%	57.69%	57.25%
Type II						
0.25	5.16%	4.19%	8.84%	9.77%	6.98%	6.99%
0.5	1.88%	1.86%	2.33%	1.40%	1.40%	1.77%

In conclusion, based on the findings of these results, the *BWSI* variable performs the best of the three for predicting the probability of a crisis in the following 12 months. These results add to the literature by confirming sentiment's incremental benefit in predicting the probability of occurrence of a crisis. They extend the literature through the

[56] The NIKKEI index data is not included as it did not converge consequently the mean would be positively affected consequently it is left out.

comparative analysis of which sentiment proxy best forecasts crises. The next section examines any cross-country comparative differences.

3.5.4 Cross-Country Comparisons

This section examines Model 3 as a cross country comparison, by comparing how the model[57] performs in predicting the probability of occurrence of a crisis in each country in light of the results of the previous section, which identifies *BWSI* proxy as the overall winner.

The model performs best in the US indices. It correctly predicts the occurrence of crises within 12 months 96% and 82% of the time at the 25% and 50% thresholds respectively (for the NASDAQ Dummy A). The incidence of false alarms (in the NASDAQ) occurs just 14% and 11% at the 25% and 50% respectively. As

[57] With the inclusion of both the *BWSI* proxy and the macroeconomic variables.

Table 58 confirms, similar results are found with the DJIA. These results are not surprising, although global markets are largely integrated and correlated, this sentiment proxy does originate in the United States. The second best model performance in errors in predicting stock market crises, is in Germany. The model accurately predicts the occurrence of crises within 12 months in the DAX 85% of the time (25% threshold Dummy A) and 47% of the time (50% threshold Dummy A). Further, the number of false alarms are low at 12% (25% threshold Dummy A) and 8% (50% threshold Dummy A). This observation presents an interesting venue for future research. For the FTSE and the CAC, the results of the error tests are similar to the DAX.

In examining the best macroeconomic predictors by region, I find reasonably similar results across all indices, with the exception of the Nikkei. Most of the macroeconomic indicators for Japan are not found significant when sentiment is included in the regression. While the US and EU markets experienced pronounced periods of crises and recover, the Nikkei experienced a relatively flat market. The macroeconomic variables seem to perform as indicators in similar ways between the NASDAQ and three European indices (FTSE, CAC, and DAX), this could be due to the similar nature of the underlying companies.

Overall, the results of Model 3 are in line with results from the vast literature published in the wake of the most recent financial crises on leading economic indicators in forecasting financial crises. While these recent advances aided in the variable selection process used in this essay, the results of the present model highlight the value added by including investor sentiment as a stock market crisis predictor, in line with previous evidence of the link between investor sentiment and stock returns.

3.6 Predicting Stock Market Recoveries

The essay has until now focused on the predictive power of sentiment in stock market crises. By construction, the dependent variables in the models were defined so as to detect significant falls in stock prices. Another useful consideration would be to determine whether the same variables could help predict a reversal in the downward trend. Following the occurrence of a crisis, it would be beneficial to determine whether the crisis has reached its trough and is expected to head towards recovery. In the context of this essay, it would be a notable contribution to determine whether investor sentiment is also an effective predictor of stock market crisis reversals. The final section of the essay builds upon the logit methodology from the literature, to examine the predictive power of the independent variables in recoveries.

3.6.1 Methods

3.6.1.1 Identifying Early Signs of Recoveries

The method previously used to detect stock market crises and create predictive dummy variables is modified to identify early signs of recoveries in the stock market. Additionally, this paper will contribute to the literature in identifying the potential incremental benefit of sentiment in predicting the probability of a recovery in the next 12 months. The same three sentiment proxies are used which further adds to the research identifying the most optimal proxy for sentiment in forecasting a recovery.

First, a new variable is created to determine when the index value is increasing significantly following the occurrence of a crisis. Mirroring the creation of the CMAX variable, CMIN is the ratio of current index value over the minimum value over a certain number of months T, as shown in Eq. (30):

$$CMIN_{i,t} = \frac{P_{i,t}}{min(P_{i,t-T} \ldots \ldots, P_{i,t})} \tag{30}$$

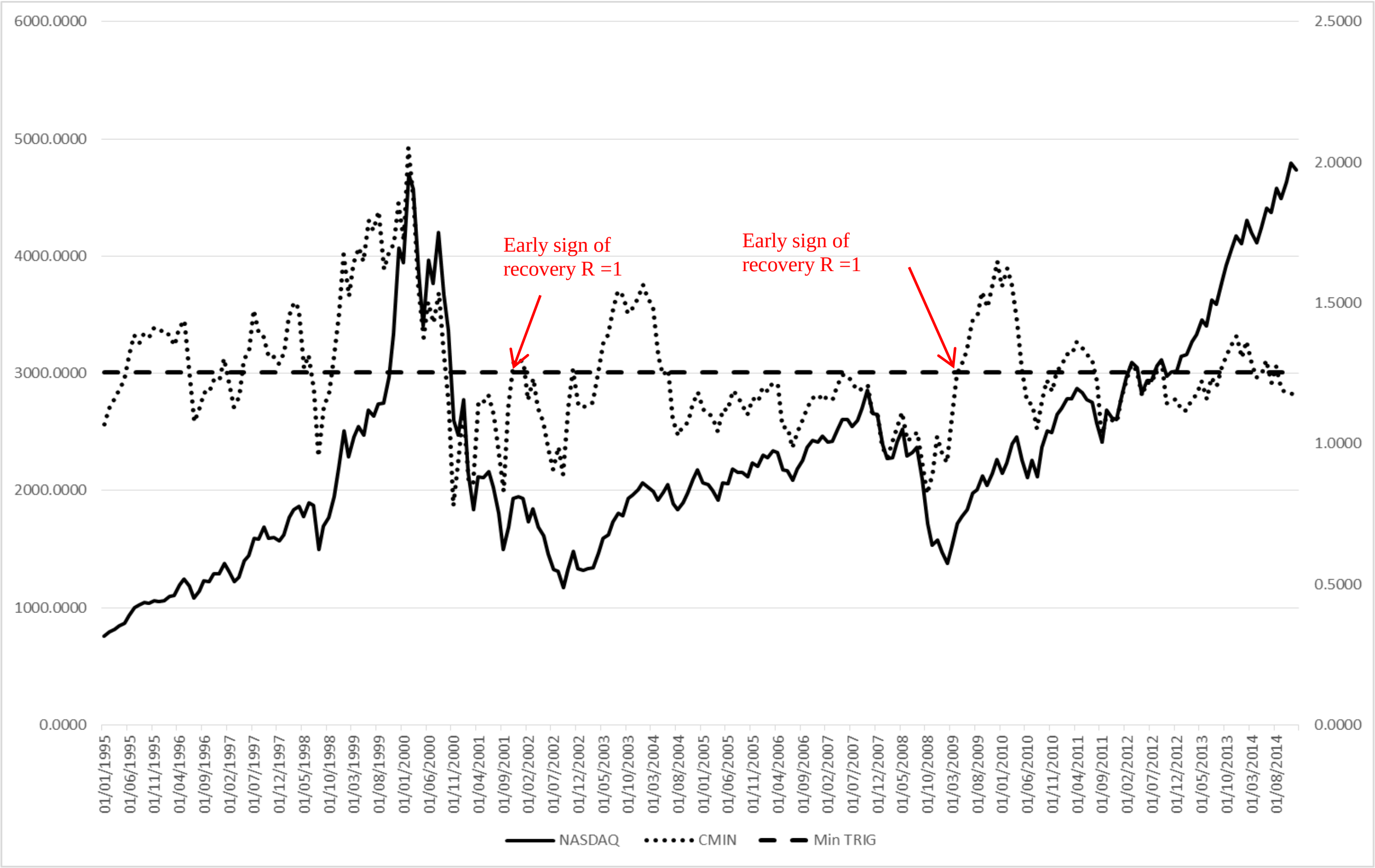
$P_{i,t}$ represents the monthly value of index i at time t, the denominator is the minimum index value over a rolling period T where T=24. CMIN is equal to 1 when the index value is at its lowest over the 2 year period. CMIN is higher than 1 in a period of rising prices. The CMIN effectively represents the increase in index value from the minimum value over a 1 year period. Following the same methodology as was previously used to detect crises, the CMIN variable is then used to identify early signs of recovery, by constructing a new dummy variable.

A new variable R is defined as 1, when the period can be considered to show signs of early recovery, when the CMIN is superior to the mean of the CMIN over the entire sample, and zero otherwise. In order to exclude double counting, a recovery is automatically eliminated if it is detected twice during the same 12 month period. Additionally, an additional condition must be met to ensure that early signs of recovery can only be detected following the detection of a crisis in the previous 12 months. This is summarised in Eq. (31).

$$R_{i,t} = \begin{cases} 1 & if\ CMIN_{i,t} > \overline{CMIN}\ \ and\ if\ C_{i,t-x} = 1\ with\ x \in [0, \ldots, 11] \\ 0 & otherwise \end{cases} \quad (31)$$

The newly created $R_{i,t}$ embodies the occurrence of early signs of recovery following a crisis. The concept is illustrated in Figure 17.

Figure 17 NASDAQ Recovery Model Illustration



3.6.1.2 Early Signs of Recoveries in the Data

Using the method described above, the CMIN and R variables are calculated for all indices over the sample period. For each of the crises, early signs of recovery are identified in a certain month. Table 60 records and compares the date where early signs of recovery were detected with the dates of the trough and the dates where the index actually recovered to pre-crisis levels.

The method does identify early signs of recovery after each crisis, often within the next few months following the trough. Over the 12 crises identified in the sample (two for each of the indices, corresponding to the Technology crash and subprime crises, see section 3.3 for more details), there are three occurrences where early signs of recovery are detected ahead of the trough: for the NASDAQ, the CAC, and the NIKKEI, the index appears to recover slightly before experiencing another period of falling prices, which lead to early detection of recovery signs (see Appendix 7 for graphical illustration). This also suggests that the recovery metric proposed in this essay could be further refined in future analyses.

Table 60 Early Sign of Recovery in the Data

Index	Beginning of Crash	Date of Trough	Early Sign of Recovery	Date of Recovery
NASDAQ	28/02/2001	30/09/2002	**30/11/2001**	28/11/2014
	27/02/2009	27/02/2009	29/05/2009	29/04/2011
DJIA	30/09/2002	28/02/2003	30/06/2003	31/12/2004
	31/10/2008	27/02/2009	29/05/2009	31/01/2013
FTSE	31/07/2002	31/01/2003	30/05/2003	31/10/2007
	31/10/2008	27/02/2009	29/05/2009	31/10/2013
CAC	31/07/2002	31/03/2003	**29/11/2002**	N/A
	28/11/2008	27/02/2009	29/05/2009	N/A
DAX	31/07/2002	31/03/2003	30/05/2003	30/04/2007
	30/01/2009	27/02/2009	30/04/2009	31/05/2013
Nikkei	28/09/2001	30/04/2003	**30/04/2002**	N/A
	31/10/2008	27/02/2009	30/04/2009	N/A

3.6.2 The Model

With the early signs of recoveries identified in the previous section, the logit model methodology can also be applied here. In order to determine the role of macroeconomic variables and both incremental benefit of investor sentiment and the optimal proxy in detecting early signs of recovery, a new dependent variable is created, mirroring the crisis variables. Dummy V is used as a dependent variable in the binomial logit model and is defined in Eq. (32):

$$V_{i,t} = \begin{cases} 1 & if\ x \in [1, \dots, 12]\ such\ that\ R_{i,t+x} = 1 \\ 0 & otherwise \end{cases} \tag{32}$$

The same three model specifications are used as previously done so, to test the role of macroeconomic variables and sentiment respectively. The specifications are re-iterated in Eq. (33) through (35)

Model 4 $\Pr(V_{i,t} = 1) = f\left(\alpha_0 + \sum_{k=1}^{n} \alpha_k M_{i,t}^{k}\right)$ (33)

Model 5 $\Pr(V_{i,t} = 1) = f(\alpha_0 + \beta\, SENT)$ (34)

Model 6 $\Pr(V_{i,t} = 1) = f\left(\alpha_0 + \sum_{k=1}^{n} \alpha_k M_{i,t}^{k} + \beta\, SENT\right)$ (35)

With a logistic function: $f(w) = \frac{e^w}{1+e^w}$

Similar to the previous models, the performance of the model is assessed through comparing the model predictions of early signs of recovery, with the actual occurrences documented above. Using the same 25% and 50% detection thresholds, the number of Type 1 and Type 2 errors are computed using the typology in Table 61.

Table 61 Recovery Metric Error Types

		Logit Model	
		Signal of early recovery generated	Signal was not Generated
Actual data	*R* predicts early signs of recovery	Correct signal	**No Signal (error 1)**
	R does not predict early signs of a recovery	**False signal (error 2)**	No signal and no crises

3.6.3 Recovery Results

The results of the logit model for the early signs of recovery are presented in this section. Table 62 through Table 72 summarise the results of Models 4, 5[58] and 6, showing the

[58] Models 5 and 6 have A, B, C versions, which represent the different proxies for sentiment.

influence of macroeconomic variables, investor sentiment variables, and both sets of variables on predicting early signs of stock market recoveries.

The results show quite strong positive predictability, acknowledging previous observations in the literature about the difficulties in modelling recoveries (Loungani, 2002)[59]. In comparison with the model predicting the probability of occurrence of crises, the results show a lower mean of both Type 1 and Type 2 errors, and show significance for the majority of the coefficients associated with the sentiment proxy variables. Overall, the signs of the coefficients present expected signs across all indices. This suggests that the macroeconomic variables highlighted in the literature as important components of early warning systems for crises, likewise show promising results for an early warning recovery model.

Table 62 presents the results of Model 4, which is comprised of only macroeconomic variables. The results show strong statistical significance for most variables, as well as expected signs associated with the coefficients. This is shown particularly with respect to *GDP* and Inflation, which show statistical significance at least at the 10% level for all in save the NIKKEI (*GDP*) and the DJIA and NIKKEI (Inflation). The macroeconomic variables on their own do an adequate job predicting the probability of a recovery as measured by the error terms presented in Table 63. At the 25% threshold, Type 1 errors are just under 30% and Type 2 errors are around 6%.

[59] Loungani (2002) argues that forecasting recoveries has been particularly successful when recessions have tended to last under one year, however, for longer recession, recovery forecasting proves a difficult task..

Table 62 Model 4 Results Summary Table

Model 4	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
IPROD	-0.342	-0.458	0.088	-0.019	-0.04	-0.117
	(0.06)*	(0.02)**	-0.27	-0.38	-0.24	(0.00)***
IFR	2.053	0.172	-3.519	-0.999	1.762	-0.454
	(0.04)**	-0.72	(0.00)***	(0.00)***	(0.00)***	-0.25
INT	-5.863	1.033	-3.187	-	-	-6.138
	(0.01)***	-0.41	(0.00)***	-	-	(0.00)***
TS	-1.184	-18.046	4.195	2.282	1.387	-0.607
	-0.68	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.47
GDP	0	0	0	0	0	0
	(0.02)**	(0.06)*	(0.00)***	(0.00)***	(0.00)***	-0.17
CAPE	0.41	0.726	-	-	-	-
	(0.06)*	(0.00)***	-	-	-	-
Constant	-15.548	87.86	186.67	39.695	-99.839	62.008
	-0.66	(0.00)***	(0.00)***	(0.00)***	(0.00)***	-0.14
Log-Likelihood	-16	-21.96	-23.59	-61.04	-55.58	-60.75
Prob>chi2	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 63 Model 4 Predictive Probabilities Errors

Dummy A Index	Model 4 NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	3.85%	7.69%	11.54%	42.31%	53.85%	50.00%	28.21%
0.5	11.54%	19.23%	15.38%	76.92%	73.08%	88.46%	47.44%
Type II							
0.25	2.80%	3.27%	2.80%	9.35%	4.67%	10.75%	5.61%
0.5	0.93%	1.87%	1.87%	1.40%	1.40%	2.34%	1.64%

Table 64 through Table 66 presents the results of Model 5, where only the sentiment variables (A. *VIX*, B. *PCR*, and C. *BWSI*) are used as a predictors of probability of occurrence of early signs of recovery. The model shows that the sentiment proxies differ greatly in significance. However, as expected the sign of the coefficient is negative across all indices, suggesting that periods of low sentiment are pre-cursors of early signs of market recoveries. Table 64 presents the results for the *VIX* variable, showing significance at the 1% level in the NASDAQ, and NIKKEI, and significance at the 5% level for the CAC.

Table **65** presents the results for the *PCR*, which shows no statistical significance associated with the coefficients. The strongest results are presented in Table 66 for the *BWSI* proxy. These results show significance at the 1% level for all indices with the exception of the NASDAQ. However, the performance of the model is tested when examining the error tests, a similar finding that was found in Model 3 to predict the probability of occurrence of crises using sentiment as the only variable. These results are shown in Table 67 through Table 69 below.

Table 64 Model 5A Results Summary Table

Model 5 A	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
VIX	-0.065	-0.006	-0.009	-0.046	-0.012	-0.082
	-(0.00)***	-0.81	-0.71	(0.04)**	-0.62	-(0.00)***
Constant	-3.567	-2.237	-2.306	-3.129	-2.364	-4.002
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	-78.16	-82.3	-82.26	-80.3	-82.21	-75.48
Prob>chi2	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 65 Model 5B Results Summary Table

Model 5 B	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
PCR	-0.62	-0.752	-0.66	-0.594	-0.239	-0.88
	-0.63	-0.56	-0.61	-0.64	-0.85	-0.5
Constant	-2.359	-2.469	-2.392	-2.337	-2.044	-2.575
	(0.03)**	(0.02)**	(0.03)**	(0.03)**	(0.06)*	(0.02)**
Log-Likelihood	-75.88	-75.82	-75.86	-75.89	-75.98	-75.76
Prob>chi2	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 66 Model 5C Results Summary Table

Model 5 C	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
BWSI	-0.43	-4.767	-5.566	-6.918	-6.002	-2.627
	-0.32	-(0.00)***	-(0.00)***	-(0.00)***	-(0.00)***	-(0.00)***
Constant	-1.804	-2.312	-2.477	-2.771	-2.57	-1.937
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	-75.58	-55.48	-51.52	-45.5	-49.48	-66.66
Prob>chi2	.	.	.	.	.	.

* p<0.1; ** p<0.05; *** p<0.01

Table 67 Model 5A Predictive Probabilities Errors

Dummy A	Model 5 A						
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	96.15%	100.00%	100.00%	100.00%	100.00%	92.31%	98.08%
0.5	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Type II							
0.25	4.67%	0.00%	0.00%	2.80%	0.00%	4.67%	2.02%
0.5	0.93%	0.00%	0.00%	0.00%	0.00%	0.93%	0.31%

Table 68 Model 5B Predictive Probabilities Errors

Dummy A	Model 5 B						
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	100.00%	100.00%	100.00%	1.00%	1.00%	1.00%	50.50%
0.5	100.00%	100.00%	100.00%	1.00%	1.00%	1.00%	50.50%
Type II							
0.25	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.5	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 69 Model 5C Predictive Probabilities Errors

Dummy A	Model 5 C

Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	1.00%	46.15%	42.31%	42.31%	38.46%	57.69%	37.99%
0.5	1.00%	76.92%	69.23%	53.85%	65.38%	96.15%	60.42%
Type II							
0.25	0.00%	7.01%	6.54%	7.01%	7.48%	5.61%	5.61%
0.5	0.00%	3.27%	2.80%	1.40%	2.34%	1.40%	1.87%

As can be seen above, sentiment alone is not a strong predictor of the probability of recoveries, perhaps save the *BWSI* proxy. The *BWSI* proxy again, as expected, presents the strongest results with an average Type 1 error of 38% (at the 25% threshold) and just under 6% for Type 2 errors at the 25% threshold. Next, the results of the sentiment variables in addition to the macroeconomic variables are shown.

Table 70 Model 6A Results Summary Table

Model 6 A	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
IPROD	-0.331	-0.394	-0.039	-0.017	-0.063	-0.074
	(0.07)*	(0.05)*	-0.78	-0.46	-0.1	-(0.03)**
IFR	1.927	-0.238	-6.007	-0.966	2.562	-1.026
	(0.08)*	-0.67	(0.00)***	(0.00)***	(0.00)***	(0.07)*
INT	-5.984	0.164	-9.326	-	-	-6.464
	(0.01)***	-0.9	(0.00)***	-	-	(0.01)***
TS	-0.729	-16.071	5.319	2.242	2.158	0.279
	-0.81	(0.00)***	(0.03)**	(0.00)***	(0.00)***	-0.78
GDP	0	0	0.001	0	0	0
	(0.05)**	0.23	(0.00)***	(0.00)***	(0.00)***	0.29
VIX	-0.029	-0.195	-0.562	-0.021	-0.15	-0.135
	-0.79	-0.12	(0.02)**	-0.47	-(0.00)***	-(0.00)***
CAPE	-0.344	-0.299	-	-	-	-
	-0.27	-0.38	-	-	-	-
Constant	-10.915	110.669	378.582	37.679	-141.12	111.287
	-0.78	(0.00)***	(0.00)***	(0.01)***	(0.00)***	(0.07)*
Log-Likelihood	-15.97	-20.51	-10.28	-60.79	-49.71	-53.4
Prob>chi2	.	.	.	.	.	.

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Table 71 Model 6A Results Summary Table

Model 6 B	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
IPROD	-0.521	-0.6	0.082	-0.017	-0.029	-0.113
	(0.03)**	(0.01)***	-0.32	-0.43	-0.43	(0.00)***
IFR	2.16	0.161	-3.611	-0.877	1.802	-0.607
	(0.05)**	-0.75	(0.00)***	(0.01)**	(0.00)***	-0.3
INT	-6.025	0.86	-2.218			-5.166
	(0.01)**	-0.51	(0.07)*			(0.07)*
TS	-1.735	-20.401	5.651	2.306	0.981	-1.422
	-0.59	(0.00)***	(0.00)***	(0.00)***	(0.01)**	-0.14
GDP	0	0	0	0	0	0
	(0.03)**	(0.06)*	(0.00)***	(0.01)**	(0.00)***	-0.78
PCR	-8.76	-8.791	-1.021	-2.336	-4.676	-2.514
	-0.22	-0.17	-0.86	-0.41	-0.13	-0.41
CAPE	0.541	0.887				
	(0.05)**	(0.00)***				
Constant	-6.665	103.387	186.219	34.502	-102.81	77.243
	-0.86	(0.00)***	(0.00)***	(0.03)**	(0.00)***	-0.23
Log-Likelihood	-15.22	-20.92	-22.81	-59.14	-49.95	-55.69
Prob>chi2	.	.	.	.	.	.

* p<0.1; ** p<0.05; *** p<0.01

Table 72 Model 6A Results Summary Table

Model 6 C	NASA	DJIAA	FTSEA	CACA	DAXA	NIKA
IPROD	-0.025	-0.855	0.114	0.002	-0.011	-0.116
	-0.89	(0.07)*	-0.19	-0.95	-0.79	(0.00)***
IFR	5.916	2.789	-3.088	0.333	1.192	-0.783
	(0.03)**	(0.08)*	(0.00)***	-0.62	(0.02)**	-0.18
INT	-10.374	8.853	-1.469	-	-	-3.687

	-0.17	-0.12	-0.2	-	-	(0.10)*
TS	3.556	-26.691	5.096	2.168	1.248	-1.659
	-0.65	(0.01)**	(0.00)***	(0.01)***	(0.05)*	-0.12
GDP	0	0	0	0	0	0
	(0.03)**	(0.02)**	(0.00)***	-0.66	(0.04)**	-1
BWSI	-1.368	-4.601	-4.443	-8.779	-6.141	-1.056
	-0.37	-0.31	(0.07)*	(0.00)***	(0.00)***	-0.12
CAPE	0.989	1.58				
	(0.02)**	(0.02)**				
Constant	-217.31	-4.646	141.784	-25.051	-74.471	71.501
	(0.05)**	-0.95	(0.00)***	-0.43	(0.01)***	-0.22
Log-Likelihood	-8.64	-13.62	-21.36	-29.37	-33.41	-50.05
Prob>chi2	.	.	.	.	.	.

* p<0.1; ** p<0.05; *** p<0.01

through Table 72 present the results of Model 6, the sentiment variable retains some significance in Models 6A and 6C. The *PCR* sentiment proxy continues to not show significance. As comparable to the results of the early warning of crises models, the addition of the sentiment variable seems to significantly improve the model performance in predicting the probability of early signs of recovery. Further examination of the results of Model 6 C show that with the inclusion of the macroeconomic variables, the sentiment proxy has lost significance in the US and Japanese indices.

Compared with the results in Section 5, these results are interesting in themselves, as they provide insight into comparable nature of predictive ability of sentiment in stock market movements. As shown in the literature, periods of high sentiment are often followed by periods of low returns.

The predictive ability of the model can be further tested through the error tests as shown in Table 73 through Table 75 below. At the 25% threshold for probabilities, the model seems to provide relatively strong results in predicting the probability of early signs of recovery. Model 6 C correctly predicts early signs of recovery during the next 12 months 93% and 74% of the time (at the 25% and 50% thresholds respectively). False alarms were quite low occurring less than 6% of the time (at both thresholds). The recovery model performs the best in the US and UK markets. The model performs quite poorly in the NIKKEI market, this is expected due to the lack of recovery during this time frame. These results present promising additions to the literature on recovery modelling.

Table 73 6 A Predictive Probabilities Errors

Dummy A Index	Model 6 A NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	3.85%	7.69%	3.85%	42.31%	46.15%	38.46%	23.72%
0.5	11.54%	15.38%	7.69%	73.08%	65.38%	73.08%	41.03%
Type II							
0.25	2.80%	3.27%	1.40%	8.88%	9.81%	6.54%	5.45%
0.5	0.93%	1.87%	0.93%	0.93%	0.93%	2.80%	1.40%

Table 74 6 B Predictive Probabilities Errors

Dummy A Index	Model 6 B NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	0.00%	7.69%	7.69%	42.31%	50.00%	38.46%	24.36%
0.5	11.54%	11.54%	15.38%	73.08%	57.69%	88.46%	42.95%
Type II							
0.25	2.34%	3.74%	3.27%	9.35%	6.54%	10.28%	5.92%
0.5	0.93%	1.40%	1.87%	0.93%	1.40%	3.27%	1.64%

Table 75 6 C Predictive Probabilities Errors

Dummy A	Model 6 C						
Index	NASDAQ	DJIA	FTSE	CAC	DAX	NIKKEI	Mean
Error							
Type 1							
0.25	0.00%	3.85%	3.85%	3.85%	11.54%	23.08%	7.69%
0.5	0.00%	3.85%	19.23%	23.08%	38.46%	69.23%	25.64%
Type II							
0.25	1.87%	3.27%	2.80%	5.14%	7.48%	10.75%	5.22%
0.5	0.47%	0.47%	2.34%	2.34%	3.74%	3.27%	2.10%

3.7 Conclusion

Numerous studies have empirically shown the role of investor sentiment in stock market returns. Findings in the literature confirm that markets are subject to waves of optimistic and pessimistic sentiment. Theoretically, these findings are supported through questioning of the theories of efficient markets by recognising the limitations of arbitrage, and the psychological concepts which might lead investors to not behave in fully rational ways. A few studies have anecdotally mentioned sentiment as a possible cause in market crises (De Long & Shleifer, 1991; Shiller, 2000), yet there has been remarkably little research explicitly examining this link.

Building on this recent literature, this essay aims to further examine this relation by comparing the predictive value of investor sentiment and importantly, the optimal proxy in stock market crises and early signs of recovery across indices and regions. Six indices are examined in this essay across five countries (two American, two European and one Japanese indices). The methodology used closely follows the literature of early warning systems, which are models attempting to use a number of independent variables to detect warning signs of future crises. Following a methodology close to Bussiere and Fratzscher (2006), a CMAX indicator is created to detect stock market crises. Through this indicator, I detect two crises in each of the six indices during the sample period 1995-2014, which correspond to the technology bubble at the end of the late 1990s and the sub-prime crisis occurring around 2007-2008.

Through these newly detected crises, it is possible to use logit model estimation on a newly created crisis dependent variable, which takes the value 1 during the 12 months preceding the occurrence of a crisis. A multinomial logit model is used to differentiate between periods immediately preceding the crisis, and periods following the crisis. The logit model aims to determine the probability of occurrence of a crisis during the next twelve months using a set of independent predictors. First, I use well established macroeconomic indicators, which have been previously linked to stock market crisis detection. Second, three investor sentiment proxies are tested individually as an explanatory variables (as proxied by *VIX*, *PCR*, and *BWSI*). Three different models are run using: the macroeconomic variables, investor sentiment variables, and then one combining both as predictors of occurrence of crises. Through this method, it is possible to test the incremental benefit of adding consumer sentiment to the EWS model as well as identifying the optimal proxy for sentiment. The performance of the model is then tested by comparing the results of the model predicting a crisis or not (from a pre-determined threshold of probability) and the actual occurrence of crisis. In the EWS methodology, the number of missed crises and false signals give a good indication of the ability of model to predict crises.

The results confirm the value of sentiment as a predictor for the occurrence of stock market crises, and identify both *VIX* and *BWSI* proxies as performing the best. Overall, after controlling for macro-economic predictors, the *BWSI* performs the best and, is found to be statistically significant (at the 1% level) and improves the model in terms of errors in predictive probability.

In terms of country comparison, the model performs best in the NASDAQ index. It correctly predicts the occurrence of crises within 12 months 96% and 71% of the time at the 25% and 50% probability thresholds, respectively. The incidence of false alarms occurs just 5% and 2% at the 25% and 50%, respectively. The second best model performance occurs for the DJIA, the other US index examined in this essay. The results show expected results from the literature in terms of expected sign and significance for the macroeconomic variables,

Further, the essay also expands this research by relating sentiment and early signs of recovery to investigate whether this link holds in periods of recovery as well. Using a similar methodology, early signs of recovery periods are identified in the dataset, and the same three multinomial logit models are run to examine the predictive power of investor sentiment in early signs of recovery as well as the optimal proxy for sentiment. The results are similar to those found in crises probability prediction. The *BWSI* variable performs the best and is found to remain statistically significant at least at the 5% level despite the addition of the macroeconomic variables[60]. In terms of model performance, results are also extremely strong and relatively similar to the crises results. The recovery model performs the best in the US and French markets. The model performs the poorest in the NIKKEI market, although it is worth noting that the inclusion of the sentiment variable improves the performance of the model in the NIKKEI.

There are many areas for further research which derive from this essay. First, the lack of consensus in the academic literature surrounding the use of sentiment proxies to capture investor sentiment suggests that this research could be strengthened by further testing the impact of various proxies for sentiment as predictors of stock market crises. However, this might be limited by issues of data availability. Second, extensions of this work could also explore improvements in the predictive model itself: more predictors could be included to reflect advances in the literature on currency and other financial market crises.

[60] *BWSI* remains statistically significant in 3 of the 6 indices FTSE, CAC, and DAX.

The relatively strong results of the recovery model also suggest that methodology improvements and future research could be quite fruitful, particularly given the practical benefits of detecting reversals in stock price movements after crisis events. Further research could focus on designing new indicators to give early warning of stock market recoveries. Further, there is a need for extending research into the most valuable predictors of stock market recoveries, as the literature has so far preferred to focus on predicting stock market crises. Finally, there would be scope to further examine the various theoretical contributions around waves of optimism, limits of arbitrage, and positive feedback in the context of stock market bubbles, crises, and recoveries.

References

A. R., & Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. Review of Financial studies, 1(1), 3-40.

Aggarwal, R., & Rivoli, P. (1989). Seasonal and Day-of-the-Week Effects in Four Emerging Stock Markets. Financial review, 24(4), 541-550.

Agrawal, A., & Tandon, K. (1994). Anomalies or illusions? Evidence from stock markets in eighteen countries. Journal of international Money and Finance, 13(1), 83-106.

Alessi, L., & Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. European Journal of Political Economy, 27(3), 520-533.

Al-Jafari, M. K. (2012). An Empirical Investigation of the Day-of-the-Week Effect on Stock Returns and Volatility: Evidence from Muscat Securities Market. International Journal of Economics and Finance, 4(7), p141.

Almazan, A., Brown, K. C., Carlson, M., & Chapman, D. A. (2004). Why constrain your mutual fund manager?. Journal of Financial Economics, 73(2), 289-321.

Alsubaie, A., & Najand, M. (2009). Trading volume, time-varying conditional volatility, and asymmetric volatility spillover in the Saudi stock market. Journal of Multinational Financial Management, 19(2), 139-159.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. The Journal of Finance, 51(1), 169-204.

Andreassen, P. B. (1990). Judgmental extrapolation and market overreaction: On the use and disuse of news. Journal of Behavioral Decision Making, 3(3), 153-174.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. The Journal of Finance, 61(1), 259-299.

Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., & Vašíček, B. (2013). Leading indicators of crisis incidence: Evidence from developed countries. Journal of International Money and Finance, 35, 1-19.
Babecký, J., Havránek, T., Matějů, J., Rusnák, M., Šmídková, K., & Vašíček, B. (2013). Leading indicators of crisis incidence: Evidence from developed countries. Journal of International Money and Finance, 35, 1-19.
Baker, M. and J. Wurgler, (2007). Investor sentiment in the stock market, Journal of Economic Perspectives 21, 129–151.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. The Journal of Finance, 61(4), 1645-1680.

Baker, M., Wurgler, J., & Yu, Y. (2009). Global and Local Investor Sentiment. Working paper,

Baklaci, H. F., Olgun, O., & Can, E. (2011). Noise traders: a new approach to understand the phantom of stock markets. Applied Economics Letters, 18(11), 1035-1041.

Ball, C. A., Torous, W. N., & Tschoegl, A. E. (1982). Gold and the "weekend effect". Journal of Futures Markets, 2(2), 175-182.

Bandopadhyaya, A., & Jones, A. L. (2006). Measuring investor sentiment in equity markets. Journal of Asset Management, 7(3), 208-215.

Bandopadhyaya, A., & Jones, A. L. (2008). Measures Of Investor Sentiment: A Comparative Analysis Put-Call Ratio Vs. Volatility Index. Journal of Business & Economics Research (JBER), 6(8).

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. Handbook of the Economics of Finance, 1, 1053-1128.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment.Journal of financial economics, 49(3), 307-343.New York University, pp. 1--39.

Barone, E. (1990). The Italian stock market: efficiency and calendar anomalies. Journal of Banking & Finance, 14(2), 483-510.

Barrell, R., Davis, E. P., Karim, D., & Liadze, I. (2010). Bank regulation, property prices and early warning systems for banking crises in OECD countries. Journal of Banking & Finance, 34(9), 2255-2264.

Bathia, D., & Bredin, D. (2013). An examination of investor sentiment effect on G7 stock market returns. The European Journal of Finance, 19(9), 909-937.

Baur, M. N., Quintero, S., & Stevens, E. (1996). The 1986–88 stock market: Investor sentiment or fundamentals?. Managerial and Decision Economics,17(3), 319-329.

Berry, T. D., & Howe, K. M. (1994). Public information arrival. The Journal of Finance, 49(4), 1331-1346.

Berument, H., & Kiymaz, H. (2001). The day of the week effect on stock market volatility. Journal of economics and finance, 25(2), 181-193.

Berument, H., & Kiymaz, H. (2001). The day of the week effect on stock market volatility. Journal of economics and finance, 25(2), 181-193.

Berument, H., Coskun, M. N., & Sahin, A. (2007). Day of the week effect on foreign exchange market volatility: Evidence from Turkey. Research in International Business and Finance, 21(1), 87-97.

Bhattacharya, K., Sarkar, N., & Mukhopadhyay, D. (2003). Stability of the day of the week effect in return and in volatility at the Indian capital market: a GARCH approach with proper mean specification. Applied Financial Economics, 13(8), 553-563.

Black, F. (1986). Noise. The journal of finance, 41(3), 529-543.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of econometrics, 31(3), 307-327.

Bollerslev, T., & Domowitz, I. (1993). Trading patterns and prices in the interbank foreign exchange market. The Journal of Finance, 48(4), 1421-1443.

Bollerslev, T., & Jubinski, D. (1999). Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. Journal of Business & Economic Statistics, 17(1), 9-21.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. Journal of econometrics, 52(1), 5-59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. Handbook of econometrics, 4, 2959-3038.

Bomfim, A. N. (2003). Pre-announcement effects, news effects, and volatility: Monetary policy and the stock market. Journal of Banking & Finance, 27(1), 133-151.

Borio, C. and P. Lowe, (2002). Asset prices, financial and monetary stability: Exploring the nexus. Working paper 114, Bank for International Settlement.

Borio, C. E., & Drehmann, M. (2009). Assessing the risk of banking crises–revisited. BIS Quarterly Review, March.

Borio, C., & Zhu, H. (2012). Capital regulation, risk-taking and monetary policy: a missing link in the transmission mechanism?. Journal of Financial Stability,8(4), 236-251.

Boubaker, A., & Makram, B. (2011). The empirical relationship between stock returns volatility and trading volume: evidence on the Tunis stock market. International Journal of Management Science and Engineering Management,6(5), 374-381.

Boucher, C., (2004). Identification et comparaison des crises boursières, Conseil d'Analyse Economique 50, 375–396.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. Journal of Empirical Finance, *11*(1), 1-27.

Brown, G. W., & Cliff, M. T. (2005). Investor Sentiment and Asset Valuation*.The Journal of Business, 78(2), 405-440.

Brown, P., Chua, A., & Mitchell, J. (2002). The influence of cultural factors on price clustering: Evidence from Asia–Pacific stock markets. Pacific-Basin Finance Journal, 10(3), 307-332.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. Review of Financial studies, 22(6), 2201-2238.

Bussiere, M. and M. Fratzscher, (2006). Towards a new early warning system of financial crises, Journal of International Money & Finance 25, 953–973.

Bussiere, M., Fratzscher, M., May 2002. Towards a new early warning system of financial crises. ECB Working Paper No. 145.

Caggiano, G., Calice, P., & Leonida, L. (2014). Early warning systems and systemic banking crises in low income countries: A multinomial logit approach. Journal of Banking & Finance, 47, 258-269.

Campbell, J. Y., & Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. The Journal of Portfolio Management, 24(2), 11-26.

Campbell, J. Y., & Shiller, R. J. (2001). Valuation ratios and the long-run stock market outlook: an update (No. w8221). National bureau of economic research.

Candelon, B., Dumitrescu, E.I., Hurlin, C., (2009). How to evaluate an Early Warning

Chang, E. C., Pinegar, J. M., & Ravichandran, R. (1993). International evidence on the robustness of the day-of-the-week effect. Journal of Financial and Quantitative Analysis, 28(04), 497-513.

Chang, S. C., Chen, S. S., Chou, R. K., & Lin, Y. H. (2012). Local sports sentiment and returns of locally headquartered stocks: A firm-level analysis.Journal of Empirical Finance, 19(3), 309-318.

Chang, S. C., Chen, S. S., Chou, R. K., & Lin, Y. H. (2012). Local sports sentiment and returns of locally headquartered stocks: A firm-level analysis.Journal of Empirical Finance, 19(3), 309-318.

Chen, G. M., Firth, M., & Rui, O. M. (2001). The dynamic relation between stock returns, trading volume, and volatility. Financial Review, 36(3), 153-174.Chicago.

Chen, J., Hong, H., & Stein, J. C. (2002). Breadth of ownership and stock returns. Journal of financial Economics, 66(2), 171-205.

Chi, L., Zhuang, X., & Song, D. (2012). Investor sentiment in the Chinese stock market: an empirical analysis. Applied Economics Letters, 19(4), 345-348.

Choudhry, T. (2000). Day of the week effect in emerging Asian stock markets: evidence from the GARCH model. Applied Financial Economics, 10(3), 235-242.

Chui, A. C.W., Titman, S. And Wei, K.C. J. (2010), Individualism and Momentum around the World. The Journal of Finance, 65: 361–392.

Chung, S. L., Hung, C. H., & Yeh, C. Y. (2012). When does investor sentiment predict stock returns?. Journal of Empirical Finance, 19(2), 217-240.

Ciccone, S. J. (2011). Investor optimism, false hopes and the January effect.Journal of Behavioral Finance, 12(3), 158-168.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. Econometrica: journal of the Econometric Society, 135-155.

Coats, W. L. (1981). The weekend eurodollar game. The Journal of Finance, 36(3), 649-659.

Comelli, F. (2014). Comparing Parametric and Non-parametric Early Warning Systems for Currency Crises in Emerging Market Economies. Review of International Economics, 22(4), 700-721.

Condoyanni, L., O'HANLON, J., & Ward, C. W. (1987). Day of the week effects on stock returns: international evidence. Journal of Business Finance & Accounting, 14(2), 159-174.

Connolly, R. A. (1989). An examination of the robustness of the weekend effect. Journal of Financial and quantitative Analysis, 24(02), 133-169.

Connolly, R. A. (1991). A posterior odds analysis of the weekend effect. Journal of Econometrics, 49(1), 51-104.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Copeland, C. P. (1976). A Model of Asset Trading Under The Assumption of Sequential Information Arrival.The Journal of Finance, 31(4), 1149-1168.

Cornell, B. (1985). The weekly pattern in stock returns: Cash versus futures: A note. Journal of Finance, 583-588.

Corredor, P., Ferrer, E., & Santamaria, R. (2013). Investor sentiment effect in stock markets: Stock characteristics or country-specific factors?. International Review of Economics & Finance, 27, 572-591.

Corredor, P., Ferrer, E., & Santamaria, R. (2013). Investor sentiment effect in stock markets: Stock characteristics or country-specific factors?. International Review of Economics & Finance, 27, 572-591.

Coudert, V. and M. Gex, (2008). Does risk aversion drive financial crises? Testing the predictive power of empirical indicators, Journal of Empirical Finance 15(2), 167–184.

Cousin, J. G., & De Launois, T. (2005). Investigation of the impact of the financial communication intensity on the conditional volatility of stock returns.Available at SSRN 648941.

Cousin, J. G., & Launois, T. (2007). News Intensity and Conditional Volatility on the US stock market. AFFI. papier C,72.

Cumperayot, P., & Kouwenberg, R. (2013). Early warning systems for currency crises: A multivariate extreme value approach. Journal of International Money and Finance, 36, 151-171.

Cutler, D. M., Poterba, J. M., & Summers, L. H. (1991). Speculative dynamics.The Review of Economic Studies, 58(3), 529-546.Ball (1978),

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. the Journal of Finance, 53(6), 1839-1885.

Davidson, (2014). Volume, Volatility, and Momentum in Financial Markets. International Research Journal of Applied Finance. 5 (2)

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. Journal of political Economy, 703-738.

DeBondt, W. F., & Thaler, R. (1985). Does the stock market overreact?. The Journal of finance, 40(3), 793-805.

DeBondt, W. F., & Thaler, R. H. (1987). Further evidence on investor overreaction and stock market seasonality. The Journal of Finance, 42(3), 557-581.

Diefenbach, R. E. (1972). How good is institutional brokerage research?. Financial Analysts Journal, 28(1), 54-60.

Dieobold, F. X. (1986). Modeling the persistence of conditional variances: A comment. Econometric Reviews, 5(1), 51-56.

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. The Journal of Finance, 57(5), 2113-2141.

Dimson, E., & Fraletti, P. (1986). Brokers' recommendations: the value of a telephone tip. The Economic Journal, 139-159.

Dong, H., & Guo, X. (2014). The Interaction between Market Sentiments in the US Financial Market and Global Equity Market. International Business Research, 8(1), p1.

Dougal, C., Engelberg, J., Garcia, D., & Parsons, C. A. (2012). Journalists and the stock market. Review of Financial Studies, hhr133.

Duffie, D., Garleanu, N., & Pedersen, L. H. (2002). Securities lending, shorting, and pricing. Journal of Financial Economics, 66(2), 307-339.

Dyl, E. A., & Holland, C. W. (1990). Why a weekend effect? Comment. The Journal of Portfolio Management, 16(4), 88-89.

Dyl, E. A., & Maberly, E. D. (1986). The weekly pattern in stock index futures: A further note. The Journal of Finance, 41(5), 1149-1152.

Ederington, L. H., & Lee, J. H. (1993). How markets process information: News releases and volatility. The Journal of Finance, 48(4), 1161-1191.ee (1993)

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. The Journal of Finance, 66(1), 67-97.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica: Journal of the Econometric Society, 987-1007.

Epps, T. W., & Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. Econometrica: Journal of the Econometric Society, 305-321.

Fama, E. F. (1965). The behavior of stock-market prices. Journal of business, 34-105.

Fama, E., (1965). The Behavior of Stock-Market Prices, The Journal of Business, 38 (1), 34-106.

Fang, L., & Peress, J. (2009). Media Coverage and the Cross-section of Stock Returns. The Journal of Finance, 64(5), 2023-2052.

Fisher, K. L., & Statman, M. (2003). Consumer confidence and stock returns.The Journal of Portfolio Management, 30(1), 115-127.

Fisher, K.L. and M. Statman, (2006). Market timing in regressions and reality, The Journal of Financial Research 29(3), 293–304.

Flannery, M. J., & Protopapadakis, A. A. (1988). From t-bills to common stocks: investigating the generality of intra-week return seasonality. The journal of finance, 43(2), 431-450.

Fleming, J., & Kirby, C. (2011). Long memory in volatility and trading volume.Journal of Banking & Finance, 35(7), 1714-1726.

Fleming, J., Kirby, C., & Ostdiek, B. (2006). Stochastic Volatility, Trading Volume, and the Daily Flow of Information*. The Journal of Business, 79(3), 1551-1590.

Foster, A. J. (1995). Volume-volatility relationships for crude oil futures markets. Journal of Futures Markets, 15(8), 929-951.

Foster, F. D., & VISWANATHAN, S. (1993). Variations in trading volume, return volatility, and trading costs; Evidence on recent price formation models.The Journal of Finance, 48(1), 187-211.

Foster, F. D., & Viswanathan, S. (1995). Can speculative trading explain the volume–volatility relation?. Journal of Business & Economic Statistics, 13(4), 379-396.

Frankel, J., & Saravelos, G. (2012). Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis. Journal of International Economics, 87(2), 216-231.

Fratzscher, M. (2002). Financial market integration in Europe: on the effects of EMU on stock markets. International Journal of Finance & Economics, 7(3), 165-193.

French, K. R., & Roll, R. (1986). Stock return variances: The arrival of information and the reaction of traders. Journal of financial economics, 17(1), 5-26.

Gibbons, M. R., & Hess, P. (1981). Day of the week effects and asset returns. Journal of business, 579-596.

Gilles, C., & LeRoy, S. L. (1991). Econometric aspects of the variance-bounds tests: A survey. Review of Financial Studies, 4(4), 753-791.Merton (1987

Goodfriend, M., (2001). Financial stability, deflation and monetary policy. Working paper 01-01, Federal Reserve Bank of Richmond.

Gregoriou, A., Kontonikas, A., & Tsitsianis, N. (2004). Does the day of the week effect exist once transaction costs have been accounted for? Evidence from the UK. Applied Financial Economics, 14(3), 215-220.

Griffin, John M., Jeffrey H. Harris and Selim Topaloglu (2003). The dynamics of institutional and individual trading. Journal of Finance, 58, p. 2285-2320.

Groth, J. C., Lewellen, W. G., Schlarbaum, G. G., & Lease, R. C. (1979). An analysis of brokerage house securities recommendations. Financial Analysts Journal, 35(1), 32-40.

Harris, L. (1986). A transaction data study of weekly and intradaily patterns in stock returns. Journal of financial economics, 16(1), 99-117.

Harris, L. (1987). Transaction data tests of the mixture of distributions hypothesis. Journal of Financial and Quantitative Analysis, 22(02), 127-141.

Harris, M., & Raviv, A. (1993). Differences of opinion make a horse race.Review of Financial studies, 6(3), 473-506.

Harvey, C. R., & Huang, R. D. (1991). Volatility in the foreign currency futures market. Review of Financial Studies, 4(3), 543-569.

Herbert, J. H. (1995). Trading volume, maturity and natural gas futures price volatility. Energy Economics, 17(4), 293-299.

Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. The Journal of Finance, 54(6), 2143-2184.

Hong, H., & Stein, J. C. (2007). Disagreement and the Stock Market (Digest Summary). Journal of Economic perspectives, 21(2), 109-128.

Hong, H., Kubik, J. D., & Stein, J. C. (2004). Social interaction and stock-market participation. The journal of finance, 59(1), 137-163.

Hong, H., Scheinkman, J., & Xiong, W. (2006). Asset float and speculative bubbles. The journal of finance, 61(3), 1073-1117.

Houge, T., Loughran, T., Yan, X. S., & Suchanek, G. (2001). Divergence of opinion, uncertainty, and the quality of initial public offerings. University of Notre Dame Working Paper.

Hui, T. K. (2005). Day-of-the-week effects in US and Asia–Pacific stock markets during the Asian financial crisis: a non-parametric approach. Omega, 33(3), 277-282.

Jaffe, J., & Westerfield, R. (1985). The Week-End Effect in Common Stock Returns: The International Evidence. The journal of finance, 40(2), 433-454.

Jaffe, J., & Westerfield, R. (1989). Is there a monthly effect in stock market returns?: Evidence from foreign countries. Journal of Banking & Finance, 13(2), 237-244.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. The Journal of Finance, 45(3), 881-898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance, 48(1), 65-91.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance, 48(1), 65-91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. The Journal of Finance, 56(2), 699-720.

Jennings, R. H., & Barry, C. B. (1983). Information dissemination and portfolio choice. Journal of Financial and Quantitative Analysis, 18(01), 1-19.

Jennings, R. H., Starks, L. T., & Fellingham, J. C. (1981). An equilibrium model of asset trading with sequential information arrival. The Journal of Finance,36(1), 143-161.

Jondeau, E., Poon, S. H., & Rockinger, M. (2007). Financial modeling under non-Gaussian distributions. Springer Science & Business Media.

Jondeau, E., Poon, S. H., & Rockinger, M. (2007). Financial modeling under non-Gaussian distributions. Springer Science & Business Media.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, 263-291.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. American psychologist, 39(4), 341.

Kalev, P. S., Liu, W. M., Pham, P. K., & Jarnecic, E. (2004). Public information arrival and volatility of intraday stock returns. Journal of Banking & Finance, 28(6), 1441-1467.

Kaminsky, G., S. Lizondo, and C.M. Reinhart, (1998). Leading indicators of currency crises, IMF Staff Papers 45, 1–48.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. Journal of Financial and quantitative Analysis, 22(01), 109-126.

Keim, D. B., & Stambaugh, R. F. (1984). A further investigation of the weekend effect in stock returns. The journal of finance, 39(3), 819-835.

Keynes, J. M. (1973). 1936. The General Theory of Employment, Interest and Money.

Kiymaz, H., & Berument, H. (2003). The day of the week effect on stock market volatility and volume: International evidence. Review of Financial Economics, 12(4), 363-380.

Kolari, J., Caputo, M., & Wagner, D. (1996). Trait recognition: An alternative approach to early warning systems in commercial banking. Journal of Business Finance & Accounting, 23(9-10), 1415-1434.

Kuhn, T. S. (1970). Logic of discovery or psychology of research. Criticism and the Growth of Knowledge, 1-23.

Kumar, M. S., & Persaud, A. (2002). Pure contagion and investors' shifting risk appetite: analytical issues and empirical evidence. International Finance, 5(3), 401-436.

Lakonishok, J., & Maberly, E. (1990). The weekend effect: Trading patterns of individual and institutional investors. Journal of Finance, 231-243.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. Journal of Business & Economic Statistics, 8(2), 225-234.

Lashgari, M. (2000). The role of TED spread and confidence index in explaining the behavior of stock prices. American Business Review, 18(2), 9.

Lee, C., Ready, M. J., & Seguin, P. J. (1994). Volume, volatility, and New York stock exchange trading halts. The Journal of Finance, 49(1), 183-214.

Lee, C.M.C., A. Shleifer, and R.H. Thaler, (1991). Investor sentiment and the closed-end fund puzzle, Journal of Finance 46, 75–109.

Lee, W. Y., Jiang, C. X., & Indro, D. C. (2002). Stock market volatility, excess returns, and the role of investor sentiment. Journal of Banking & Finance,26(12), 2277-2299.

Lehmann, B., (1990). Fads, Martingales, and Market Efficiency. The Quarterly Journal of Economics. 105, 1-28.

Lemmon, M., & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. Review of Financial Studies, 19(4), 1499-1529.

LeRoy, S. F., & Porter, R. D. (1981). The present-value relation: Tests based on implied variance bounds. Econometrica: Journal of the Econometric Society, 555-574.

Levy, R. A. (1967). Random walks: Reality or myth. Financial Analysts Journal,23(6), 69-77.

Li, W. K., Ling, S., & McAleer, M. (2002). Recent theoretical results for time series models with GARCH errors. Journal of Economic Surveys, 16(3), 245-269.

Liesenfeld, R. (2001). A generalized bivariate mixture model for stock price volatility and trading volume. Journal of Econometrics, 104(1), 141-178.

Loungani, P. (2002). There Will Be Growth in the Spring: How Credible Are Forecasts of Recovery?. World Economics, 3(1).

Ma, C. K. (1986). A further investigation of the day-of-the-week effect in the gold market. Journal of Futures Markets, 6(3), 409-419.

Majand, M., & Yung, K. (1991). A GARCH examination of the relationship between volume and price variability in futures markets. Journal of Futures Markets, 11(5), 613-621.

Majand, M., & Yung, K. (1991). A GARCH examination of the relationship between volume and price variability in futures markets. Journal of Futures Markets, 11(5), 613-621.

Mandelbrot, B., (1963). The variation of certain speculative prices. Journal of Business, 36 pp. 394-419

Marsh, E. J., & Tversky, B. (2004). Spinning the stories of our lives. Applied Cognitive Psychology, 18(5), 491-503.

Martikainen, T., & Puttonen, V. (1996). Finnish Day-of-the-Week Effects. Journal of Business Finance & Accounting, 23(7), 1019-1032.

McFarland, J. W., Pettit, R. R., & Sung, S. K. (1982). The distribution of foreign exchange price changes: trading day effects and risk measurement. the Journal of Finance, 37(3), 693-715.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information. The journal of finance, 42(3), 483-510.

Miller, E. M. (1977). Risk, uncertainty, and divergence of opinion. The Journal of Finance, 32(4), 1151-1168.

Miller, E. M. (1988). Why a weekend effect. The Journal of Portfolio Management, 14(4), 43-48.

Mills, T. C., & Andrew Coutts, J. (1995). Calendar effects in the London Stock Exchange FT–SE indices. The European Journal of Finance, 1(1), 79-93.

Mitchell, M. L., & Mulherin, J. H. (1994). The impact of public information on the stock market. The Journal of Finance, 49(3), 923-950.

Monteiro, J. (2012). Sensitivity of seasonality effects on mean and conditional volatility to error distributional assumptions: evidence from French stock market. Ecorfan Journal, 3(8), 879-898.

Montes, G. C., & Peixoto, G. B. T. (2014). Risk-taking channel, bank lending channel and the "paradox of credibility": Evidence from Brazil. Economic Modelling, 39, 82-94.

Morse, D. (1980). Asymmetrical information in securities markets and trading volume. Journal of Financial and Quantitative Analysis, 15(05), 1129-1148.

Mougoué, M., & Aggarwal, R. (2011). Trading volume and exchange rate volatility: Evidence for the sequential arrival of information hypothesis. Journal of Banking & Finance, 35(10), 2690-2703.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica: Journal of the Econometric Society, 347-370.

Odean, T. (1998). Are investors reluctant to realize their losses?. The Journal of finance, 53(5), 1775-1798.

Otoo, M. W. (1999). Consumer sentiment and the stock market. US Federal Reserve Working Paper.

Park, Y. S., & Lee, J. J. (2003). An empirical study on the relevance of applying relative valuation models to investment strategies in the Japanese stock market. Japan and the World Economy, 15(3), 331-339.

Patel, S.A. and A. Sarkar, (1998). Crises in developed and emerging stock markets, Financial Analysts Journal 54, 50–61.Mishkin and White (2002

Patell, J. M., & Wolfson, M. A. (1982). Good news, bad news, and the intraday timing of corporate disclosures. Accounting Review, 509-527.

Penman, S. H. (1987). The distribution of earnings news over time and seasonalities in aggregate stock returns. Journal of Financial Economics, 18(2), 199-228.

Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. Journal of financial economics, 22(1), 27-59.

Randall, M. R., Suk, D. Y., & Tully, S. W. (2003). Mutual fund cash flows and stock market performance. The Journal of Investing, 12(1), 78-80.

Redman, A., Manakyan, H., & Liano, K. (1997). Real estate investment trusts and calendar anomalies. Journal of Real Estate Research, 14(1), 19-28.

Rogalski, R. J. (1984). New findings regarding day-of-the-week returns over trading and non-trading periods: a note. The Journal of Finance, 39(5), 1603-1614.

Rystrom, D. S., & Benson, E. D. (1989). Investor psychology and the day-of-the-week effect. Financial Analysts Journal, 45(5), 75-78.

Sapienza, P., Zingales, L., & Guiso, L. (2006). Does culture affect economic outcomes? (No. w11999). National Bureau of Economic Research.

Scheinkman, J. A., & Xiong, W. (2003). Overconfidence and speculative bubbles. Journal of political Economy, 111(6), 1183-1220.

Schmeling, M., (2009). Investor sentiment and stock returns: Some international evidence, Journal of Empirical Finance 16, 394–408.

Schwert, G. W. (2011). Stock volatility during the recent financial crisis. European Financial Management, 17(5), 789-805.

Sharma, J. L., Mougoue, M., & Kamath, R. (1996). Heteroscedasticity in stock market indicator return data: volume versus GARCH effects. Applied Financial Economics, 6(4), 337-342.

Shefrin, H. (1999). Beyond greed and fear: Understanding behavioral finance and the psychology of investing. Oxford University Press.

Shiller, R. J. (1981). The Use of Volatility Measures in Assessing Market Efficiency*. The Journal of Finance, 36(2), 291-304.

Shiller, R. J. (1987). Investor behavior in the October 1987 stock market crash: Survey evidence.

Shiller, R. J. (2005). Irrational exuberance. Random House Digital, Inc.

Shiller, R. J. (2014). Speculative asset prices. American Economic Review, 104(6), 1486-1517.

Shleifer, A. (2000). Inefficient markets: An introduction to behavioral finance. Oxford university press.

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance.The Journal of Economic Perspectives, 19-33.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. The Journal of Finance, 52(1), 35-55.

Siegel, J. J. (1992). Equity risk premia, corporate profit forecasts, and investor sentiment around the stock crash of October 1987. Journal of Business, 557-570.

Singleton, J. C., & Wingender, J. R. (1994). The nonparallel weekend effect in the stock and bond markets. Journal of Financial Research, 17(4), 531-538.

Smidt, S. (1968). A new look at the random walk hypothesis. Journal of Financial and Quantitative Analysis, 3(3), 235-261.

Smith, A. (1776). The wealth of Nations. New York: The Modern Library.

Solnik, B., & Bousquet, L. (1990). Day-of-the-week effect on the Paris Bourse. Journal of Banking & Finance, 14(2), 461-468.

Srinivasan, P. and Kalaivani, M.(2014), Day-of-the-Week Effects in the Indian Stock Market, International Journal of Economics and Management, 8(1), 1

Srinivasan, P., & Kalaivani, M. (2015). Day-of-the-Week Effects in the Indian stock market.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. Journal of Financial Economics, 104(2), 288-302.

Stone, A. A., Hedges, S. M., Neale, J. M., & Satin, M. S. (1985). Prospective and cross-sectional mood reports offer no evidence of a" blue Monday" phenomenon. Journal of Personality and Social Psychology, 49(1), 129.

System ? Towards a Unified Statistical Framework for Assessing Financial Crises

Tauchen, G. E., & Pitts, M. (1983). The price variability-volume relationship on speculative markets. Econometrica: Journal of the Econometric Society, 485-505.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information?. Review of Financial Studies, 24(5), 1481-1512.

Thatcher, J. S., & Blenman, L. P. (2001). Synthetic Trades and Calendar Day Patterns The Case of the Dollar/Sterling Markets. Financial Review, 36(2), 177-200.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. science, 185(4157), 1124-1131.

Vega, C. (2006). Stock price reaction to public and private information. Journal of Financial Economics, 82(1), 103-133.

Venables, P. H., & Christie, M. J. (1974). Neuroticism, physiological state and mood: An exploratory study of Friday/Monday changes. Biological psychology, 1(3), 201-211.

Wong, J., Wong, T. C., & Leung, P. (2010). Predicting banking distress in the EMEAP economies. Journal of Financial Stability, 6(3), 169-179.

Wurgler, J., Baker, M., & Yuan, Y. (2009). Global, local, and contagious investor sentiment.

Xie, L., Zhou, M., & Sun, M. (2012). Hierarchical structure based hybrid approach to sentiment analysis of chinese micro blog and its feature extraction. Journal of Chinese Information Processing, 26(1), 73-82.

Xing, Y., Zhang, X., & Zhao, R. (2010). What does the individual option volatility smirk tell us about future equity returns?.

Yüksel, A. (2002). The performance of the Istanbul stock exchange during the Russian crisis. Emerging Markets Finance & Trade, 78-99.losed-end fund premiums. Journal of Finance, 67-78.

Ziemba, W. T. (1991). Japanese security market regularities: Monthly, turn-of-the-month and year, holiday and golden week effects. Japan and the World Economy, 3(2), 119-146.

Zou, H., & Sun, L. (2012, August). The Influence of Investor Sentiment on Stock Return and its Volatility under Different Market States. In Business Intelligence and Financial Engineering (BIFE), 2012 Fifth International Conference on (pp. 337-341). IEEE.

Zouaoui, M., Nouyrigat, G., & Beer, F. (2011). How does investor sentiment affect stock market crises? Evidence from panel data. Financial Review, 46(4), 723-747.

Zweig, M. E. (1973). An investor expectations stock price predictive model using c Admati,

Appendix 1: Essay 1- De-Trended Volume of Trades

Equation	FTSE-Vol.	SP 500-Vol.	DJIA-Vol.
Date	-374.165	1,883,745.85	-20,549.18
	(0.00)***	(0.00)***	(0.00)***
Constant	7,978,533.28	-29224682364	615,955,820.34
	(0.00)***	(0.00)***	(0.00)***
R^2	0.31	0.5	0.02
N	1,485	1,486	1,486

* p<0.1; ** p<0.05; *** p<0.01

Appendix 2: Essay 1- E-GARCH Results

GARCH (1,1)

$r_t = \omega + \mu r_{t-1} + \varepsilon_t$

$\sigma_t^2 = \alpha + \beta \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2$

T=1486	Estimates of parameters						Diagnostics	Schwartz Tests
Returns	ω	μ	α	β	γ	β+γ	LR test	AIC/BIC
FTSE	0.00057	-0.05794	1.24e-06	0.12145	0.87591	0.99736	4690.97	-9371.95
p-value	(0.006)	(0.039)	(0.003)	(0.000)	(0.000)		(0.0387)	-9345.43
NASDAQ	0.00070	-0.02933	1.18e-06	0.12070	0.87802	0.99872	4659.68	-9309.376
p-value	(0.001)	(0.298)	(0.005)	(0.000)	(0.000)		(0.2981)	-9282.856
S&P 500	0.00053	-0.09500	1.44e-06	0.08910	0.90117	0.99027	4672.936	-9335.873
p-value	(0.011)	(0.002)	(0.000)	(0.000)	(0.000)		(0.0021)	-9309.354
Russell 2000	0.00074	-0.06741	3.14e-06	0.08058	0.90631	0.98689	4176.442	-8342.885

p-value	(0.020)	(0.024)	(0.000)	(0.000)	(0.000)		(0.0240)	-8316.366

**A p-value inferior to 0.01 corresponds to a coefficient statistically significant at the 1% level*

**A p-value inferior to 0.05 corresponds to a coefficient statistically significant at the 5% level*

**A p-value inferior to 0.10 corresponds to a coefficient statistically significant at the 10% level*

EGARCH (1,1)

$$r_t = \omega + \mu r_{t-1} + \varepsilon_t$$

$$\sigma_t^2 = \alpha + \beta\varepsilon_{t-1}^2 + \gamma\sigma_{t-1}^2$$

T=1486	Estimates of parameters						Diagnostics	Schwartz Tests
Returns	ω	μ	α	β	γ	β+γ	LR test	AIC/BIC
FTSE	0.00051	-0.06747	-0.98702	0.89666	313.107	314.003	4574.498	-9138.996
p-value	(0.052)	(0.005)	(0.000)	(0.000)	(0.000)		(0.0046)	-9112.477
NASDAQ	0.00060	-0.04138	-0.93670	0.90149	289.882	290.7834	4547.838	-9085.676
p-value	(0.031)	(0.089)	(0.000)	(0.000)	(0.000)		(0.0886)	-9059.157
S&P 500	0.00041	-0.10430	-0.68441	0.92804	182.29	183.21	4558.553	-9107.106

p-value	(0.098)	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	-9080.587
Russell 2000	0.00066	-0.07019	-0.73143	0.91832	137.79	138.70	4123.759	-8237.519
p-value	(0.055)	(0.008)	(0.000)	(0.000)	(0.000)		(0.000)	-8211

**A p-value inferior to 0.01 corresponds to a coefficient statistically significant at the 1% level*

**A p-value inferior to 0.05 corresponds to a coefficient statistically significant at the 5% level*

**A p-value inferior to 0.10 corresponds to a coefficient statistically significant at the 10% level*

Appendix 3: Essay 2 - ARCH Effects

Table 76 Test Results for Autoregressive Conditional Heteroskedasticity

		NASDAQ	RUSSEL
AR Specification		AR(4)	AR(3)
Cameron and Trivedi's Information Matrix	Heteroskedasticity	382.79	413.04
	p-value	(0.000)	(0.000)
	Skewness	38.35	16.53
	p-value	(0.000)	(0.000)
	Kurtosis	11.72	13.86
	p-value	(0.000)	(0.000)
Engle's Test	F-Test specification	F (1464)	F (1,478)
	F-Test result	47.79	131.39
	Prob > F	(0.000)	(0.000)

Appendix 4: Essay 2 - Additional Results

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
FTSE_Returns	Interaction V.	0	0	0	0	0
		-0.53	-0.3	-0.27	-0.29	-0.8
	Dummy	-0.001	0	-0.002	-0.001	0.001
		-0.61	-0.99	-0.14	-0.54	-0.37
	Constant	0.001	0.001	0.001	0.001	0
		(0.03)**	(0.00)***	(0.01)***	(0.04)**	-0.19
ARMA	L.ar	-0.051	-0.04	-0.047	-0.048	-0.048
		-0.23	-0.35	-0.27	-0.26	-0.25
ARCH	L.arch	0.117	0.134	0.128	0.131	0.117
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.83	0.801	0.811	0.806	0.828
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	Interaction V.	-13.378	0.02	0.005	-0.817	0.082
		(0.00)***	-0.37	-0.9	-0.91	-0.15
	Dummy	136.536	2.478	1.55	-3.159	-0.668
			-0.19	-0.27		-0.81
	Constant	-12.638	-13.931	-13.253	-12.381	-12.835
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,305.91	2,308.47	2,306.32	2,305.97	2,307.36
Prob>chi2		0	0	0	0	0

Number of obs	643	643	643	643	643

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
FTSE_Returns	Interaction V.	0	0	0	0	0
		-0.69	-0.28	(0.09)*	-0.86	-0.44
	Dummy	0	-0.002	-0.003	-0.001	-0.002
		-0.91	-0.24	-0.23	-0.82	-0.4
	Constant	0	0.001	0	0.001	0.001
		-0.38	(0.05)*	-0.27	(0.05)*	(0.07)*
ARMA	L.ar	-0.06	-0.062	-0.064	-0.065	-0.069
		-0.12	(0.09)*	(0.08)*	(0.07)*	(0.06)*
ARCH	L.arch	0.106	0.107	0.107	0.104	0.111
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.877	0.877	0.875	0.878	0.873
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	Interaction V.	0.012	-0.159	-1.423	-0.438	0.756
		-0.63	-0.36	-0.25	-0.3	-0.57
	Dummy	11.188	5.68	12.639	4.739	-26.16
			-0.63	-0.13	-0.21	-0.62
	Constant	-22.221	-14.737	-12.214	-12.315	-12.201
		(0.00)***	-0.22	(0.00)***	(0.00)***	(0.00)***

Log-Likelihood	2,400.06	2,399.93	2,400.71	2,399.05	2,399.33
Prob>chi2	0	0	0	0	0
Number of obs	843	843	843	843	843

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
nasdaq_return	Interaction V.	0	0	0	0	0
		-0.24	-0.26	-0.21	-0.6	-0.26
	Dummy	0.002	0	-0.002	0	-0.001
		-0.29	-0.92	(0.08)*	-0.79	-0.4
	Constant	0.001	0.001	0.001	0.001	0.001
		(0.01)***	(0.04)**	(0.00)***	(0.01)***	(0.01)***
ARMA	L.ar	-0.028	-0.022	-0.029	-0.025	-0.024
		-0.52	-0.6	-0.49	-0.56	-0.57
ARCH	L.arch	0.154	0.169	0.167	0.16	0.158
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.803	0.776	0.778	0.797	0.801
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HET	Interaction V.	0.021	-0.007	-0.436	-0.028	-0.03
		-0.36	-0.86	-0.96	-0.86	-0.77
	Dummy	1.41	1.66	-6.093	-0.191	0.874
		-0.43	-0.18		-0.95	-0.63
	Constant	-13.394	-13.121	-12.31	-12.638	-12.909

	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood	2,267.92	2,268.69	2,269.09	2,266.51	2,267.02
Prob>chi2	0	0	0	0	0
Number of obs	643	643	643	643	643

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
nasdaq_return	Interaction V.	0	0	0	0	0
		-0.43	-0.79	-0.14	-0.76	-0.36
	Dummy	0.004	0	-0.004	0.002	0
		-0.17	-0.86	-0.15	-0.48	-0.85
	Constant	0	0.001	0.001	0.001	0.001
		-0.32	(0.10)*	(0.07)*	-0.22	(0.01)***
ARMA	L.ar	-0.022	-0.028	-0.031	-0.027	-0.023
		-0.55	-0.45	-0.4	-0.47	-0.54
ARCH	L.arch	0.109	0.094	0.096	0.103	0.097
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.879	0.886	0.881	0.877	0.883
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HET	Interaction V.	0.071	-0.524	-0.431	-0.396	-0.349
		-0.51	-0.17	-0.14	(0.10)*	-0.31
	Dummy	-3.175	6.507	6.37	5.595	5.735
		-0.68	(0.03)**	(0.01)***	(0.01)**	(0.01)***

	Constant	-12.431	-12.588	-12.704	-12.523	-12.949
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,406.12	2,405.66	2,407.87	2,405.81	2,408.62
Prob>chi2		0	0	0	0	0
Number of obs		843	843	843	843	843

* *p*<0.1; ** *p*<0.05; *** *p*<0.01

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
sp_return	Interaction V.	0	0	0	0	0
		-0.67	-0.5	-0.2	-0.51	-0.96
	Dummy	0	0	0	-0.001	-0.001
		-0.79	-0.87	-0.96	-0.41	-0.61
	Constant	0	0	0	0.001	0.001
		-0.11	(0.05)*	-0.47	(0.06)*	(0.04)**
ARMA	L.ar	-0.061	-0.049	-0.056	-0.063	-0.063
		-0.15	-0.24	-0.18	-0.14	-0.14
ARCH	L.arch	0.054	0.075	0.053	0.053	0.052
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.889	0.818	0.887	0.893	0.889
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HET	Interaction V.	-0.306	-0.003	-0.879	0.036	0.115
		-0.98	-0.87	-1	-0.75	-0.78

	Dummy	-5.751	14.844	-8.574	-1.066	-3.739
sp_return			-0.87	-0.94	-0.82	-0.81
	Constant	-12.61	-25.44	-12.549	-12.751	-12.583
		(0.00)***	-0.78	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,307.06	2,315.70	2,309.33	2,306.54	2,306.87
Prob>chi2		0	0	0	0	0
Number of obs		643	643	643	643	643

* p<0.1; ** p<0.05; *** p<0.01

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
sp_return	Interaction V.	0	0	0	0	0
		(0.07)*	-0.57	-0.1	-0.59	-0.85
	Dummy	-0.003	-0.001	-0.003	0.001	-0.002
		-0.21	-0.66	-0.25	-0.79	-0.51
	Constant	0.001	0.001	0.001	0.001	0.001
		-0.18	-0.12	-0.23	(0.05)*	(0.03)**
ARMA	L.ar	-0.111	-0.115	-0.11	-0.113	-0.112
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
ARCH	L.arch	0.104	0.081	0.099	0.101	0.099
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.885	0.905	0.89	0.892	0.889
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	Interaction V.	-0.394	0.936	-0.74	0.064	-0.43

		-0.25	-0.39	-0.37	-0.34	-0.45
	Dummy	8.457	-59.587	7.687	6.448	5.622
		-0.12	-0.43	-0.3	-0.84	-0.19
	Constant	-12.595	-12.651	-12.708	-18.758	-12.912
		(0.00)***	(0.00)***	(0.00)***	-0.55	(0.00)***
Log-Likelihood		2,381.23	2,383.03	2,380.85	2,380.39	2,380.50
Prob>chi2		0	0	0	0	0
Number of obs		843	843	843	843	843

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
russel_return	Interaction V.	0	0	0	0	0
		-0.93	-0.8	-0.13	-0.22	-0.55
	Dummy	-0.001	-0.001	0	-0.002	-0.001
		-0.83	-0.76	-0.88	-0.27	-0.44
	Constant	0.001	0.001	0	0.001	0.001
		-0.2	-0.14	-0.79	-0.21	-0.16
ARMA	L.ar	-0.004	-0.001	0	-0.007	-0.008
		-0.93	-0.98	-1	-0.87	-0.85
ARCH	L.arch	0.049	0.053	0.048	0.051	0.043
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.01)***
	L.garch	0.896	0.886	0.899	0.89	0.901
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
HET	Interaction V.	-0.007	-0.016	-0.433	0.009	0.278

		-0.96	-0.5	-0.99	-1	-0.52
	Dummy	-0.587	13.255	-6.03	-5.766	-7.902
		-0.9	-0.88	-0.96	-0.94	-0.59
	Constant	-11.889	-23.391	-11.809	-11.685	-11.817
		(0.00)***	-0.79	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,023.33	2,027.33	2,026.50	2,024.18	2,024.69
Prob>chi2		0	0	0	0	0
Number of obs		643	643	643	643	643

* p<0.1; ** p<0.05; *** p<0.01

Equation	Variable	Monday	Tuesday	Wednesday	Thursday	Friday
russel_return	Interaction V.	0	0	0	0	0
		(0.06)*	-0.67	-0.42	-0.29	-0.57
	Dummy	-0.005	0	-0.001	0.002	0.001
		-0.12	-0.87	-0.72	-0.55	-0.85
	Constant	0.001	0.001	0.001	0.001	0.001
		(0.04)**	(0.06)*	-0.1	(0.01)**	(0.01)**
ARMA	L.ar	-0.107	-0.11	-0.108	-0.108	-0.109
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
ARCH	L.arch	0.101	0.083	0.097	0.095	0.097
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
	L.garch	0.889	0.906	0.894	0.895	0.894
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***

HET	Interaction V.	-0.326	0.525	0.364	-0.245	-0.407
		-0.33	-0.48	-0.77	-0.65	-0.53
	Dummy	7.327	-31.391	-9.661	4.614	4.945
		-0.21	-0.54	-0.81	-0.14	-0.33
	Constant	-12.25	-12.331	-12.3	-12.995	-12.499
		(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***
Log-Likelihood		2,164.34	2,164.94	2,163.45	2,163.83	2,163.17
Prob>chi2		0	0	0	0	0
Number of obs		843	843	843	843	843

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

Appendix 5: Essay 3 - Crisis in Indices and CMAX Indicator Threshold Graphs

Figure 18 FTSE Crises Chart

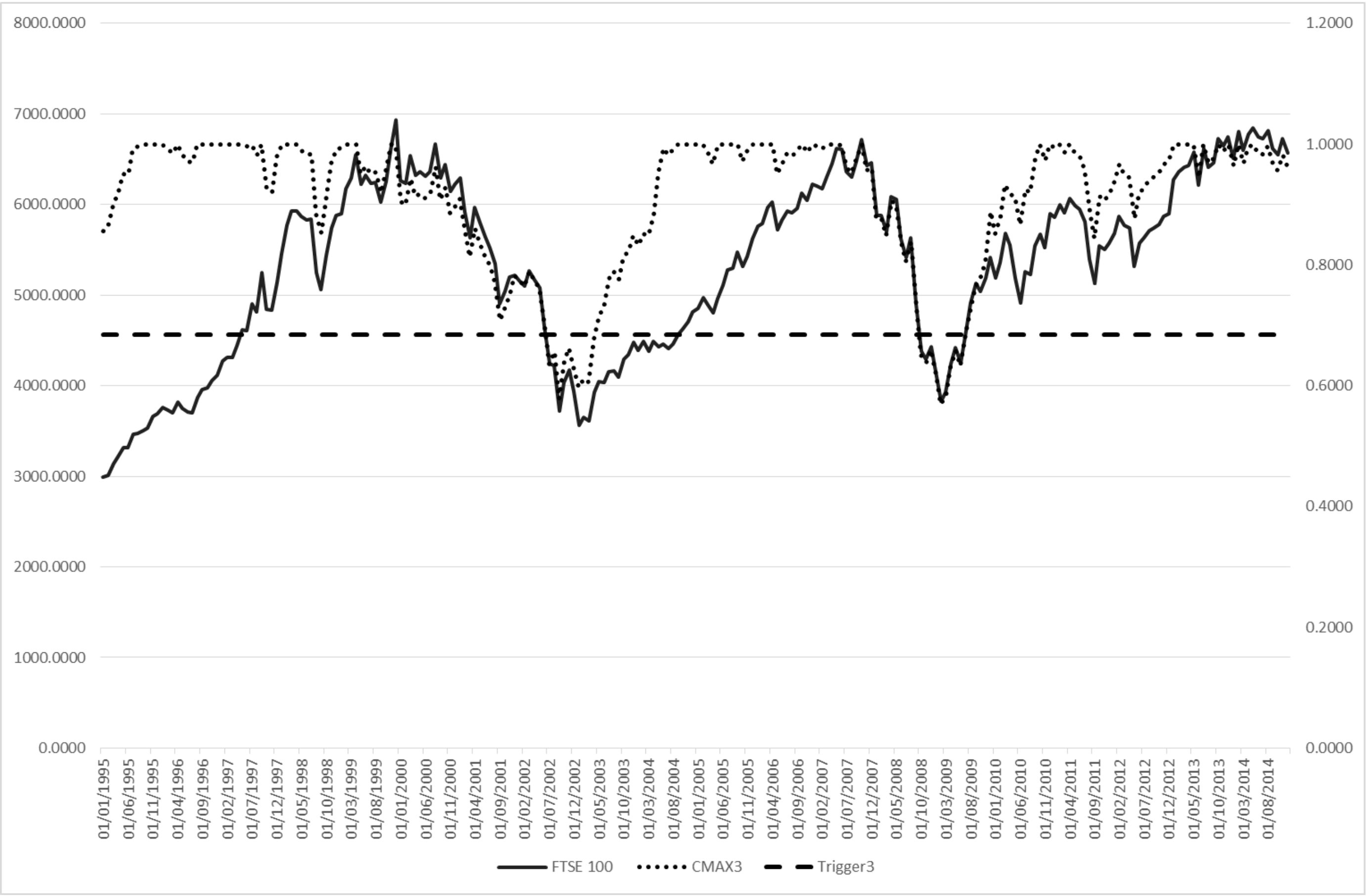


Figure 19 DJIA Crises Chart

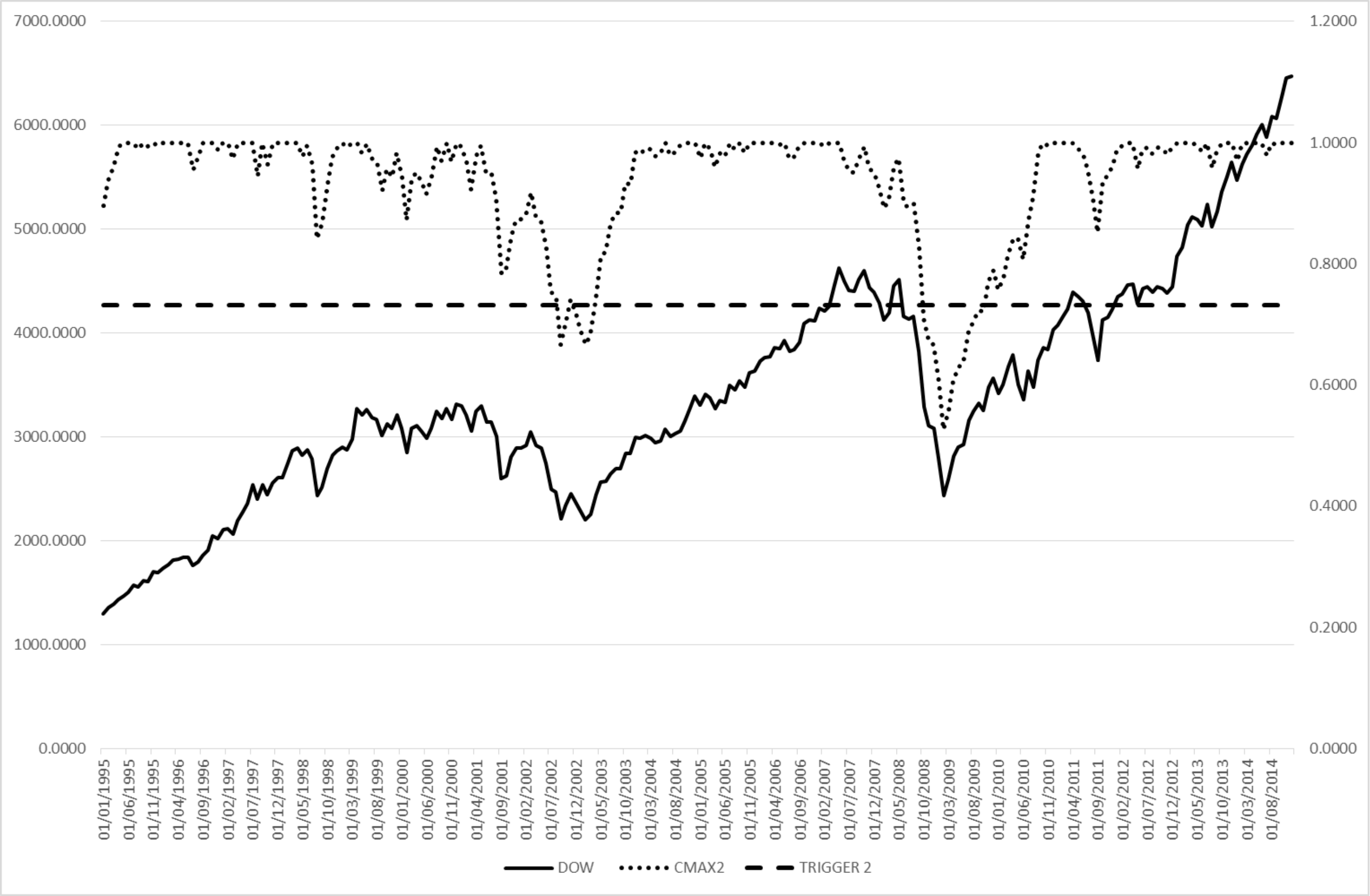


Figure 20 FTSE Crises Graph

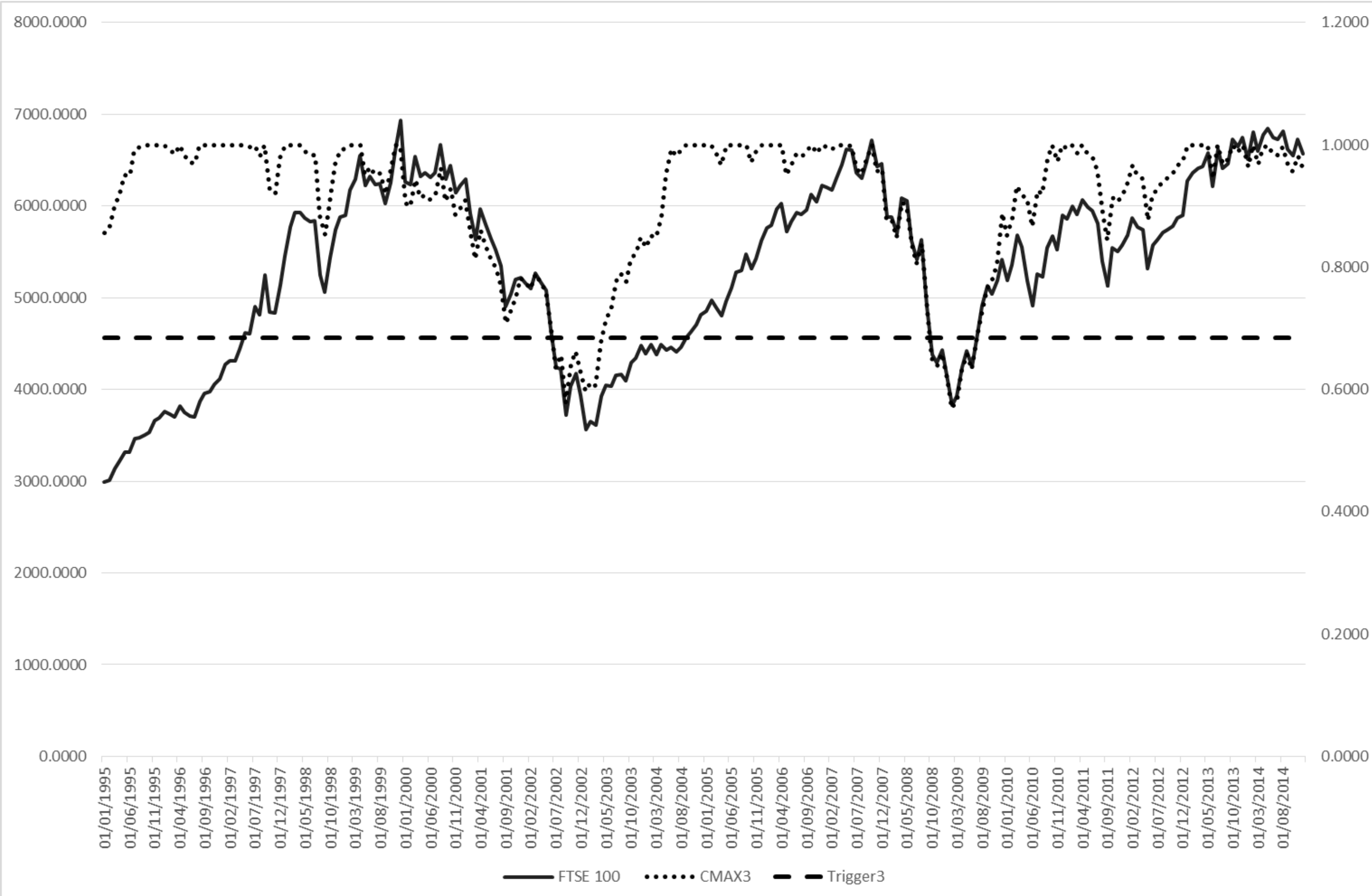


Figure 21 DAX Crises Graph

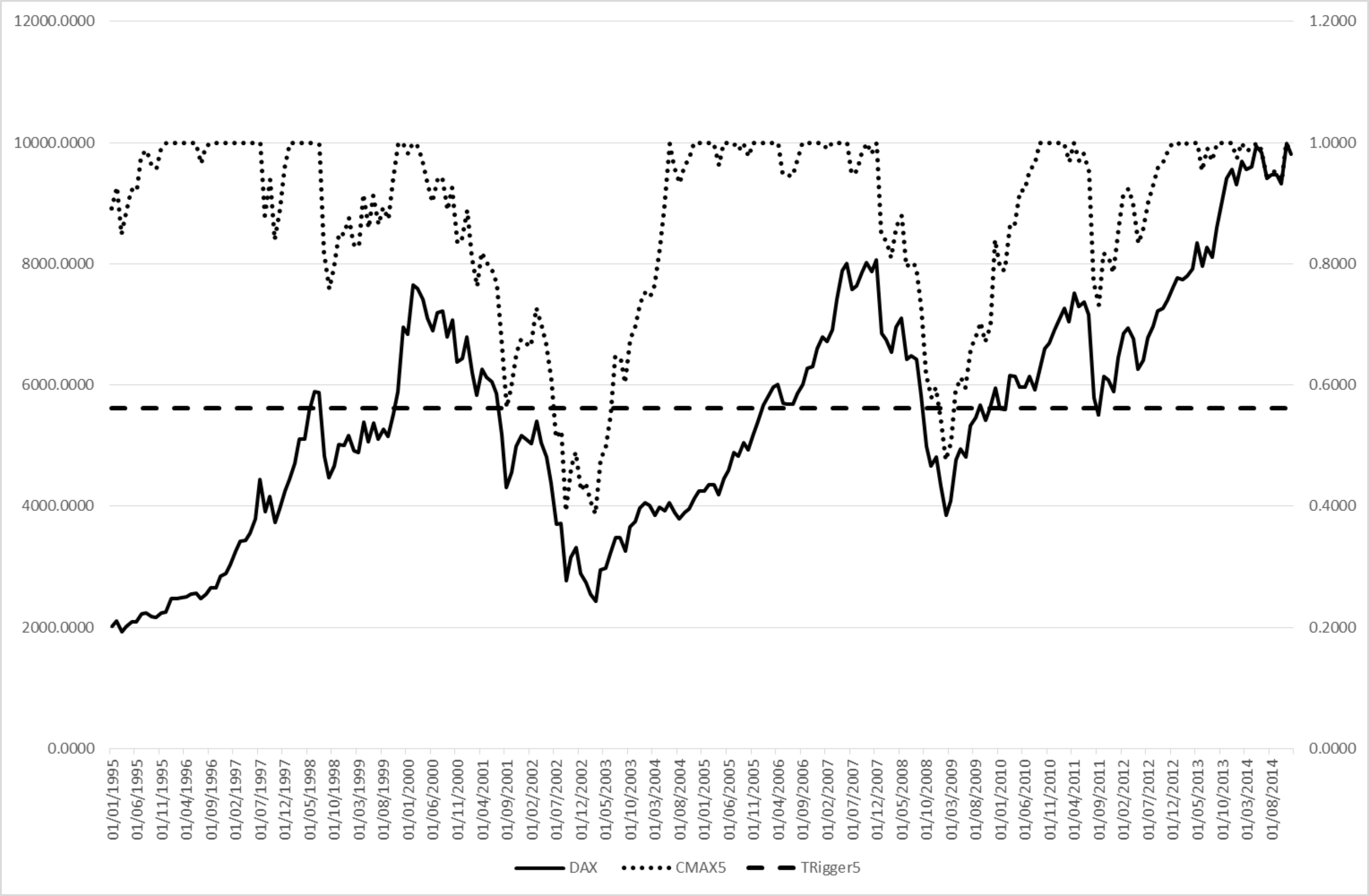
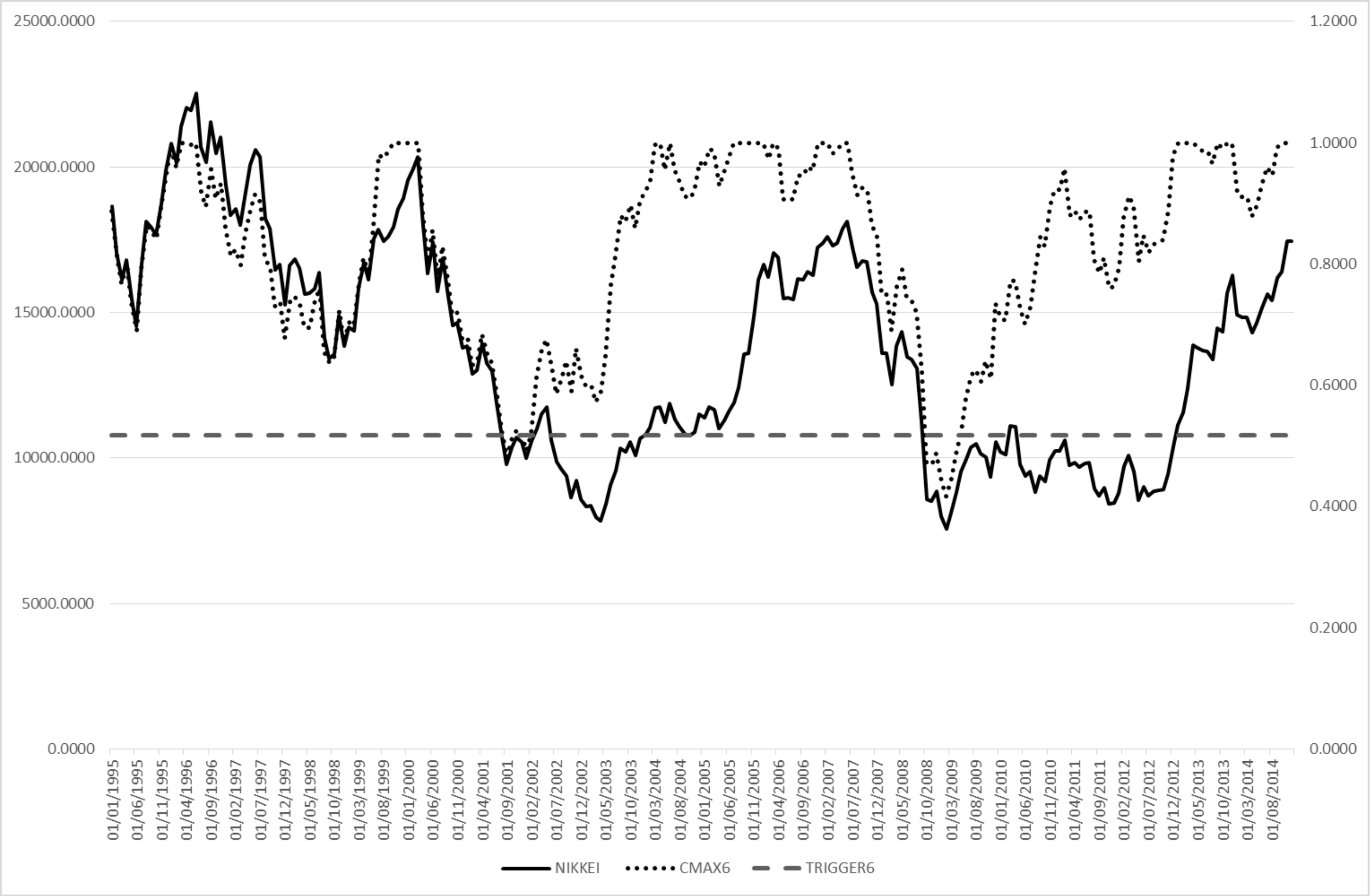


Figure 22 NIKKEI Crises Chart



Appendix 6: Essay 3 - Sources of Data

ABBREVIATION	Variable	Quantifies	Source of Data
STOCK MARKET VARIABLES			
P	Stock price index	Stock Prices	Datastream
CAPE	Cyclically adjusted Price to Earnings ratio	Cyclically adjusted price to earnings data for the US markets	Yale
INVESTOR SENTIMENT VARIABLES			
VIX	Consumer Sentiment	Monthly data assessing volatility in options index	Chicago Board of Exchange
PCR	Consumer Sentiment	Monthly data assessing ratio of put to call options	Chicago Board of Exchange
BWSI	Orthogonalised Investor Sentiment Components	Stock Market Riskiness	Baker and Wurgler, 2006,2008
MACROECONOMIC VARIABLES			
IPROD	Industrial Production	Change in log of Industrial production	International Financial Statistics
TS	Term Spread	Difference between yields on 10 Government bonds and 3 month treasury bills	International Financial Statistics
GDP	Gross domestic Production	Gross Domestic Production	US Federal Reserve Bank, European Central Bank, Japan Central Bank
IFR	Inflation Rate	Change in the natural log of the consumer price index	International Financial Statistics
INT	Interest Rates	Money market rate using the CPI	International Financial Statistics

Appendix 7: Essay 3- Recovery in Indices Graphs

Figure 23 DJIA Recovery Metric Chart

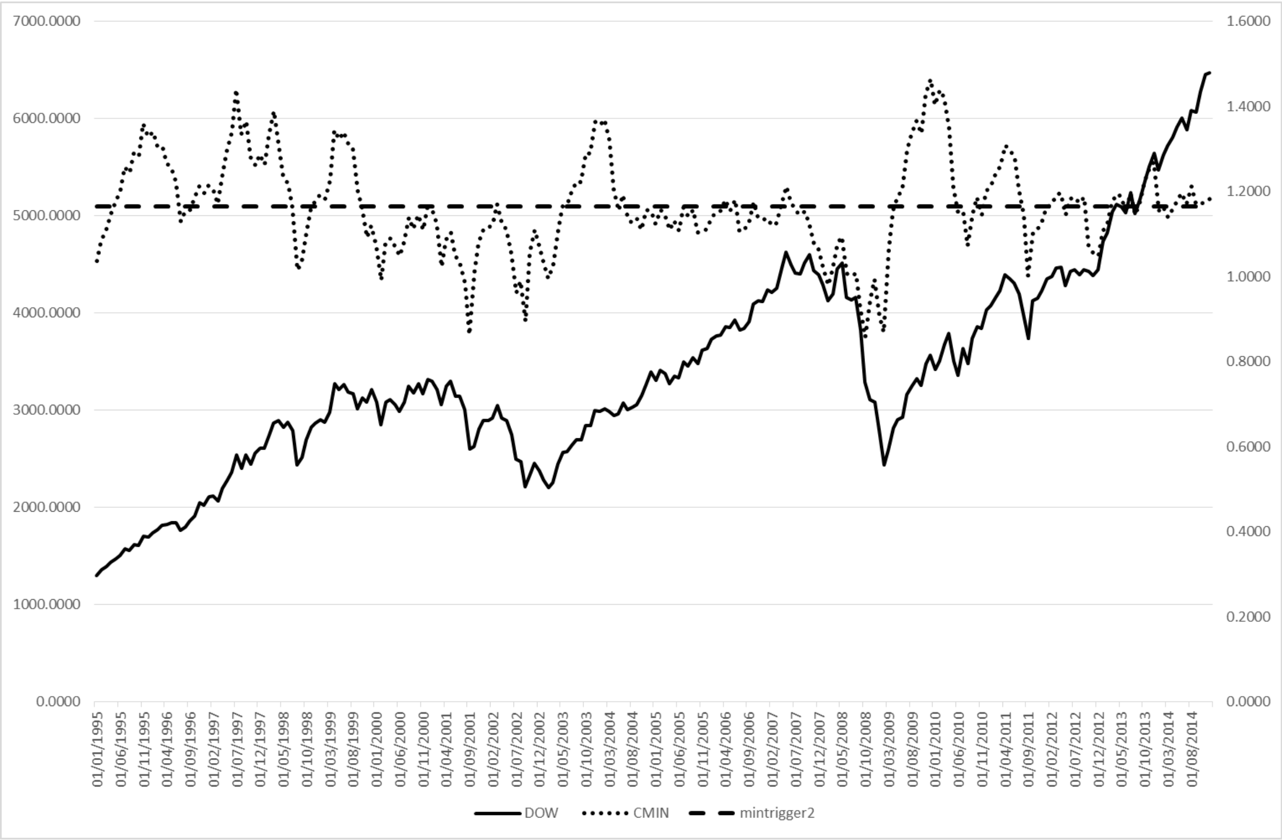


Figure 24 FTSE Recovery Metric Chart

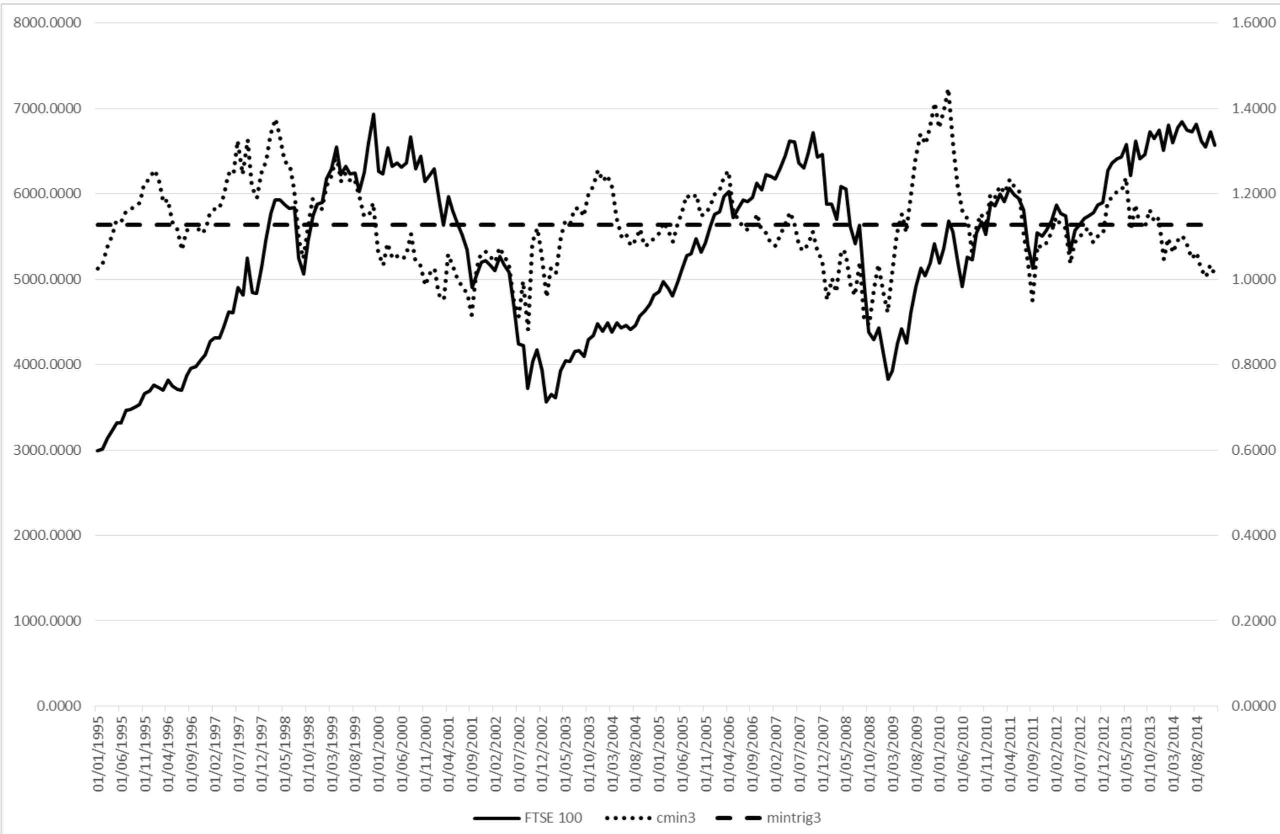


Figure 25 CAC Recovery Metric Chart

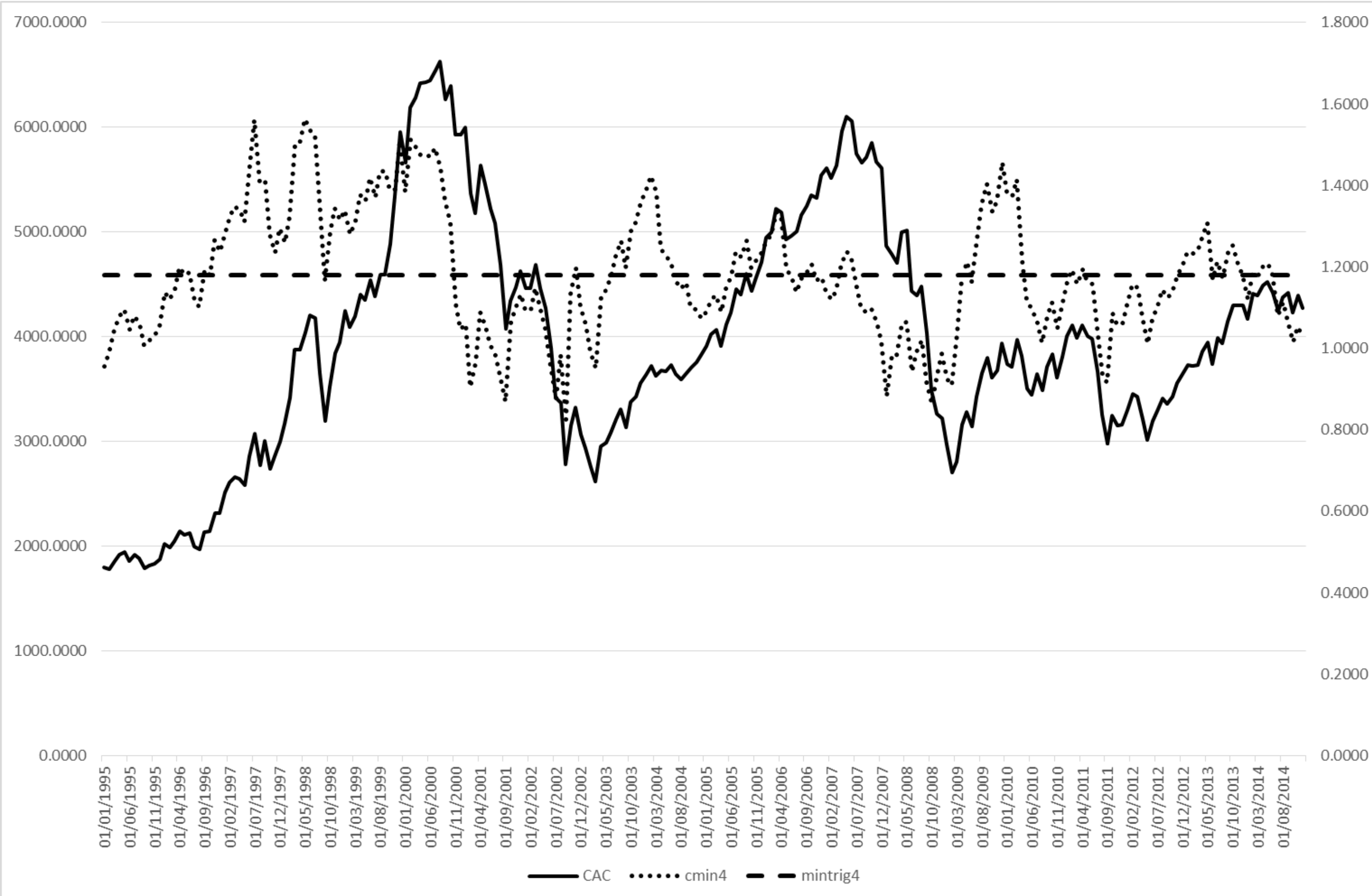


Figure 26 DAX Recovery Metric Chart

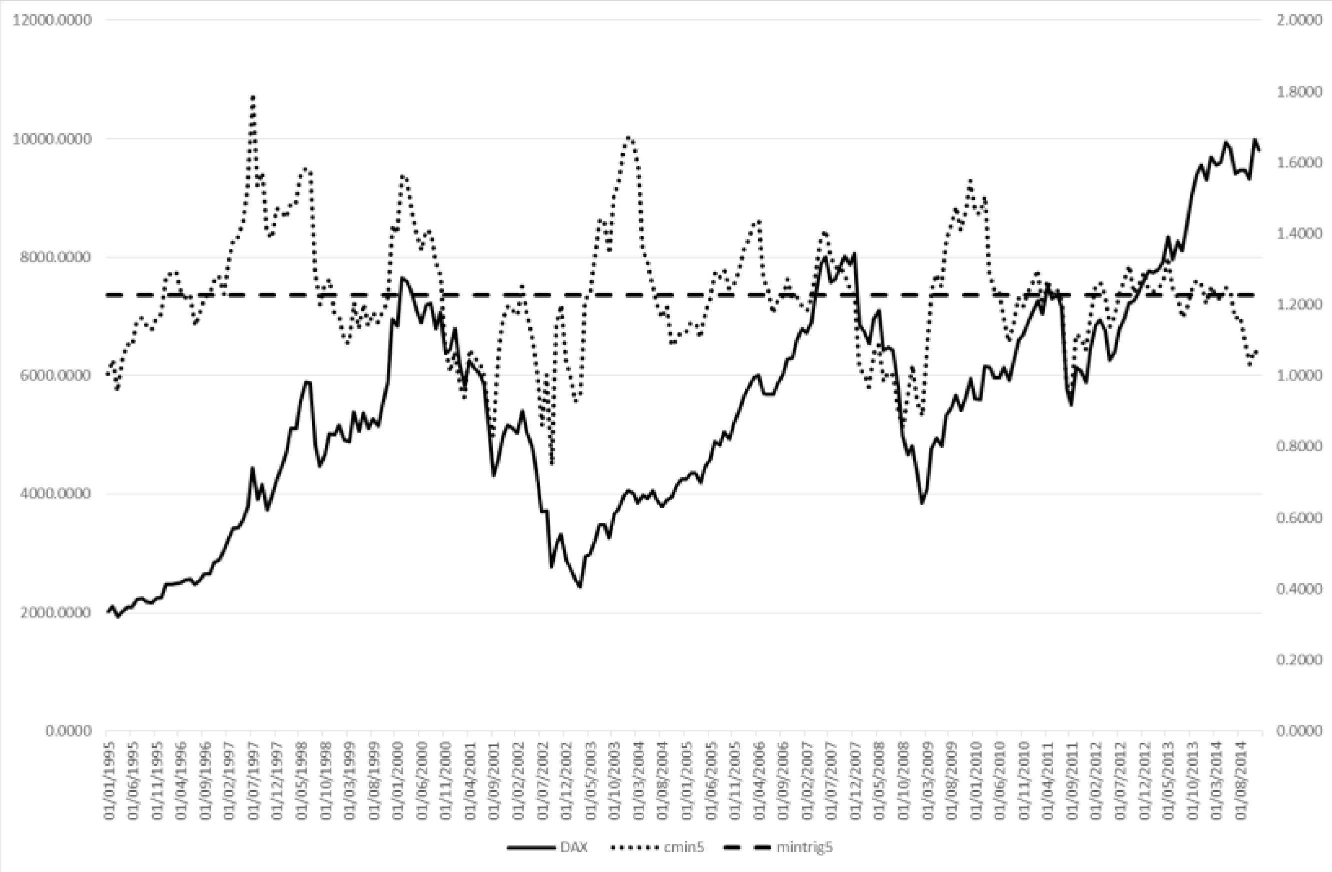
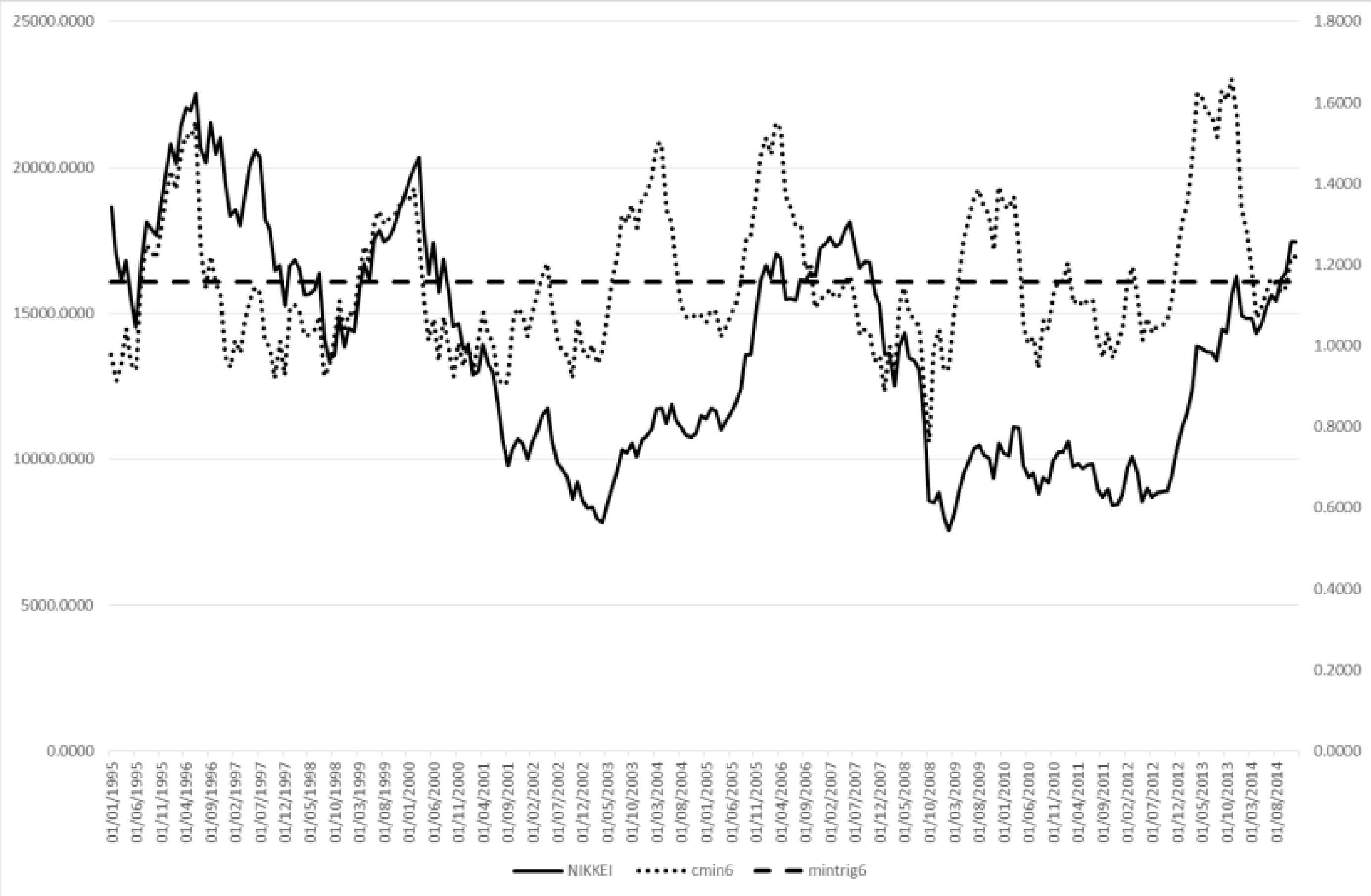


Figure 27 NIKKEI Recovery Metric Chart



This Page was intentionally left blank.